STOCHASTIC SIGNALING FOR POWER CONSTRAINED COMMUNICATION SYSTEMS

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND

ELECTRONICS ENGINEERING

AND GRADUATE SCHOOL OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

By

Çağrı Göken

June 2011

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

Asst. Prof. Dr. Sinan Gezici (Supervisor)

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

Prof. Dr. Orhan Arıkan

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

Asst. Prof. Dr. Ali Cafer Gürbüz

Approved for the Graduate School of Engineering and Science:

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

ABSTRACT

STOCHASTIC SIGNALING FOR POWER CONSTRAINED COMMUNICATION SYSTEMS

Çağrı Göken
M.S. in Electrical and Electronics Engineering
Supervisor: Asst. Prof. Dr. Sinan Gezici
June 2011

In this thesis, optimal stochastic signaling problem is studied for power constrained communications systems. In the first part, optimal stochastic signaling problem is investigated for binary communications systems under second and fourth moment constraints for any given detector structure and noise probability distribution. It is shown that an optimal signal can be represented by randomization among at most three signal levels for each symbol. Next, stochastic signaling problem is studied in the presence of an average power constraint instead of second and fourth moment constraints. It is shown that an optimal signal can be represented by randomization between at most two signal levels for each symbol in this case. For both scenarios, sufficient conditions are obtained to determine the improvability and nonimprovability of conventional deterministic signaling via stochastic signaling. In the second part of the thesis, the joint design of optimal signals and optimal detector is studied for binary communications systems under average power constraints in the presence of additive non-Gaussian noise. It is shown that the optimal solution involves randomization between at most two signal levels and the use of the corresponding maximum *a posteriori* probability (MAP) detector. In the last part of the thesis, stochastic signaling

is investigated for power-constrained scalar valued binary communications systems in the presence of uncertainties in channel state information (CSI). First, stochastic signaling is performed based on the available imperfect channel coefficient at the transmitter to examine the effects of imperfect CSI. The sufficient conditions are derived for improvability and nonimprovability of deterministic signaling via stochastic signaling in the presence of CSI uncertainty. Then, two different stochastic signaling strategies, namely, robust stochastic signaling and stochastic signaling with averaging, are proposed for designing stochastic signals under CSI uncertainty. For the robust stochastic signaling problem, sufficient conditions are derived to obtain an equivalent form which is simpler to solve. In addition, it is shown that optimal signals for each symbol can be written as randomization between at most two signal levels for stochastic signaling using imperfect channel coefficient and stochastic signaling with averaging as well as for robust stochastic signaling under certain conditions. The solutions of the optimal stochastic signaling problems are obtained by using global optimization techniques, specifically, Particle Swarm Optimization (PSO), and by employing convex relaxation approaches. Numerical examples are presented to illustrate the theoretical results at the end of each part.

Keywords: Stochastic signaling, probability of error, additive noise channels, detection, binary communications, MAP decision rule, global optimization, channel state information.

ÖZET

GÜÇ KISITLAMALI HABERLEŞME SİSTEMLERİ İÇİN STOKASTİK İŞARETLEME

Çağrı Göken
Elektrik ve Elektronik Mühendisliği Bölümü Yüksek Lisans
Tez Yöneticisi: Asst. Prof. Dr. Sinan Gezici
Haziran 2011

Bu tezde, güç kısıtlı haberleşme sistemleri için optimal stokastik işaretleme problemi çalışılmaktadır. İlk kısımda, herhangi bir sezici ve gürültü olasılık dağılımı ele alınarak, ikinci ve dördüncü moment kısıtlamaları altında ikili haberleşme sistemleri için optimal stokastik işaretleme problemi incelenmektedir. Her bir sembol için, optimal işaretlemenin, en fazla üç işaret seviyesi arasındaki rastgeleleştirme ile ifade edilebileceği gösterilmektedir. Sonrasında, stokastik işaretleme problemi ikinci ve dördüncü moment kısıtlamaları yerine, ortalama güç kısıtlaması altında çalışılmaktadır. Bu durumda, her sembol için, optimal bir işaretin en fazla iki işaret seviyesi arasındaki rastgeleleştirme ile ifade edilebileceği gösterilmektedir. Her iki senaryo için de, klasik deterministik işaretlemenin stokastik işaretleme vasıtasıyla geliştirilebilmesi ve geliştirilememesine karar veren yeter koşullar elde edilmektedir. Tezin ikinci kısmında, ortalama güç kısıtı ve Gauss'tan farklı bir gürültü altında çalışan ikili haberleşme sistemleri için optimal sezici ve işaretlerin ortak tasarlanması çalışılmaktadır. Optimal çözümün en fazla iki işaret seviyesi arasında rastgeleleştirme ve buna karşılık gelen maksimum sonsal olasılık (MAP) sezicisinin kullanımını içerdiği gösterilmektedir. Tezin en

son kısmında stokastik işaretleme, güç kısıtlı sayıl değerli ikili haberleşme sistemleri için kanal durum bilgisi (CSI) belirsizliği altında incelenmektedir. İlk olarak, halihazırdaki hatalı kanal katsayısı kullanımına dayalı stokastik işaretleme uygulanarak, hatalı kanal durum bilgisinin etkileri incelenmektedir. CSI belirsizliği altında, deterministik işaretlemenin stokastik işaretleme vasıtasıyla geliştirilebilmesi ve geliştirilememesi için yeter koşullar elde edilmektedir. Sonrasında, CSI belirsizliği altında stokastik işaretleme tasarımı için gürbüz stokastik işaretleme ve ortalamayla stokastik işaretleme isimli iki farklı işaretleme stratejisi önerilmektedir. Gürbüz stokastik işaretleme probleminin, çözümü daha kolay olan eşdeğer bir formunun elde edilebilmesi için yeter koşullar sunulmaktadır. Ayrıca, hatalı kanal katsayısına dayalı stokastik işaretleme, ortalamayla stokastik işaretleme ve bazı koşullar altında gürbüz stokastik işaretleme için, her bir sembole özel optimal işaretin en fazla iki işaret değeri arasındaki rastgeleleştirme ile ifade edilebileceği gösterilmektedir. Optimal stokastik işaretleme problemlerinin çözümü, parçacık sürü optimizasyonu (PSO) gibi küresel optimizasyon yöntemleri veya konveks gevşetme teknikleri kullanılarak elde edilebilmektedir. Her bir kısmın sonunda, kuramsal sonuçları açıklamak için sayısal örnekler sunulmaktadır.

Anahtar Kelimeler: Stokastik işaretleme, ortalama hata olasılığı, toplanır gürültü kanalı, sezimleme, ikili haberleşme, maksimum sonsal olasılık (MAP) kuralı, küresel optimizasyon, kanal durum bilgisi.

ACKNOWLEDGMENTS

I would like to thank Asst. Prof. Dr. Sinan Gezici for his valuable guidance, time and continuous support throughout this study. It was a great pleasure and experience for me to work with him. I would also like to thank Prof. Dr. Orhan Arıkan for his constructive comments and advices to our study. In addition, I would like to thank Asst. Prof. Dr. Ali Cafer Gürbüz for agreeing to serve in my thesis committee. Finally, I would like to thank my parents and my sister for their support throughout my life.

Contents

1 INTRODUCTION **1**

1.1 Objectives and Contributions of the Thesis 1

1.2 Organization of the Thesis 6

2 OPTIMAL STOCHASTIC SIGNALING FOR POWER CONSTRAINED COMMUNICATION SYSTEMS **7**

2.1 Stochastic Signaling Under Second and Fourth Moment Constraints 8

2.1.1 System Model and Motivation 8

2.1.2 Optimal Stochastic Signaling 11

2.1.2.1 On the Optimality of the Conventional Signaling 13

2.1.2.2 Sufficient Conditions for Improvability 14

2.1.2.3 Statistical Characteristics of Optimal Signals . . 18

2.1.2.4 Calculation of the Optimal Signal 23

2.1.2.4.1 Global Optimization Approach 23

2.1.2.4.2 Convex Optimization Approach 25

2.1.3 Simulation Results 26

2.1.4 Extensions to M-ary Pulse Amplitude Modulation (PAM) 36

2.1.5 Concluding Remarks and Extensions 37

2.2 Stochastic Signaling Under an Average Power Constraint 39

2.2.1 System Model and Motivation 39

2.2.2 Optimal Stochastic Signaling 41

2.2.2.1 On the Optimality of Conventional Signaling . . 42

2.2.2.2 Sufficient Conditions for Improvability 46

2.2.2.3 Statistical Characteristics of Optimal Signals . . 48

2.2.3 Numerical Results 50

2.2.4 Concluding Remarks 54

3 OPTIMAL SIGNALING AND DETECTOR DESIGN FOR POWER CONSTRAINED COMMUNICATION SYSTEMS 55

3.1 Optimal Signaling and Detector Design 56

3.2 Numerical Results and Conclusions 60

4 STOCHASTIC SIGNALING UNDER CHANNEL STATE INFORMATION UNCERTAINTIES 64

4.1 System Model and Motivation 65

4.2 Effects of Channel Uncertainties on the Stochastic Signaling . . . 67

4.2.1 Stochastic Signaling with Imperfect Channel Coefficient . . 67

4.2.2 Stochastic Signaling versus Conventional Signaling 69

4.3 Design of Stochastic Signals Under CSI Uncertainty 74

4.3.1 Robust Stochastic Signaling 75

4.3.2 Stochastic Signaling with Averaging 78

4.4 Performance Evaluation 80

4.5 Conclusions 91

List of Figures

2.1 Probability mass functions (PMFs) of the PSO and the convex optimization algorithms for the noise PDF in (2.19). 28

2.2 Error probability versus A/σ^2 for $\kappa = 1.1$. A symmetric Gaussian mixture noise, which has its mass points at $\pm[0.3\ 0.455\ 1.011]$ with corresponding weights [0.1 0.317 0.083], is considered. 30

2.3 $G(x)$ in (2.9) for the sign detector in Fig. 2.2 at A/σ^2 values of 0, 20 and 40 dB. 30

2.4 PMFs of the PSO and the convex optimization algorithms for the sign detector in Fig. 2.2 at $A/\sigma^2 = 20$ dB. 32

2.5 PMFs of the PSO and the convex optimization algorithms for the sign detector in Fig. 2.2 at $A/\sigma^2 = 40$ dB. 32

2.6 Error probability versus A/σ^2 for $\kappa = 1.5$. A symmetric Gaussian mixture noise, which has its mass points at $\pm[0.19\ 0.39\ 0.83\ 1.03]$, each with equal weight, is considered. 33

2.7 $G(x)$ in (2.9) for the sign detector in Fig. 2.6 at A/σ^2 values of 0, 25 and 40 dB. 34

2.8 PMFs of the PSO and the convex optimization algorithms for the sign detector in Fig. 2.6 at $A/\sigma^2 = 25$ dB. 34

2.9 PMFs of the PSO and the convex optimization algorithms for the sign detector in Fig. 2.6 at $A/\sigma^2 = 40$ dB. 35

2.10 The region in which the inequality $Q(x_1) - Q(x_0) < Q(1) - Q(-1)$ is satisfied is outside of the circle $0.5\,x_0^2 + 0.5\,x_1^2 = 1$. 45

2.11 Average probability of error versus A/σ^2 for conventional, optimal deterministic, and optimal stochastic signaling. 52

3.1 Average probability of error versus A/σ^2 for the three algorithms. 62

4.1 Average probability of error versus A/σ^2 for conventional signaling and stochastic signaling with various ε values. 82

4.2 $\mathrm{P}_{\mathrm{e}}^{\hat{\alpha}}$ versus $\hat{\alpha}$ for $A/\sigma^2 = 40\,\mathrm{dB}$. The second condition in Proposition 4.1 is satisfied for $\kappa_1 = 0.04354$, $\kappa_2 = 0.01913$, $\gamma_{th} = 0.1135$, $\theta_{th} = 0.8$, $\beta_{th} = 1.038$, and $G(\sqrt{A}, \alpha)) = 0.03884$. 83

4.3 Average probability of error versus ε for stochastic signaling. At $\varepsilon_{th} = 0.413$, stochastic signaling has the same average probability of error as conventional signaling. 84

4.4 Average probability of error versus A/σ^2 for various signaling strategies. 87

4.5 Average probability of error versus Δ for stochastic signaling with averaging when $A/\sigma^2 = 40dB$ and $\epsilon = 0.05$. Stochastic signaling with averaging performs same with conventional signaling when $\Delta = 0.0078$. It has the same average probability of error as robust stochastic signaling at $\Delta = 0.0236$ and $\Delta = 0.1684$. 88

4.6 Average probability of error versus α for various signaling strategies when $A/\sigma^2 = 40dB$. 89

List of Tables

2.1 Optimal stochastic signals for the ML detectors in Fig. 2.2 (top block) and Fig. 2.6 (bottom block). 36

2.2 Optimal stochastic signaling. 53

2.3 Optimal deterministic signaling. 53

3.1 Optimal stochastic and deterministic signals for symbol 1. 63

4.1 Optimal signals for stochastic signaling for various α and robust design for symbol 1. 91

4.2 Optimal signals for stochastic signaling with averaging for symbol 1 when $A/\sigma^2 = 40dB$. 91

Dedicated to my family.

Chapter 1

INTRODUCTION

1.1 Objectives and Contributions of the Thesis

Optimal signaling in the presence of zero-mean Gaussian noise has been studied extensively in the literature [1], [2]. In binary communications systems over additive white Gaussian noise channels and under average power constraints in the form of $\mathrm{E}\{|S_i|^2\} \leq A$ for $i = 0, 1$, the average probability of error is minimized when deterministic antipodal signals ($S_0 = -S_1$) are used at the power limit ($|S_0|^2 = |S_1|^2 = A$) and a maximum *a posteriori* probability (MAP) decision rule is employed at the receiver [2]. In addition, for vector observations, selecting the deterministic signals along the eigenvector of the covariance matrix of the Gaussian noise corresponding to the minimum eigenvalue minimizes the average probability of error under power constraints in the form of $\|\mathbf{S}_0\|^2 \leq A$ and $\|\mathbf{S}_1\|^2 \leq A$ [2, pp.61–63]. In [3], the optimal deterministic signaling is investigated for nonequal prior probabilities under an average power constraint in the form of $\sum_{i=1}^{2} \pi_i \mathrm{E}\{|S_i|^2\} \leq A$, where π_i represents the prior probability of symbol i, when the noise is zero-mean Gaussian and the MAP decision rule is employed at the receiver. It is shown that the optimal signaling strategy is on-off

keying for coherent receivers when the signals have nonnegative correlation and for noncoherent receivers with any arbitrary correlation value. In addition, it is also concluded from [3] that, for coherent systems, the best performance is achieved when the signals have a correlation of -1 and the power is distributed among the signals in such a way that the Euclidean distance between them is maximized under the given power constraint. In [4], a source-controlled turbo coding algorithm is proposed for nonuniform binary memoryless sources over AWGN channels by utilizing asymmetric nonbinary signal constellations.

Although the average probability of error expressions and optimal signaling techniques are well-known when the noise is Gaussian, the noise can have significantly different probability distribution than the Gaussian distribution in some cases due to effects such as multiuser interference and jamming [5]-[7]. In [8], additive noise channels with binary inputs and scalar outputs are studied, and the worst-case noise distribution is characterized. Specifically, it is shown that the least-favorable noise distribution that maximizes the average probability of error and minimizes the channel capacity is a mixture of discrete lattices [8]. A similar problem is considered in [9] for a binary communications system in the presence of an additive jammer, and properties of optimal jammer distribution and signal distribution are obtained.

In [6], the convexity properties of the average probability of error are investigated for binary-valued scalar signals over additive noise channels under an average power constraint. It is shown that the average probability of error is a convex nonincreasing function for unimodal differentiable noise probability density functions (PDFs) when the receiver employs maximum likelihood (ML) detection. Based on this result, it is concluded that randomization of signal values (or, stochastic signal design) cannot improve error performance for the considered communications system. Then, the problem of maximizing the average probability of error is studied for an average power-constrained jammer,

and it is shown that the optimal solution can be obtained when the jammer randomizes its power between at most two power levels. Finally, the results are applied to multiple additive noise channels, and optimum channel switching strategy is obtained as time-sharing between at most two channels and power levels [6]. In [10], the results in [6] are generalized by exploring the convexity properties of the error rates for constellations with arbitrary shape, order and dimensionality for ML detector in additive white Gaussian noise (AWGN) with no fading or frequency flat slowly fading channels. Also, the discussion in [6] for optimum power/time sharing for a jammer to maximize average probability of error and optimum transmission strategy to minimize average probability of error is extended to arbitrary multidimensional constellations for AWGN channels.

Optimal randomization between two deterministic signal pairs and the corresponding ML decision rules is studied in [11] for an average power-constrained antipodal binary communications system, and it is shown that power randomization can result in significant performance improvement. In [12], the problem of pricing and transmission scheduling is investigated for an access point in a wireless network, and it is proven that the randomization between two business decision and price pairs maximizes the time-average profit of the access point. Although the problem studied in [12] is in a different context, its theoretical approach is similar to those in [6] and [11] for obtaining optimal signal distributions.

Although the average probability of error of a binary communications system is minimized by conventional *deterministic* signaling in additive Gaussian noise channels [2], the studies in [6, 9, 11, 12] imply that *stochastic* signaling can sometimes achieve lower average probability of error when the noise is non-Gaussian. Therefore, a more generic formulation of the optimal signaling problem for binary communications systems can be stated as obtaining the optimal probability distributions of signals S_0 and S_1 such that the average probability of error of

the system is minimized under certain constraints on the moments of S_0 and S_1. It should be noted that the main difference of this optimal stochastic signaling approach from the conventional (deterministic) approach [1, 2] is that signals S_0 and S_1 are considered as random variables in the former whereas they are regarded as deterministic quantities in the latter.

In the first section of Chapter 2, optimal stochastic signaling is studied under second and fourth moment constraints for a given decision rule (detector) at the receiver. Firstly, a generic formulation (i.e., for arbitrary receivers and noise probability distributions) of the optimal stochastic signaling problem is performed under *both* average power *and* peakedness constraints on individual signals. Then, sufficient conditions to determine whether stochastic signaling can provide error performance improvement compared to the conventional (deterministic) signaling are derived. Also, the statistical characterization of optimal signals is provided and it is shown that an optimal stochastic signal can be expressed as a randomization of at most three different signals levels. The power constraints achieved by optimal signals are specified under various conditions. In addition, two optimization techniques, namely particle swarm optimization (PSO) [13] and convex relaxation [14], are studied to obtain optimal and close-to-optimal solutions to the stochastic signaling problem. Also, simulation results are presented to investigate the theoretical results. Finally, it is explained that the results obtained for minimizing the average probability of error for a binary communications system can be extended to M-ary systems, as well as to other performance criteria than the average probability of error, such as the Bayes risk [2, 15]. In the second section of Chapter 2, optimal stochastic signaling based on an average power constraint in the form of $\sum_{i=1}^{2} \pi_i \mathrm{E}\{|S_i|^2\} \leq A$ is studied. Similarly to the first section, optimal stochastic signaling problem is formulated for any given fixed receiver and noise probability distribution and sufficient conditions for improvability and nonimprovability of conventional deterministic signaling via stochastic approach are obtained. In addition, the statistical structure

of the optimal stochastic signals is investigated and it is shown that an optimal stochastic signal can be represented by a randomization between at most two signal levels for each symbol. Finally, by using particle swarm optimization (PSO), optimal stochastic signals are calculated and numerical examples are presented to illustrate the theoretical results.

In Chapter 3, the joint optimization of stochastic signaling and the decision rule (detector) is studied under average power constraints on individual signals. Firstly, the joint optimization problem, which involves optimization over a function space, is formulated. Then, theoretical results are provided to show that the optimal solution can be obtained by searching over a number of variables instead of functions, which greatly simplifies the original formulation. In addition, particle swarm optimization (PSO) is employed to obtain the optimal signals with the decision rule and a numerical example is provided.

In Chapter 4, the effects of imperfect channel state information (CSI) on the performance of stochastic signaling and the design of stochastic signals under CSI uncertainty are studied. Firstly, stochastic signaling based on imperfect CSI information at the transmitter is considered to observe the effects of imperfect channel state information. It is shown that an optimal stochastic signal involves randomization between at most two signal levels for the formulated problem. Then by deriving upper and lower bounds on the average probability of error for stochastic signaling under CSI uncertainty, sufficient conditions are obtained to specify when the use of stochastic signaling can or cannot improve the performance of conventional signaling. Secondly, two different methods, namely robust stochastic signaling and stochastic signaling with averaging, are considered for designing stochastic signals under CSI uncertainty. In robust stochastic signaling, signals are designed for the worst-case channel coefficients, and the optimal signaling problem is formulated as a minimax problem [2, 16]. Then, sufficient conditions under which the generic minimax problem is equivalent to

designing signals for the smallest possible magnitude of the channel coefficient are obtained. In stochastic signaling with averaging approach, the transmitter assumes a probability distribution for the channel coefficient, and stochastic signals are designed by averaging over different channel coefficient values based on that probability distribution. It is shown that optimal signals obtained after this averaging method and those for the equivalent form of robust signaling method can be represented by at most two signal levels for each symbol. Solutions for the optimization problems can be calculated by using Particle Swarm Optimization (PSO) or convex relaxation approaches can be employed as in [14, 17, 18, 19]. Finally, simulations are performed and two numerical examples are presented to illustrate the theoretical results.

1.2 Organization of the Thesis

The organization of this thesis is as follows. In Chapter 2, optimal stochastic signaling is studied for any given detector for binary communications systems under second and fourth moment constraints on individual signals firstly and under an average power constraint secondly.

In Chapter 3, joint design of optimal signals and optimal detector for power constrained communication systems is investigated.

In Chapter 4, stochastic signaling is studied for power constrained scalar valued binary communications systems in the presence of uncertainties in channel state information (CSI).

Chapter 2

OPTIMAL STOCHASTIC SIGNALING FOR POWER CONSTRAINED COMMUNICATION SYSTEMS

In this chapter, optimal stochastic signaling is studied for the detection of scalar-valued binary signals in additive noise channels for a given decision rule. In the first section, optimization of the signals is performed under second and fourth moment constraints. For this scenario, sufficient conditions are obtained to specify when the use of stochastic signals instead of deterministic ones can or cannot improve the error performance of a given binary communications system. Also, statistical characterization of optimal signals is presented, and it is shown that an optimal stochastic signal can be represented by a randomization of at most three different signal levels. In addition, the power constraints achieved by optimal stochastic signals are specified under various conditions. Furthermore, two approaches for solving the optimal stochastic signaling problem are proposed; one

based on particle swarm optimization (PSO) and the other based on convex relaxation of the original optimization problem. Finally, simulations are performed to investigate the theoretical results, and extensions of the results to M-ary communications systems and to other criteria than the average probability of error are discussed.

In the second section, optimal signaling is studied in the presence of an average power constraint. Sufficient conditions are derived to determine the cases in which stochastic signaling can or cannot outperform the conventional signaling in this case as well. Also, statistical characterization of the optimal signals is provided and it is obtained that an optimal stochastic signal can be represented by a randomization of at most two different signal levels for each symbol for this scenario. In addition, via global optimization techniques, the solution of the generic optimal stochastic signaling problem is obtained, and theoretical results are investigated via numerical examples.

2.1 Stochastic Signaling Under Second and Fourth Moment Constraints

2.1.1 System Model and Motivation

Consider a scalar binary communications system, as in [6], [8] and [20], in which the received signal is expressed as

$$Y = S_i + N \ , \qquad i \in \{0, 1\} \ , \tag{2.1}$$

where S_0 and S_1 represent the transmitted signal values for symbol 0 and symbol 1, respectively, and N is the noise component that is independent of S_i. In addition, the prior probabilities of the symbols, which are represented by π_0 and π_1, are assumed to be known.

As stated in [6], the scalar channel model in (2.1) provides an abstraction for a continuous-time system that processes the received signal by a linear filter and samples it once per symbol interval. In addition, although the signal model in (2.1) is in the form of a simple additive noise channel, it also holds for flat-fading channels assuming perfect channel estimation. In that case, the signal model in (2.1) can be obtained after appropriate equalization [1].

It should be noted that the probability distribution of the noise component in (2.1) is not necessarily Gaussian. Due to interference, such as multiple-access interference, the noise component can have a significantly different probability distribution from the Gaussian distribution [5], [6], [21].

A generic decision rule is considered at the receiver to determine the symbol in (2.1). That is, for a given observation $Y = y$, the decision rule $\phi(y)$ is specified as

$$\phi(y) = \begin{cases} 0 \ , & y \in \Gamma_0 \\ 1 \ , & y \in \Gamma_1 \end{cases} \ , \tag{2.2}$$

where Γ_0 and Γ_1 are the decision regions for symbol 0 and symbol 1, respectively [2].

The aim is to design signals S_0 and S_1 in (2.1) in order to minimize the average probability of error for a given decision rule, which is expressed as

$$\mathrm{P}_{\mathrm{avg}} = \pi_0 \mathrm{P}_0(\Gamma_1) + \pi_1 \mathrm{P}_1(\Gamma_0) \ , \tag{2.3}$$

where $\mathrm{P}_i(\Gamma_j)$ is the probability of selecting symbol j when symbol i is transmitted. In practical systems, there are constraints on the average power and the peakedness of signals, which can be expressed as [22]

$$\mathrm{E}\{|S_i|^2\} \leq A \ , \qquad \mathrm{E}\{|S_i|^4\} \leq \kappa A^2 \ , \tag{2.4}$$

for $i = 0, 1$, where A is the average power limit and the second constraint imposes a limit on the peakedness of the signal depending on the $\kappa \in (1, \infty)$ parameter.[1] Therefore, the average probability of error in (2.3) needs to be minimized under the second and fourth moment constraints in (2.4).

The main motivation for the optimal stochastic signaling problem is to improve the error performance of the communications system by considering the signals at the transmitter as random variables and finding the optimal probability distributions for those signals [6]. Therefore, the generic problem can be formulated as obtaining the optimal probability distributions of the signals S_0 and S_1 for a given decision rule at the receiver under the average power and peakedness constraints in (2.4).

Since the optimal signal design is performed at the transmitter, the transmitter is assumed to have the knowledge of the statistics of the noise at the receiver and the channel state information. Although this assumption may not hold in some cases, there are certain scenarios in which it can be realized.[2] Consider, for example, the downlink of a multiple-access communications system, in which the received signal can be modeled as $Y = S^{(1)} + \sum_{k=2}^{K} \xi_k S^{(k)} + \eta$, where $S^{(k)}$ is the signal of the kth user, ξ_k is the correlation coefficient between user 1 and user k, and η is a zero-mean Gaussian noise component. For the desired signal component $S^{(1)}$, $N = \sum_{k=2}^{K} \xi_k S^{(k)} + \eta$ forms the total noise, which has Gaussian mixture distribution. When the receiver sends via feedback the variance of noise η and the signal-to-noise ratio (SNR) to the transmitter, the transmitter can fully characterize the PDF of the total noise N, as it knows the transmitted signal levels of all the users and the correlation coefficients.

[1] Note that for $\mathrm{E}\{|S_i|^2\} = A$, the second constraint becomes $\mathrm{E}\{|S_i|^4\}/(\mathrm{E}\{|S_i|^2\})^2 \leq \kappa$, which limits the kurtosis of the signal [22].

[2] As discussed in Section 2.1.5, the problem studied in this section can be considered for other systems than communications; hence, the practicality of the assumption depends on the specific application domain.

In the conventional signal design, S_0 and S_1 are considered as deterministic signals, and they are set to $S_0 = -\sqrt{A}$ and $S_1 = \sqrt{A}$ [1], [2]. In that case, the average probability of error expression in (2.3) becomes

$$\mathrm{P}_{\mathrm{avg}}^{\mathrm{conv}} = \pi_0 \int_{\Gamma_1} p_N\left(y + \sqrt{A}\right) dy + \pi_1 \int_{\Gamma_0} p_N\left(y - \sqrt{A}\right) dy \ , \tag{2.5}$$

where $p_N(\cdot)$ is the PDF of the noise in (2.1). As investigated in Section 2.1.2.1, the conventional signal design is optimal for certain classes of noise PDFs and decision rules. However, in some cases, use of stochastic signals instead of deterministic ones can improve the system performance. In the following section, conditions for optimality and suboptimality of the conventional signal design are derived, and properties of optimal signals are investigated.

2.1.2 Optimal Stochastic Signaling

Instead of employing constant levels for S_0 and S_1 as in the conventional case, consider a more generic scenario in which the signal components can be stochastic. The aim is to obtain the optimal PDFs for S_0 and S_1 in (2.1) that minimize the average probability of error under the constraints in (2.4).

Let $p_{S_0}(\cdot)$ and $p_{S_1}(\cdot)$ represent the PDFs for S_0 and S_1, respectively. Then, the average probability of error for the decision rule in (2.2) can be expressed from (2.3) as

$$\mathrm{P}_{\mathrm{avg}}^{\mathrm{stoc}} = \pi_0 \int_{-\infty}^{\infty} p_{S_0}(t) \int_{\Gamma_1} p_N(y-t)\, dy\, dt + \pi_1 \int_{-\infty}^{\infty} p_{S_1}(t) \int_{\Gamma_0} p_N(y-t)\, dy\, dt \ . \tag{2.6}$$

Therefore, the optimal stochastic signal design problem can be stated as

$$\begin{aligned} &\min_{p_{S_0}, p_{S_1}} \ \mathrm{P}_{\mathrm{avg}}^{\mathrm{stoc}} \\ &\text{subject to} \ \ \mathrm{E}\{|S_i|^2\} \leq A \ , \quad \mathrm{E}\{|S_i|^4\} \leq \kappa A^2 \ , \quad i = 0, 1 \ . \end{aligned} \tag{2.7}$$

Note that there are also implicit constraints in the optimization problem in (2.7), since $p_{S_i}(t)$ represents a PDF. Namely, $p_{S_i}(t) \geq 0 \ \forall t$ and $\int_{-\infty}^{\infty} p_{S_i}(t)dt = 1$ should also be satisfied by the optimal solution.

Since the aim is to obtain optimal stochastic signals for a given receiver, the decision rule in (2.2) is fixed (i.e., predefined Γ_0 and Γ_1). For a given decision rule (detector) and a noise PDF, changing p_{S_0} has no effect on the second term in (2.6) and the constraints for S_1 in (2.7). Similarly, changing p_{S_1} has no effect on the first term in (2.6) and the constraints for S_0 in (2.7). Therefore, the problem of minimizing the expression in (2.6) over p_{S_0} and p_{S_1} under the constraints for S_0 and S_1 in (2.7) is equivalent to minimizing the first term in (2.6) over p_{S_0} under the constraints for S_0 in (2.7) and minimizing the second term in (2.6) over p_{S_1} under the constraints for S_1 in (2.7). Therefore, the signal design problems for S_0 and S_1 can be separated and expressed as two decoupled optimization problems. For example, the optimal signal for symbol 1 can be obtained from the solution of the following optimization problem:

$$\begin{aligned} &\min_{p_{S_1}} \int_{-\infty}^{\infty} p_{S_1}(t) \int_{\Gamma_0} p_N(y-t)\, dy\, dt \\ &\text{subject to } \ \mathrm{E}\{|S_1|^2\} \leq A \ , \quad \mathrm{E}\{|S_1|^4\} \leq \kappa A^2 \ . \end{aligned} \tag{2.8}$$

A similar problem can be formulated for S_0 as well. Since the signals can be designed separately, the remainder of the discussion focuses on the design of optimal S_1 according to (2.8).

The objective function in (2.8) can be expressed as the expectation of

$$G(S_1) \triangleq \int_{\Gamma_0} p_N(y - S_1)\, dy \tag{2.9}$$

over the PDF of S_1. Then, the optimization problem in (2.8) becomes

$$\begin{aligned} &\min_{p_{S_1}} \mathrm{E}\{G(S_1)\} \\ &\text{subject to } \ \mathrm{E}\{|S_1|^2\} \leq A \ , \quad \mathrm{E}\{|S_1|^4\} \leq \kappa A^2 \ . \end{aligned} \tag{2.10}$$

It is noted that (2.10) provides a generic formulation that is valid for any noise PDF and detector structure. In the following sections, the signal subscripts are dropped for notational simplicity. Note that $G(x)$ in (2.9) represents the probability of deciding symbol 0 instead of symbol 1 when signal S_1 takes a constant value of x; that is, $S_1 = x$.

2.1.2.1 On the Optimality of the Conventional Signaling

Under certain circumstances, using the conventional signaling approach, i.e., setting $S = \sqrt{A}$ (or, $p_S(x) = \delta(x - \sqrt{A})$), solves the optimization problem in (2.10). For example, if $G(x)$ achieves its minimum at $x = \sqrt{A}$; that is, $\arg\min_x G(x) = \sqrt{A}$, then $p_S(x) = \delta(x - \sqrt{A})$ becomes the optimal solution since it yields the minimum value for $\mathrm{E}\{G(S_1)\}$ and also satisfies the constraints. However, this case is not very common as $G(x)$, which is the probability of deciding symbol 0 instead of symbol 1 when $S = x$, is usually a decreasing function of x; that is, when a larger signal value x is used, smaller error probability can be obtained. Therefore, the following more generic condition is derived for the optimality of the conventional algorithm.

Proposition 2.1: *If $G(x)$ is a strictly convex and monotone decreasing function, then $p_S(x) = \delta(x - \sqrt{A})$ solves the optimization problem in* (2.10).

Proof: The proof is obtained via contradiction. First, it is assumed that there exists a PDF $p_{S2}(x)$ for signal S that makes the conventional solution suboptimal; that is, $\mathrm{E}\{G(S)\} < G(\sqrt{A})$ under the constraints in (2.10).

Since $G(x)$ is a strictly convex function, Jensen's inequality implies that $\mathrm{E}\{G(S)\} > G(\mathrm{E}\{S\})$. Therefore, as $G(x)$ is a monotone decreasing function, $\mathrm{E}\{S\} > \sqrt{A}$ must be satisfied in order for $\mathrm{E}\{G(S)\} < G(\sqrt{A})$ to hold true.

On the other hand, Jensen's inequality also states that $\mathrm{E}\{S\} > \sqrt{A}$ implies $\mathrm{E}\{S^2\} > (\mathrm{E}\{S\})^2 > A$; that is, the constraint on the average power is violated (see (2.10)). Therefore, it is proven that no PDF can provide $\mathrm{E}\{G(S)\} < G(\sqrt{A})$ and satisfy the constraints under the assumptions in the proposition. □

As an example application of Proposition 2.1, consider a zero-mean Gaussian noise N in (2.1) with $p_N(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right)$, and a decision rule of the form $\Gamma_0 = (-\infty, 0]$ and $\Gamma_1 = [0, \infty)$; i.e., the sign detector. Then, $G(x)$ in (2.9) can be obtained as

$$G(x) = \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(y-x)^2}{2\sigma^2}\right) dy = Q\left(\frac{x}{\sigma}\right) , \tag{2.11}$$

where $Q(x) = (1/\sqrt{2\pi}) \int_x^\infty \exp(-t^2/2)\, dt$ defines the Q-function. It is observed that $G(x)$ in (2.11) is a monotone decreasing and strictly convex function for $x > 0$.[3] Therefore, the optimal signal is specified by $p_S(x) = \delta(x - \sqrt{A})$ from Proposition 2.1. Similarly, the optimal signal for symbol 0 can be obtained as $p_S(x) = \delta(x + \sqrt{A})$. Hence, the conventional signaling is optimal in this scenario.

2.1.2.2 Sufficient Conditions for Improvability

In this section, the aim is to determine when it is possible to improve the performance of the conventional signaling approach via stochastic signaling. A simple observation of (2.10) reveals that if the minimum of $G(x) = \int_{\Gamma_0} p_N(y-x)dy$ is achieved at $x_{\min}$ with $x_{\min}^2 < A$, then $p_S(x) = \delta(x - x_{\min})$ becomes a better solution than the conventional one. In other words, if the noise PDF is such that the probability of selecting symbol 0 instead of symbol 1 is minimized for a signal value of $S_1 = x_{\min}$ with $x_{\min}^2 < A$, then the conventional solution can be improved. Another sufficient condition for the conventional algorithm to be suboptimal is to have a positive first-order derivative of $G(x)$ at $x = \sqrt{A}$, which can

[3] It is sufficient to consider the positive signal values only since $G(x)$ is monotone decreasing and the constraints x^2 and x^4 are even functions. In other words, no negative signal value can be optimal since its absolute value has the same constraint value but smaller $G(x)$.

also be expressed from (2.9) as $-\int_{\Gamma_0} p_N^{'}(y-\sqrt{A}\,)\,dy > 0$, where $p_N^{'}(\cdot)$ denotes the derivative of $p_N(\cdot)$. In this case, $p_{S2}(x) = \delta(x-\sqrt{A}+\epsilon)$ yields a smaller average probability of error than the conventional solution for infinitesimally small $\epsilon > 0$ values.

Although both of the conditions above are sufficient for improvability of the conventional algorithm, they are rarely met in practice since $G(x)$ is commonly a decreasing function of x as discussed before. Therefore, in the following, a sufficient condition is derived for more generic and practical conditions.

Proposition 2.2: *Assume that $G(x)$ is twice continuously differentiable around $x = \sqrt{A}$. Then, if $\int_{\Gamma_0} \big(p_N^{''}(y-\sqrt{A}\,) + p_N^{'}(y-\sqrt{A}\,)/\sqrt{A}\,\big)dy < 0$ is satisfied, $p_S(x) = \delta(x-\sqrt{A})$ is not an optimal solution to* (2.10).

Proof: It is first observed from (2.9) that the condition in the proposition is equivalent to $G^{''}(\sqrt{A}) < G^{'}(\sqrt{A})/\sqrt{A}$. Therefore, in order to prove the suboptimality of the conventional solution $p_S(x) = \delta(x-\sqrt{A})$, it is shown that when $G^{''}(\sqrt{A}) < G^{'}(\sqrt{A})/\sqrt{A}$, there exists $\lambda \in (0,1)$, $\epsilon > 0$ and $\Delta > 0$ such that $p_{S2}(x) = \lambda\,\delta(x-\sqrt{A}+\epsilon) + (1-\lambda)\,\delta(x-\sqrt{A}-\Delta)$ has a lower error probability than $p_S(x)$ while satisfying all the constraints in (2.10). More specifically, the existence of $\lambda \in (0,1)$, $\epsilon > 0$ and $\Delta > 0$ that satisfy

$$\lambda\, G(\sqrt{A}-\epsilon) + (1-\lambda)\, G(\sqrt{A}+\Delta) < G(\sqrt{A}) \tag{2.12}$$

$$\lambda(\sqrt{A}-\epsilon)^2 + (1-\lambda)(\sqrt{A}+\Delta)^2 = A \tag{2.13}$$

$$\lambda(\sqrt{A}-\epsilon)^4 + (1-\lambda)(\sqrt{A}+\Delta)^4 \leq \kappa A^2 \tag{2.14}$$

is sufficient to prove the suboptimality of the conventional signal design.

From (2.13), the following equation is obtained.

$$\lambda\,\epsilon^2 + (1-\lambda)\Delta^2 = -2\sqrt{A}\;[(1-\lambda)\Delta - \lambda\,\epsilon]\;. \tag{2.15}$$

If infinitesimally small ϵ and Δ values are selected, (2.12) can be approximated as

$$\lambda \left[G(\sqrt{A}) - \epsilon\, G^{'}(\sqrt{A}) + \frac{\epsilon^2}{2}\, G^{''}(\sqrt{A}) \right] + (1-\lambda) \left[G(\sqrt{A}) + \Delta\, G^{'}(\sqrt{A}) + \frac{\Delta^2}{2}\, G^{''}(\sqrt{A}) \right]$$
$$< G(\sqrt{A}) + G^{'}(\sqrt{A})\, [(1-\lambda)\Delta - \lambda\, \epsilon] + \frac{G^{''}(\sqrt{A})}{2} \left[\lambda\, \epsilon^2 + (1-\lambda)\Delta^2 \right] < 0 \tag{2.16}$$

When the condition in (2.15) is employed, (2.16) becomes

$$[(1-\lambda)\Delta - \lambda\, \epsilon] \left(G^{'}(\sqrt{A}) - \sqrt{A}\, G^{''}(\sqrt{A}) \right) < 0\ . \tag{2.17}$$

Since $(1-\lambda)\Delta - \lambda\, \epsilon$ is always negative as can be noted from (2.15), the $G^{'}(\sqrt{A}) - \sqrt{A}\, G^{''}(\sqrt{A})$ term in (2.17) must be positive to satisfy the condition. In other words, when $G^{''}(\sqrt{A}) < G^{'}(\sqrt{A})/\sqrt{A}$, $p_{S_2}(x)$ can have a smaller error value than that of the conventional algorithm for infinitesimally small ϵ and Δ values that satisfy (2.15). To complete the proof, the condition in (2.14) needs to be verified for the specified ϵ and Δ values. From (2.15), (2.14) can be expressed, after some manipulation, as

$$A^2 + 16A\sqrt{A}\, [(1-\lambda)\Delta - \lambda\, \epsilon] - 4\sqrt{A}\, \left[\lambda\, \epsilon^3 - (1-\lambda)\Delta^3 \right]$$
$$+\ \left[\lambda\, \epsilon^4 - (1-\lambda)\Delta^4 \right] \leq \kappa A^2\ . \tag{2.18}$$

Since $(1-\lambda)\Delta - \lambda\, \epsilon$ is negative, the inequality can be satisfied for infinitesimally small ϵ and Δ, for which the third and the fourth terms on the left-hand-side become negligible compared to the first two. □

The condition in Proposition 2.2 can be expressed more explicitly in practice. For example, if Γ_0 is the form of an interval, say $[\tau_1, \tau_2]$, then the condition in the proposition becomes $p_N^{'}(\tau_2 - \sqrt{A}\,) - p_N^{'}(\tau_1 - \sqrt{A}\,) + \big(p_N(\tau_2 - \sqrt{A}\,) - p_N(\tau_1 - \sqrt{A}\,)\big)/\sqrt{A} < 0$. This inequality can be generalized in a straightforward manner when Γ_0 is the union of multiple intervals.

Since the condition in Proposition 2.2 is equivalent to $G^{''}(\sqrt{A}) < G^{'}(\sqrt{A})/\sqrt{A}$ (see (2.9)), the intuition behind the proposition can be explained as

follows. As the optimization problem in (2.10) aims to minimize $\mathrm{E}\{G(S)\}$ while keeping $\mathrm{E}\{S^2\}$ and $\mathrm{E}\{S^4\}$ below thresholds A and κA^2, respectively, a better solution than $p_S(x) = \delta(x - \sqrt{A})$ can be obtained with multiple mass points if $G(x)$ is decreasing at an increasing rate (i.e., with a negative second derivative) such that an increase from $x = \sqrt{A}$ causes a fast decrease in $G(x)$ but relatively slow increase in x^2 and x^4, and a decrease from $x = \sqrt{A}$ causes a fast decrease in x^2 and x^4 but relatively slow increase in $G(x)$. In that case, it becomes possible to use a PDF with multiple mass points and to obtain a smaller $\mathrm{E}\{G(S)\}$ while satisfying $\mathrm{E}\{S^2\} \leq A$ and $\mathrm{E}\{S^4\} \leq \kappa A^2$.

Proposition 2.2 provides a simple sufficient condition to determine if there is any possibility for performance improvement over the conventional signal design. For a given noise PDF and a decision rule, the condition in Proposition 2.2 can be evaluated in a straightforward manner. In order to provide an illustrative example, consider the noise PDF

$$p_N(y) = \begin{cases} y^2 \ , & |y| \leq 1.1447 \\ 0 \ , & |y| > 1.1447 \end{cases} , \tag{2.19}$$

and a sign detector at the receiver; that is, $\Gamma_0 = (-\infty, 0]$. Then, the condition in Proposition 2.2 can be evaluated as

$$p_N^{'}(-\sqrt{A}\,) + p_N(-\sqrt{A}\,)/\sqrt{A} < 0 \ . \tag{2.20}$$

Assuming that the average power is constrained to $A = 0.64$, the inequality in (2.20) becomes $2(-0.8) + (-0.8)^2/0.8 < 0$. Hence, Proposition 2.2 implies that the conventional solution is not optimal for this problem. For example, $p_S(x) = 0.391\,\delta(x - 0.988) + 0.333\,\delta(x - 0.00652) + 0.276\,\delta(x - 0.9676)$ yields an average error probability of 0.2909 compared to 0.3293 corresponding to the conventional solution $p_S(x) = \delta(x - 0.8)$, as studied in Section 2.1.3.

Although the noise PDF in (2.19) is not common in practice, improvements over the conventional algorithm are possible and Proposition 2.2 can be applied

also for certain types of Gaussian mixture noise (see Section 2.1.3), which is observed more frequently in practical scenarios [21]-[24]. For example, in multiuser wireless communications, the desired signal is corrupted by interfering signals from other users as well as zero-mean Gaussian noise, which altogether result in Gaussian mixture noise [21].

2.1.2.3 Statistical Characteristics of Optimal Signals

In this section, PDFs of optimal signals are characterized and it is shown that an optimal signal can be represented by a randomization of at most three different signal levels. In addition, it is proven that the optimal signal achieves at least one of the second and fourth moment constraints in (2.10) for most practical cases.

In the following proposition, it is stated that, in most practical scenarios, an optimal stochastic signal can be represented by a discrete random variable with no more than three mass points.

Proposition 2.3: *Assume that the possible signal values are specified by* $|S| \leq \gamma$ *for a finite* $\gamma > 0$*, and* $G(\cdot)$ *in* (2.9) *is continuous. Then, an optimal solution to* (2.10) *can be expressed in the form of* $p_S(x) = \sum_{i=1}^{3} \lambda_i \, \delta(x - x_i)$*, where* $\sum_{i=1}^{3} \lambda_i = 1$ *and* $\lambda_i \geq 0$ *for* $i = 1, 2, 3$ *.*

Proof: In order to prove Proposition 2.3, we take an approach similar to those in [12] and [25]. First, the following set is defined:

$$U = \Big\{ (u_1, u_2, u_3) \ : \ u_1 = G(x), \ u_2 = x^2, \ u_3 = x^4, \ \text{for } |x| \leq \gamma \Big\} \ . \tag{2.21}$$

Since $G(x)$ is continuous, the mapping from $[-\gamma, \gamma]$ to $\mathbb{R}^3$ defined by $F(x) = (G(x), x^2, x^4)$ is continuous. Since the continuous image of a compact set is compact, U is a compact set [26].

Let V represent the convex hull of U. Since U is compact, the convex hull V of U is closed [26]. Also, the dimension of V should be smaller than or equal to 3, since $V \subseteq \mathbb{R}^3$. In addition, let W be the set of all possible conditional error probability $\mathrm{P}_1(\Gamma_0)$, second moment, and fourth moment triples; i.e.,

$$W = \Big\{(w_1, w_2, w_3) \,:\, w_1 = \int_{-\infty}^{\infty} p_S(x)G(x)dx,\ w_2 = \int_{-\infty}^{\infty} p_S(x)x^2dx,$$
$$w_3 = \int_{-\infty}^{\infty} p_S(x)x^4dx,\ \forall\, p_S(x),\ |x| \leq \gamma\Big\}. \tag{2.22}$$

where $p_S(x)$ is the signal PDF.

Similar to [25], $V \subseteq W$ can be proven as follows. Since V is the convex hull of U, each element of V can be expressed as $\mathbf{v} = \sum_{i=1}^{L} \lambda_i \left(G(x_i), x_i^2, x_i^4\right)$, where $\sum_{i=1}^{L} \lambda_i = 1$, and $\lambda_i \geq 0\ \forall i$. Considering set W, it has an element that is equal to $\mathbf{v}$ for $p_S(x) = \sum_{i=1}^{L} \lambda_i\, \delta(x - x_i)$. Hence, each element of V also exists in W. On the other hand, since for any vector random variable $\mathbf{\Theta}$ that takes values in set Ω, its expected value $\mathrm{E}\{\mathbf{\Theta}\}$ is in the convex hull of Ω [12], it is concluded from (2.21) and (2.22) that W is in the convex hull V of U; that is, $V \supseteq W$ [19].

Since $W \supseteq V$ and $V \supseteq W$, it is concluded that $W = V$. Therefore, Carathéodory's theorem [27], [28] implies that any point in V (hence, in W) can be expressed as the convex combination of at most 4 points in U. Since an optimal PDF should minimize the average probability of error, it corresponds to the boundary of V. Since V is a closed set as discussed at the beginning of the proof, it contains its own boundary. Since any point at the boundary of V can be expressed as the convex combination of at most 3 elements in U [27], an optimal PDF can be represented by a discrete random variable with 3 mass points $\square$.

The assumption in the proposition, which states that the possible signal values belong to set $[-\gamma, \gamma]$, is realistic for practical communications systems since arbitrarily large positive and negative signal values cannot be generated at the transmitter. In addition, for most practical scenarios, $G(\cdot)$ in (2.9) is continuous

since the noise at the receiver, which is commonly the sum of zero-mean Gaussian thermal noise and interference terms that are independent from the thermal noise, has a continuous PDF.

The result in Proposition 2.3 can be extended to the problems with more constraints. Let $\mathrm{E}\{G(S)\}$ be the objective function to minimize over possible PDFs $p_S(x)$, subject to $\mathrm{E}\{H_i(S)\} \leq A_i$ for $i = 1, \ldots, N_c$. Then, under the conditions in the proposition, this proof implies that there exists an optimal PDF with at most $N_c + 1$ mass points.[4]

The significance of Proposition 2.3 lies in the fact that it reduces the optimization problem in (2.10) from the space of all PDFs that satisfy the second and fourth moment constraints to the space of discrete PDFs with at most 3 mass points that satisfy the second and fourth moment constraints. In other words, instead of optimization over functions, an optimization over a vector of 6 elements (namely, 3 mass point locations and their weights) can be considered for the optimal signaling problem as a result of Proposition 2.3. In addition, this result facilitates a convex relaxation of the optimization problem in (2.10) for any noise PDF and decision rule as studied in Section 2.1.2.4.

Next, the second and the fourth moments of the optimal signals are investigated. Let $x_{\min}$ represent the signal level that yields the minimum value of $G(x)$ in (2.9); that is, $x_{\min} = \arg\min_x G(x)$. If $x_{\min} < \sqrt{A}$, the optimal signal has the constant value of $x_{\min}$ and the second and fourth moments are given by $x_{\min}^2 < A$ and $x_{\min}^4 < \kappa A^2$, respectively. However, it is more common to have $x_{\min} > \sqrt{A}$ since larger signal values are expected to reduce $G(x)$ as discussed before. In that case, the following proposition states that at least one of the constraints in (2.10) is satisfied.

[4]It is assumed that $H_1(x), \ldots, H_{N_c}(x)$ are bounded functions for the possible values of the signal.

Proposition 2.4: *Let* $x_{\min} = \arg\min_x G(x)$ *be the unique minimum of* $G(x)$ *.*

a) *If* $A^2 < x_{\min}^4 < \kappa A^2$*, then the optimal signal satisfies* $\mathrm{E}\{S^2\} = A$*.*

b) *If* $x_{\min}^4 > \kappa A^2$*, then the optimal signal satisfies at least one of* $\mathrm{E}\{S^2\} = A$ *and* $\mathrm{E}\{S^4\} = \kappa A^2$*.*

Proof: **a)** Let $A^2 < x_{\min}^4 < \kappa A^2$ and $p_{S1}(x)$ represent an optimal signal PDF with $w_1 \triangleq \mathrm{E}\{G(S)\}$, $w_2 \triangleq \mathrm{E}\{S^2\}$ and $w_3 \triangleq \mathrm{E}\{S^4\}$, where $w_2 < A$ and $w_3 \leq \kappa A^2$. In the following, it is shown that such a signal cannot be optimal (hence, a contradiction), and an optimal signal needs to satisfy $\mathrm{E}\{S^2\} = A$. To that aim, define another signal PDF as follows:

$$p_{S2}(x) = \frac{A - w_2}{x_{\min}^2 - w_2}\,\delta(x - x_{\min}) + \frac{x_{\min}^2 - A}{x_{\min}^2 - w_2}\,p_{S1}(x)\ . \tag{2.23}$$

It can be shown for $p_{S2}(x)$ that

$$\mathrm{E}\{G(S)\} = \frac{A - w_2}{x_{\min}^2 - w_2}\,G(x_{\min}) + \frac{x_{\min}^2 - A}{x_{\min}^2 - w_2}\,w_1 < w_1\ , \tag{2.24}$$

$$\mathrm{E}\{S^2\} = \frac{A - w_2}{x_{\min}^2 - w_2}\,x_{\min}^2 + \frac{x_{\min}^2 - A}{x_{\min}^2 - w_2}\,w_2 = A\ , \tag{2.25}$$

$$\mathrm{E}\{S^4\} = \frac{A - w_2}{x_{\min}^2 - w_2}\,x_{\min}^4 + \frac{x_{\min}^2 - A}{x_{\min}^2 - w_2}\,w_3 < \kappa A^2\ . \tag{2.26}$$

The inequality in (2.24) is obtained by observing that $G(x_{\min})$ is the unique minimum value of $G(x)$ and that no signals can achieve $\mathrm{E}\{G(S)\} = G(x_{\min})$ since $x_{\min} > \sqrt{A}$. The inequality in (2.26) is achieved since $x_{\min}^4 < \kappa A^2$ and $w_3 \leq \kappa A^2$. From (2.24)-(2.26), it is concluded that $p_{S2}(x)$ defines a better signal than $p_{S1}(x)$ does. In other words, the optimal signal cannot have a smaller average power than A; that is, $\mathrm{E}\{S^2\} = A$ must be satisfied by the optimal signal.

b) Now assume $x_{\min}^4 > \kappa A^2$ and $p_{S1}(x)$ represents an optimal signal PDF with $w_1 \triangleq \mathrm{E}\{G(S)\}$, $w_2 \triangleq \mathrm{E}\{S^2\}$ and $w_3 \triangleq \mathrm{E}\{S^4\}$, where $w_2 < A$ and $w_3 < \kappa A^2$. In the following, it is proven that $w_2 < A$ and $w_3 < \kappa A^2$ cannot be satisfied at the same time for an optimal signal.

Consider $p_{S_2}(x)$ in (2.23) and $p_{S_3}(x)$ below:

$$p_{S_3}(x) = \frac{\kappa A^2 - w_3}{x_{\min}^4 - w_3}\,\delta(x - x_{\min}) + \frac{x_{\min}^4 - \kappa A^2}{x_{\min}^4 - w_3}\,p_{S_1}(x)\ . \tag{2.27}$$

For both $p_{S_2}(x)$ and $p_{S_3}(x)$, it can be shown that $\mathrm{E}\{G(S)\} < w_1$ since $G(x_{\min}) < w_1$. For $p_{S_2}(x)$, the second and fourth moment constraints can be expressed as

$$\mathrm{E}\{S^2\} = \frac{A - w_2}{x_{\min}^2 - w_2}\,x_{\min}^2 + \frac{x_{\min}^2 - A}{x_{\min}^2 - w_2}\,w_2 = A\ , \tag{2.28}$$

$$\mathrm{E}\{S^4\} = \frac{A - w_2}{x_{\min}^2 - w_2}\,x_{\min}^4 + \frac{x_{\min}^2 - A}{x_{\min}^2 - w_2}\,w_3 \triangleq \beta_1\ . \tag{2.29}$$

On the other hand, for $p_{S_3}(x)$, the constraints are given by

$$\mathrm{E}\{S^2\} = \frac{\kappa A^2 - w_3}{x_{\min}^4 - w_3}\,x_{\min}^2 + \frac{x_{\min}^4 - \kappa A^2}{x_{\min}^4 - w_3}\,w_2 \triangleq \beta_2\ , \tag{2.30}$$

$$\mathrm{E}\{S^4\} = \frac{\kappa A^2 - w_3}{x_{\min}^4 - w_3}\,x_{\min}^4 + \frac{x_{\min}^4 - \kappa A^2}{x_{\min}^4 - w_3}\,w_3 = \kappa A^2\ . \tag{2.31}$$

Now it is claimed that at least one of the conditions $\beta_1 \le \kappa A^2$ or $\beta_2 \le A$ must be true. In other words, it is not possible to have $\beta_1 > \kappa A^2$ and $\beta_2 > A$ at the same time. To prove this, the condition for $\beta_1 > \kappa A^2$ is considered first. Since $x_{\min}^4 > \kappa A^2$ and $w_3 < \kappa A^2$, $\beta_1 > \kappa A^2$ can be expressed from (2.29) as

$$\frac{x_{\min}^4 - \kappa A^2}{\kappa A^2 - w_3} > \frac{x_{\min}^2 - A}{A - w_2}\ . \tag{2.32}$$

Next, the $\beta_2 > A$ condition is considered. Since $x_{\min}^2 > A$ and $w_2 < A$, that condition can be expressed, from (2.30), as

$$\frac{x_{\min}^4 - \kappa A^2}{\kappa A^2 - w_3} < \frac{x_{\min}^2 - A}{A - w_2}\ . \tag{2.33}$$

Since (2.32) and (2.33) cannot be true at the same time, at least one of the conditions $\beta_1 \le \kappa A^2$ or $\beta_2 \le A$ is true. This implies that at least one of $p_{S_2}(x)$ or $p_{S_3}(x)$ provides a signal that has a smaller average probability of error than that for $p_{S_1}(x)$. In addition, such a signal satisfies at least one of the constraints with equality as can be observed from (2.28) and (2.31). Therefore, an optimal signal cannot be in the form of $p_{S_1}(x)$, which satisfies both inequalities as $\mathrm{E}\{S^2\} < A$ and $\mathrm{E}\{S^4\} < \kappa A^2$. $\square$

An important implication of Proposition 2.4 is that when $x_{\min} > \sqrt{A}$, any solution that results in second and fourth moments that are smaller than A and κA^2, respectively, cannot be optimal. In other words, it is possible to improve that solution by increasing the second and/or the fourth moment of the signal until at least one of the constraints become active.

After characterizing the structure and the properties of optimal signals, two approaches are proposed in the next section to obtain optimal and close-to-optimal signal PDFs.

2.1.2.4 Calculation of the Optimal Signal

In order to obtain the PDF of an optimal signal, the constrained optimization problem in (2.10) should be solved. In this section, two approaches are studied in order to obtain optimal and close-to-optimal solutions to that optimization problem.

2.1.2.4.1 Global Optimization Approach Since Proposition 2.3 states that the optimal signaling problem in (2.10) can be solved over PDFs in the form of $p_S(x) = \sum_{j=1}^{3} \lambda_j \, \delta(x - x_j)$, (2.10) can be expressed as

$$\begin{aligned} &\min_{\boldsymbol{\lambda},\mathbf{x}} \sum_{j=1}^{3} \lambda_j \, G(x_j) \\ &\text{subject to} \sum_{j=1}^{3} \lambda_j \, x_j^2 \leq A \ , \quad \sum_{j=1}^{3} \lambda_j \, x_j^4 \leq \kappa A^2 \ , \\ &\qquad \sum_{j=1}^{3} \lambda_j = 1 \ , \quad \lambda_j \geq 0 \ \ \forall j \ , \end{aligned} \tag{2.34}$$

where $\mathbf{x} = [x_1 \ x_2 \ x_3]^T$ and $\boldsymbol{\lambda} = [\lambda_1 \ \lambda_2 \ \lambda_3]^T$.

Note that the optimization problem in (2.34) is a not convex problem in general due to both the objective function and the first two constraints. Therefore, global optimization techniques, such as PSO, differential evolution and genetic

algorithms [29] should be employed to obtain the optimal PDF. In this study, the PSO approach [13], [30]-[32] is used since it is based on simple iterations with low computational complexity and has been successfully applied to numerous problems in various fields [33]-[37].

In order to describe the PSO algorithm, consider the minimization of an objective function over parameter $\boldsymbol{\theta}$. In PSO, first a number of parameter values $\{\boldsymbol{\theta}_i\}_{i=1}^{M}$, called *particles*, are generated, where M is called the population size (i.e., the number of particles). Then, iterations are performed, where at each iteration new particles are generated as the summation of the previous particles and velocity vectors $\boldsymbol{\upsilon}_i$ according to the following equations [13]:

$$\boldsymbol{\upsilon}_i^{k+1} = \chi\left(\omega\boldsymbol{\upsilon}_i^k + c_1\rho_{i1}^k\left(\mathbf{p}_i^k - \boldsymbol{\theta}_i^k\right) + c_2\rho_{i2}^k\left(\mathbf{p}_g^k - \boldsymbol{\theta}_i^k\right)\right) \tag{2.35}$$

$$\boldsymbol{\theta}_i^{k+1} = \boldsymbol{\theta}_i^k + \boldsymbol{\upsilon}_i^{k+1} \tag{2.36}$$

for $i = 1, \ldots, M$, where k is the iteration index, χ is the constriction factor, ω is the inertia weight, which controls the effects of the previous history of velocities on the current velocity, c_1 and c_2 are the cognitive and social parameters, respectively, and ρ_{i1}^k and ρ_{i2}^k are independent uniformly distributed random variables on $[0, 1]$ [30]. In (2.35), $\mathbf{p}_i^k$ represents the position corresponding to the smallest objective function value until the kth iteration of the ith particle, and $\mathbf{p}_g^k$ denotes the position corresponding to the global minimum among all the particles until the kth iteration. After a number of iterations, the position with the lowest objective function value, $\mathbf{p}_g^k$, is selected as the optimizer of the optimization problem.

In order to extend PSO to constrained optimization problems, various approaches, such as penalty functions and keeping feasibility of particles, can be taken [31], [32]. In the penalty function approach, a particle that becomes infeasible is assigned a large value (considering a minimization problem), which forces migration of particles to the feasible region. In the constrained optimization approach that preserves the feasibility of the particles, no penalty is applied to any

particles; but for the positions $\mathbf{p}_i^k$ and $\mathbf{p}_g^k$ in (2.35) corresponding to the lowest objective function values, only the feasible particles are considered [32].

In order to employ PSO for the optimal stochastic signaling problem in (2.34), the optimization variable is defined as $\boldsymbol{\theta} \triangleq [x_1\ x_2\ x_3\ \lambda_1\ \lambda_2\ \lambda_3]^T$, and the iterations in (2.35) and (2.36) are used while using a penalty function approach to impose the constraints. The results are presented in Section 2.1.3.

2.1.2.4.2 Convex Optimization Approach In order to provide an alternative approximate solution with lower complexity, consider a scenario in which the PDF of the signal is modeled as

$$p_S(x) = \sum_{j=1}^{K} \tilde{\lambda}_j\, \delta(x - \tilde{x}_j)\ , \tag{2.37}$$

where $\tilde{x}_j$'s are the *known* mass points of the PDFs, and $\tilde{\lambda}_j$'s are the weights to be estimated. This scenario corresponds to the cases with a finite number of possible signal values. For example, in a digital communications system, if the transmitter can only send one of K pre-determined $\tilde{x}_j$ values for a specific symbol, then the problem becomes calculating the optimal probability assignments, $\tilde{\lambda}_j$'s, for the possible signal values for each symbol. Note that since the optimization is performed over PDFs as in (2.37), the optimal solution can include more than three mass points in general. In other words, the solution in this case is expected to approximate the optimal PDF, which includes at most three mass points, with a PDF with multiple mass points.

The solution to the optimal signal design problem in (2.10) over the set of signals with their PDFs as in (2.37) can be obtained from the solution of the

following convex optimization problem:[5]

$$\min_{\tilde{\boldsymbol{\lambda}}} \mathbf{g}^T \tilde{\boldsymbol{\lambda}} \tag{2.38}$$

$$\text{subject to} \quad \mathbf{B}\tilde{\boldsymbol{\lambda}} \preceq \mathbf{C} \ ,$$

$$\mathbf{1}^T \tilde{\boldsymbol{\lambda}} = 1 \ , \quad \tilde{\boldsymbol{\lambda}} \succeq \mathbf{0} \ ,$$

where $\mathbf{g} \triangleq [G(\tilde{x}_1) \cdots G(\tilde{x}_K)]^T$, with $G(x)$ as in (2.9),

$$\mathbf{B} \triangleq \begin{bmatrix} \tilde{x}_1^2 & \cdots & \tilde{x}_K^2 \\ \tilde{x}_1^4 & \cdots & \tilde{x}_K^4 \end{bmatrix} \ , \quad \mathbf{C} \triangleq \begin{bmatrix} A \\ \kappa A^2 \end{bmatrix} \ , \tag{2.39}$$

and $\mathbf{1}$ and $\mathbf{0}$ represent vectors of all ones and all zeros, respectively.

It is observed from (2.38) that the optimal weight assignments can be obtained as the solution of a convex optimization problem, specifically, a linearly constrained linear programming problem. Therefore, the solution can be obtained in polynomial time [14].

Note that if the set of possible signal values $\tilde{x}_j$'s include the deterministic signal value for the conventional algorithm, i.e., $\sqrt{A}$, then the performance of the convex algorithm in (2.38) can never be worse than that of the conventional one. In addition, as the number of possible signal values, K in (2.37), increases, the convex algorithm can approximate the exact optimal solution more closely.

2.1.3 Simulation Results

In this section, numerical examples are presented for a binary communications system with equal priors ($\pi_0 = \pi_1 = 0.5$) in order to investigate the theoretical results in the previous section. In the implementation of the PSO algorithm specified by (2.35) and (2.36), $M = 50$ particles are employed and 10000 iterations are performed. In addition, the parameters are set to $c_1 = c_2 = 2.05$ and

[5]For K-dimensional vectors $\mathbf{x}$ and $\mathbf{y}$, $\mathbf{x} \preceq \mathbf{y}$ means that the ith element of $\mathbf{x}$ is smaller than or equal to the ith element of $\mathbf{y}$ for $i = 1, \ldots, K$.

$\chi = 0.72984$, and the inertia weight ω is changed from 1.2 to 0.1 linearly with the iteration number [13]. Also, a penalty function approach is implemented to impose the constraints in (2.34); namely, the objective function is set to 1 whenever a particle becomes infeasible [33].

First, the noise in (2.1) is modeled by the PDF in (2.19), $A = 0.64$ and $\kappa = 1.5$ are employed for the constraints in (2.10), and the decision rule at the receiver is specified by $\Gamma_0 = (-\infty, 0]$ and $\Gamma_1 = [0, \infty)$ (that is, a sign detector). As stated after (2.20), the conventional signaling is suboptimal in this case based on Proposition 2.2. In order to calculate optimal signals via the PSO and the convex optimization algorithms in Section 2.1.2.4, the optimization problems in (2.34) and (2.38) are solved, respectively. For the convex algorithm, the mass points $\tilde{x}_j$ in (2.37) are selected uniformly over the interval $[0, 2]$ with a step size of Δ, and the results for $\Delta = 0.01$ and $\Delta = 0.1$ are considered. Fig. 2.1 illustrates the optimal probability distributions obtained from the PSO and the convex optimization algorithms.[6]

It is calculated that the conventional algorithm, which uses a deterministic signal value of 0.8, has an average error probability of 0.3293, whereas the PSO and the convex optimization algorithms with $\Delta = 0.01$ and $\Delta = 0.1$ have average error probabilities of 0.2909, 0.2911 and 0.2912, respectively. It is noted that the PSO algorithm achieves the lowest error probability with three mass points and the convex algorithms approximate the PSO solution with multiple mass points around those of the PSO solution. In addition, the calculations indicate that the optimal solutions achieve both the second and the fourth moment constraints in accordance with Proposition 2.4-b .

[6]For the probability distributions obtained from the convex optimization algorithms, the signal values that have zero probability are not marked in the figures to clarify the illustrations.

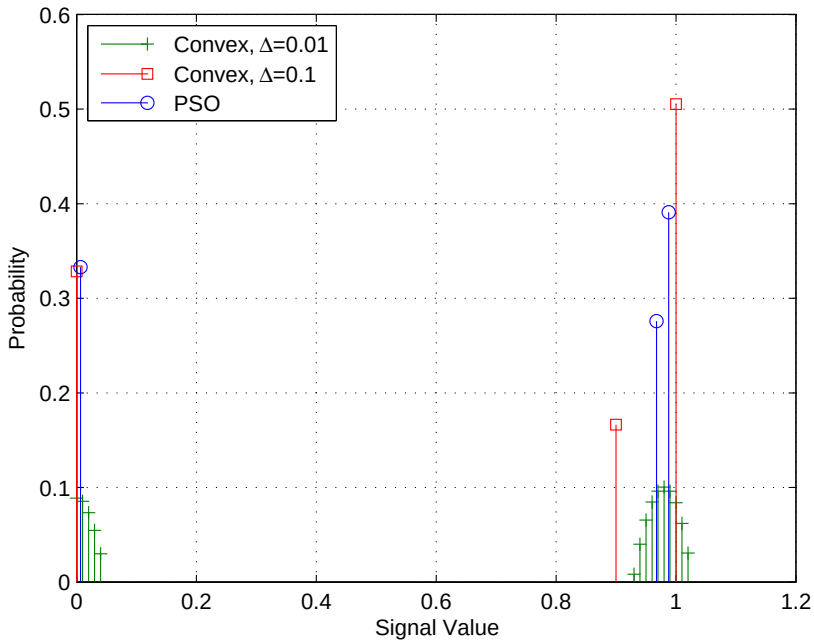

Figure 2.1: Probability mass functions (PMFs) of the PSO and the convex optimization algorithms for the noise PDF in (2.19).

Next, the optimal signaling problem is studied in the presence of Gaussian mixture. The Gaussian mixture noise can be used to model the effects of co-channel interference, impulsive noise and multiuser interference in communications systems [5], [7]. In the simulations, the Gaussian mixture noise is specified by $p_N(y) = \sum_{l=1}^{L} v_l\, \psi_l(y - y_l)$, where $\psi_l(y) = \mathrm{e}^{-y^2/(2\sigma_l^2)}/(\sqrt{2\pi}\,\sigma_l)$. In this case, $G(x)$ can be obtained from (2.9) as $G(x) = \sum_{l=1}^{L} v_l\, Q\left(\frac{x+y_l}{\sigma_l}\right)$. In all the scenarios, the variance parameter for each mass point of the Gaussian mixture is set to σ^2 (i.e., $\sigma_l^2 = \sigma^2\ \forall l$), and the average power constraint A is set to 1. Note that the average power of the noise can be calculated as $\mathrm{E}\{N^2\} = \sigma^2 + \sum_{l=1}^{L} v_l\, y_l^2$. First, we consider a symmetric Gaussian mixture noise which has its mass points at $\pm[0.3\ 0.455\ 1.011]$ with corresponding weights $[0.1\ 0.317\ 0.083]$ in order to illustrate the improvements that can be obtained via stochastic signaling. In Fig. 2.2, the average error probabilities of various algorithms are plotted against A/σ^2 when $\kappa = 1.1$ for both the sign detector and the ML detector. For the sign detector, the decision rule at the receiver is specified by $\Gamma_0 = (-\infty, 0]$ and

$\Gamma_1 = [0, \infty)$. In this case, it is observed from Fig. 2.2 that the conventional algorithm, which uses a constant signal value of 1, has a large error floor compared to the PSO and convex optimization algorithms at high A/σ^2. Also, the average probability of error of the conventional signaling increases as A/σ^2 increases after a certain value. This seemingly counterintuitive result is observed because the average probability of error is related to the area under the two shifted noise PDFs as in (2.5). Since the noise has a multi-modal PDF, that area is a non-monotonic function of A/σ^2 and can increase in some cases as A/σ^2 increases. It is also observed that the convex optimization algorithm performs very closely to the PSO algorithm for densely spaced possible signal values, i.e., for $\Delta = 0.01$. For the ML detector, the receiver compares $p_N(y - \sqrt{A})$ and $p_N(y + \sqrt{A})$, and decides symbol 0 if the latter is larger, and decides 1 otherwise. It is observed for small σ^2 values that the ML receiver performs significantly better than the other receivers that are based on the sign detector. However, stochastic signaling causes the sign detector to perform better than the conventional ML receiver, which uses deterministic signaling, for medium A/σ^2 values. For example, the PSO and convex optimization algorithms for $\Delta = 0.01$ have better performance than the ML receiver for A/σ^2 values from 20 dB to 40 dB. This is mainly due to the fact that the conventional ML detector uses deterministic signaling whereas the others employ stochastic signaling. However, when the stochastic signaling is applied to the ML detector as well, it achieves the lowest probabilities of error for all A/σ^2 values as observed in Fig. 2.2 (labeled as "ML (Stochastic)").

Another observation from Fig. 2.2 is that improvements over the conventional algorithm disappear as σ^2 increases (i.e., for small A/σ^2 values). This result can be explained from Propositions 2.1 and 2.2, based on the plots of $G(x)$ at various A/σ^2 values. For example, Fig. 2.3 illustrates the plots of $G(x)$ at A/σ^2 of 0, 20 and 40 dB for the sign detector. The function is decreasing and convex for 0 dB for the positive signal values, which are practically the domain of optimization since $G(x)$ is a decreasing function and the constraint functions x^2 and x^4 are even

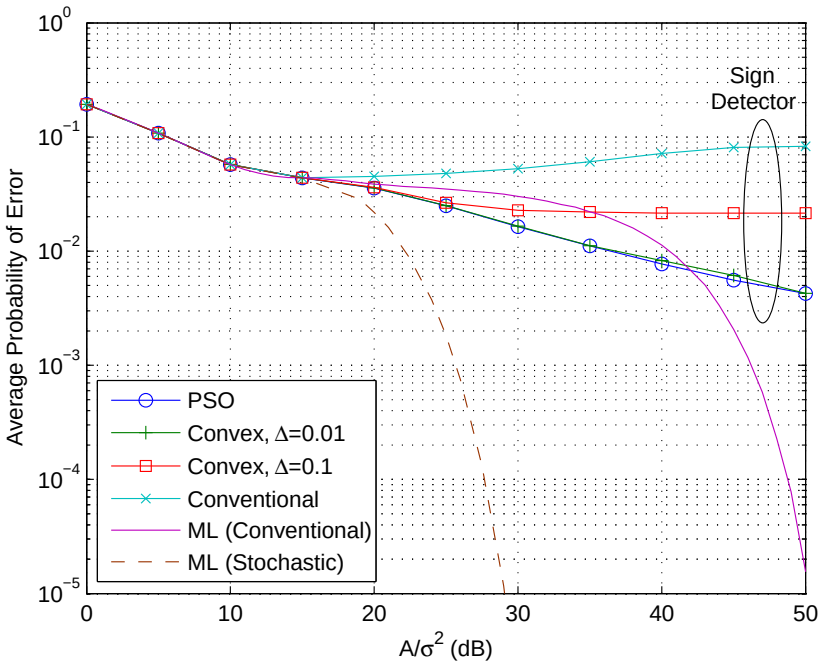

Figure 2.2: Error probability versus A/σ^2 for $\kappa = 1.1$. A symmetric Gaussian mixture noise, which has its mass points at $\pm[0.3\ 0.455\ 1.011]$ with corresponding weights $[0.1\ 0.317\ 0.083]$, is considered.

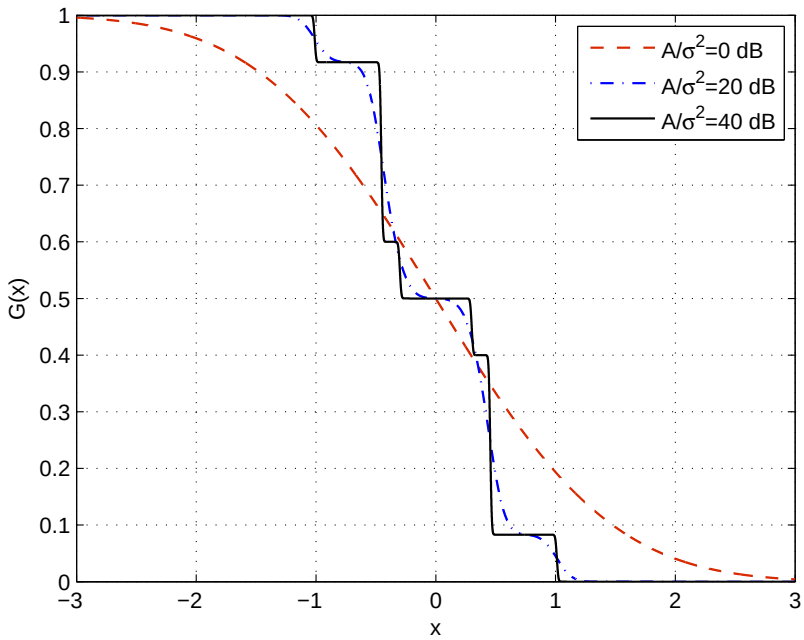

Figure 2.3: $G(x)$ in (2.9) for the sign detector in Fig. 2.2 at A/σ^2 values of 0, 20 and 40 dB.

functions.[7] Therefore, Proposition 2.1 implies that the conventional algorithm that uses a constant signal value of 1 is optimal in this case, as observed in Fig. 2.2. On the other hand, at 20 dB and 40 dB, the calculations show that the condition in Proposition 2.2 is satisfied; hence, the conventional algorithm cannot be optimal in that case, and improvements are observed in Fig. 2.2 at $A/\sigma^2 = 20$ dB and $A/\sigma^2 = 40$ dB. Another result obtained from the numerical studies for Fig. 2.2 is that all the solutions achieve at least one of the second moment or the fourth moment constraints with equality as a result of Proposition 2.4.

For the scenario in Fig. 2.2, the probability distributions of the optimal signals for the sign detector are shown in Fig. 2.4 and Fig. 2.5 for $A/\sigma^2 = 20$ dB and $A/\sigma^2 = 40$ dB, respectively, where both the PSO and the convex optimization algorithms are considered. In the first case, the convex optimization algorithm with $\Delta = 0.1$ approximates the probability mass function (PMF) obtained from the PSO algorithm with two mass points (with nonzero probabilities), whereas the convex optimization algorithm with $\Delta = 0.01$ results in 8 mass points. In the second case, the convex optimization algorithms with $\Delta = 0.1$ and $\Delta = 0.01$ result in PMFs with two and three mass points, respectively, as shown in Fig. 2.5. Since the convex optimization algorithm with $\Delta = 0.1$ does not provide a PMF that is very close to those of the other algorithms in this case, the resulting error probability becomes significantly higher for that algorithm, as observed from Fig. 2.2 at $A/\sigma^2 = 40$ dB.

Finally, a symmetric Gaussian mixture noise which has its mass points at $\pm[0.19\ 0.39\ 0.83\ 1.03]$ each with a weight of 1/8 is considered. Such a noise PDF can be considered to model the effects of co-channel interference [7], or a system that operates under the effect of multiuser interference [5]. For example, in the

[7] In other words, negative signal values are never selected for symbol 1 since selecting the absolute value of a negative signal value always gives a smaller average probability of error without changing the signal moments.

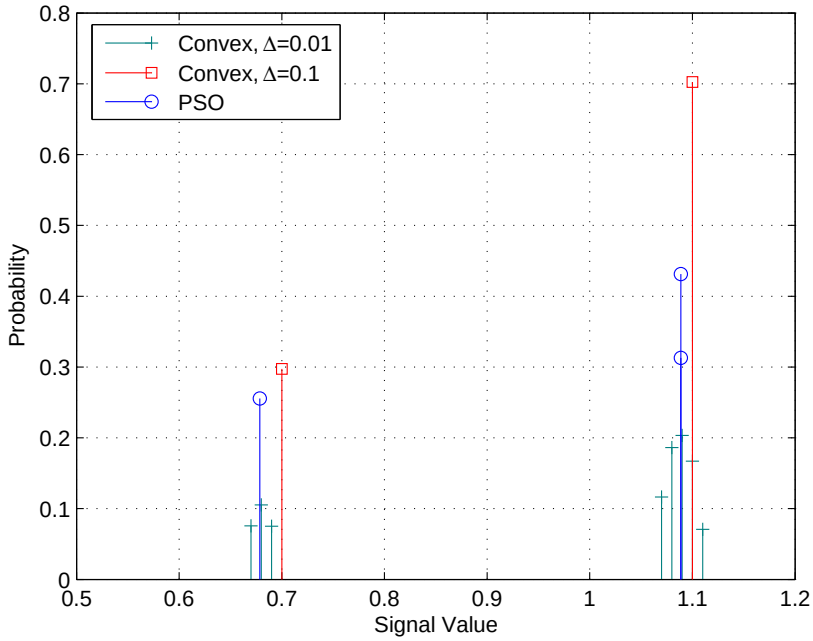

Figure 2.4: PMFs of the PSO and the convex optimization algorithms for the sign detector in Fig. 2.2 at $A/\sigma^2 = 20$ dB.

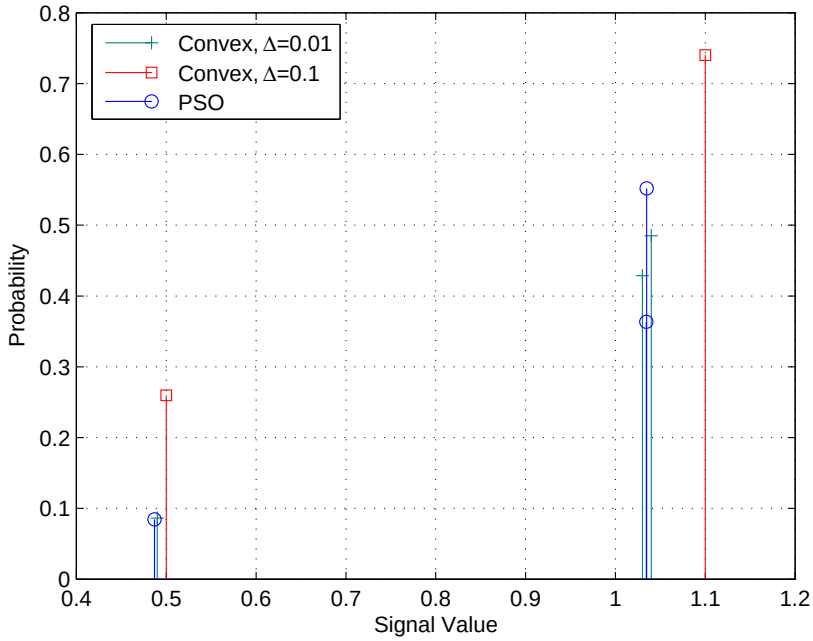

Figure 2.5: PMFs of the PSO and the convex optimization algorithms for the sign detector in Fig. 2.2 at $A/\sigma^2 = 40$ dB.

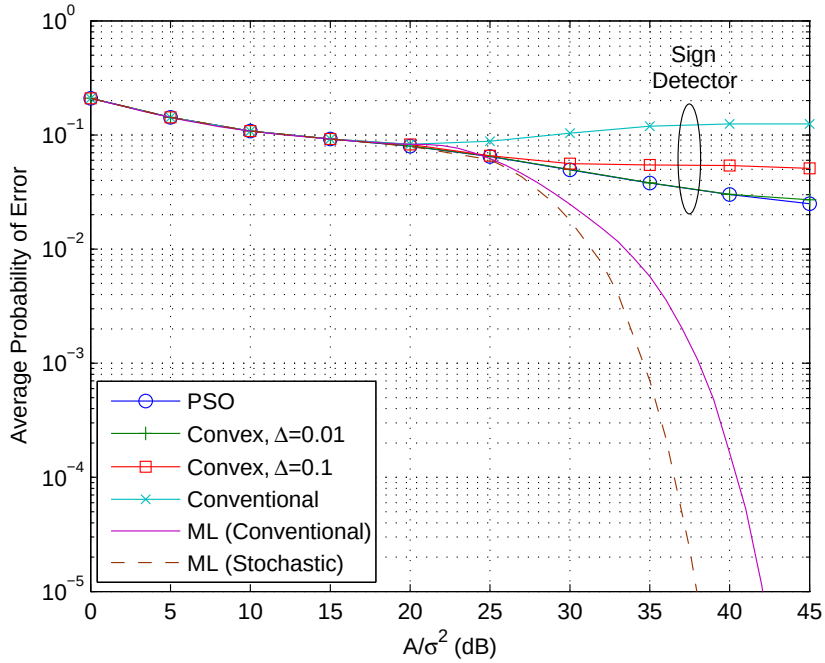

Figure 2.6: Error probability versus A/σ^2 for $\kappa = 1.5$. A symmetric Gaussian mixture noise, which has its mass points at $\pm[0.19\ 0.39\ 0.83\ 1.03]$, each with equal weight, is considered.

presence of multiple users, the noise can be modeled as $N = \sum_{k=2}^{K} A_k b_k + \eta$, where $b_k \in \{-1, 1\}$ with equal probabilities and η is a zero-mean Gaussian thermal noise component with variance σ^2. Then, for $K = 4$, $A_2 = 0.1$, $A_3 = 0.61$ and $A_2 = 0.32$, the noise becomes Gaussian mixture noise with 8 mass points as specified at the beginning of the paragraph. In Fig. 2.6, the average error probabilities of various algorithms are plotted against the A/σ^2 for $\kappa = 1.5$.

Also the plots of $G(x)$ at $A/\sigma^2 = 0, 25, 40$ dB are presented in Fig. 2.7, and the probability distributions at $A/\sigma^2 = 25$ dB and $A/\sigma^2 = 40$ dB are illustrated in Fig. 2.8 and Fig. 2.9, respectively, for the sign detector. Although similar observations as in the previous scenario can be made, a number of differences are also noticed. The improvements achieved via the stochastic signaling over the conventional (deterministic) signaling are less than those observed in Fig. 2.2. In addition, since $\kappa = 1.5$ in this scenario, only the second moment constraint is achieved with equality in all the solutions.

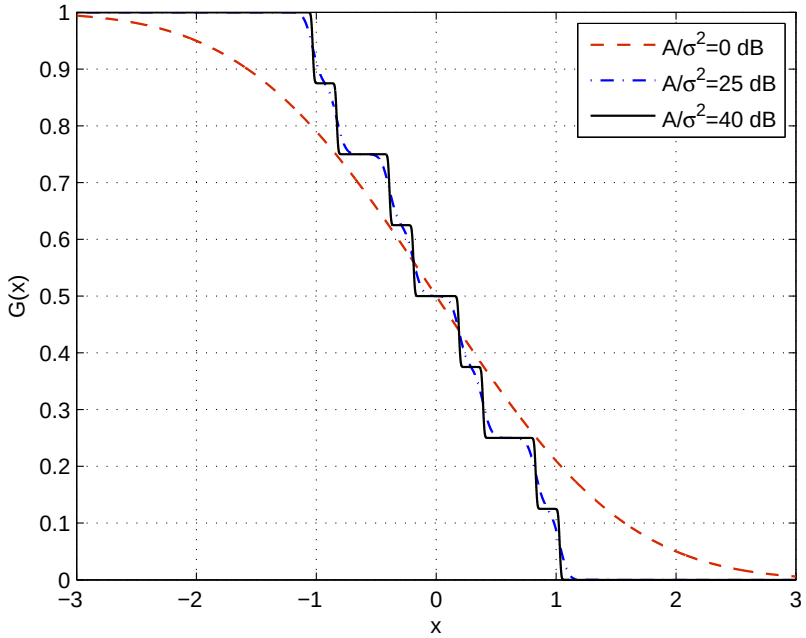

Figure 2.7: $G(x)$ in (2.9) for the sign detector in Fig. 2.6 at A/σ^2 values of 0, 25 and 40 dB.

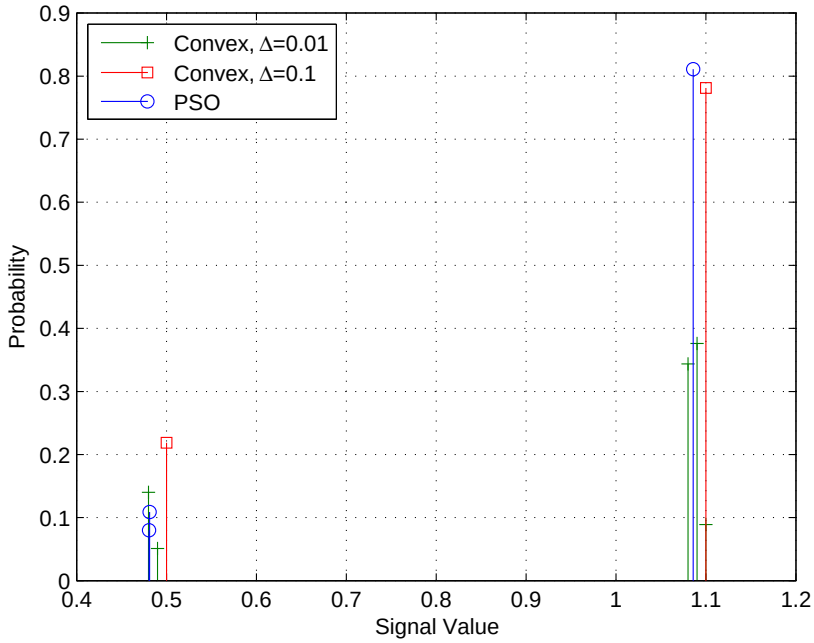

Figure 2.8: PMFs of the PSO and the convex optimization algorithms for the sign detector in Fig. 2.6 at $A/\sigma^2 = 25$ dB.

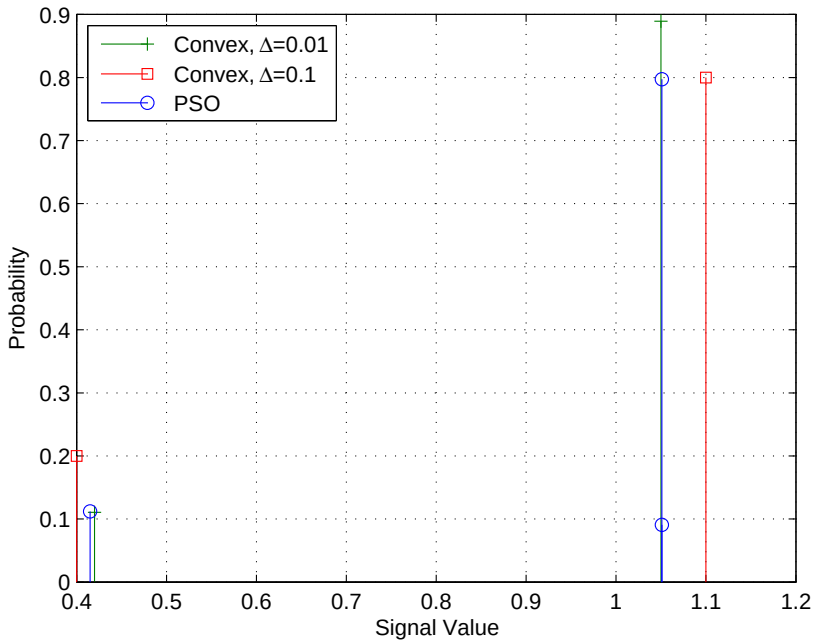

Figure 2.9: PMFs of the PSO and the convex optimization algorithms for the sign detector in Fig. 2.6 at $A/\sigma^2 = 40$ dB.

In order to investigate the optimal stochastic signaling for the ML detectors studied in Fig. 2.2 and Fig. 2.6, Table 2.1 presents the PDFs of the optimal stochastic signals in those scenarios, where the optimal PDFs are expressed in the form of $p_S(x) = \lambda_1\,\delta(x - x_1) + \lambda_2\,\delta(x - x_2) + \lambda_3\,\delta(x - x_3)$. It is observed from the table that the conventional deterministic signaling is optimal at low A/σ^2 values, which can also be verified from Fig. 2.2 and Fig. 2.6 since there is no improvement via the stochastic signaling over the conventional one for those A/σ^2 values. However, as A/σ^2 increases, the optimal signaling is achieved via randomization between two signal values. In those cases, significant improvements over the conventional signaling can be achieved as observed from Fig. 2.2 and Fig. 2.6. Finally, it is noted from the table that the optimal solutions result in randomization between at most two different signal levels in this example. This is in compliance with Proposition 2.3 since the proposition does not guarantee the existence of three different signal levels in general but states that an optimal

Table 2.1: Optimal stochastic signals for the ML detectors in Fig. 2.2 (top block) and Fig. 2.6 (bottom block).

A/σ^2 (dB)	λ_1	λ_2	λ_3	x_1	x_2	x_3
10	1	0	0	1	N/A	N/A
15	1	0	0	1	N/A	N/A
20	0.1181	0.8819	0	1.4211	0.9151	N/A
25	0.1264	0.8736	0	1.4494	0.8876	N/A
27.5	0.1317	0.8683	0	1.4465	0.8811	N/A
10	1	0	0	1	N/A	N/A
15	1	0	0	1	N/A	N/A
20	0.1272	0.8728	0	0.5073	1.0527	N/A
25	0.9791	0.0209	0	0.9950	1.2116	N/A
30	0.9415	0.0585	0	0.9859	1.2047	N/A
35	0.9236	0.0764	0	0.9823	1.1936	N/A

signal can be represented by a randomization of *at most* three different signal levels.

2.1.4 Extensions to M-ary Pulse Amplitude Modulation (PAM)

The results in the study can be extended to M-ary PAM communications systems for $M > 2$ as well. To that aim, consider a generic detector which chooses the ith symbol if the observation is in decision region Γ_i for $i = 0, 1, \ldots, M-1$. In other words, the decision rule is defined as

$$\phi(y) = i \ , \quad \text{if } y \in \Gamma_i \ , \quad i = 0, 1, \ldots, M-1 \ . \tag{2.40}$$

Then, the average probability of error for an M-ary system can be expressed as

$$\mathrm{P}_{\text{avg}} = \sum_{i=0}^{M-1} \pi_i \left(1 - \mathrm{P}_i(\Gamma_i)\right) \ , \tag{2.41}$$

where π_i denotes the prior probability of the ith symbol.

If signals $S_0, S_1, \ldots, S_{M-1}$ are modeled as stochastic signals with PDFs $p_{S_0}, p_{S_1}, \ldots, p_{S_{M-1}}$, respectively, the average probability of error in (2.41) can

be expressed, similarly to (2.6), as

$$\mathrm{P}_{\mathrm{avg}}^{\mathrm{stoc}} = \sum_{i=0}^{M-1} \pi_i \left(1 - \int_{-\infty}^{\infty} p_{S_i}(t) \int_{\Gamma_i} p_N(y-t)\, dy\, dt\right) . \tag{2.42}$$

Then, the optimal stochastic signaling problem can be stated as

$$\begin{aligned} &\min_{p_{S_0},\ldots,p_{S_{M-1}}} \sum_{i=0}^{M-1} \pi_i \left(1 - \int_{-\infty}^{\infty} p_{S_i}(t) \int_{\Gamma_i} p_N(y-t)\, dy\, dt\right) \\ &\text{subject to } \ \mathrm{E}\{|S_i|^2\} \le A \ , \quad \mathrm{E}\{|S_i|^4\} \le \kappa A^2 \ , \quad i = 0, 1, \ldots, M-1 \ . \end{aligned} \tag{2.43}$$

Due to the structure of the objective function in (2.43) and the individual constraints on each signal, M separate optimization problems, similar to (2.8), can be obtained. Namely, for $i = 0, 1, \ldots, M-1$,

$$\begin{aligned} &\min_{p_{S_i}} \ 1 - \int_{-\infty}^{\infty} p_{S_i}(t) \int_{\Gamma_i} p_N(y-t)\, dy\, dt \\ &\text{subject to } \ \mathrm{E}\{|S_i|^2\} \le A \ , \quad \mathrm{E}\{|S_i|^4\} \le \kappa A^2 \ . \end{aligned} \tag{2.44}$$

In addition, if auxiliary functions $G_i(x)$ are defined as $G_i(x) \triangleq 1 - \int_{\Gamma_i} p_N(y-x)\, dy$ for $i = 0, 1, \ldots, M-1$, the optimization problem in (2.44) can be expressed as

$$\begin{aligned} &\min_{p_{S_i}} \ \mathrm{E}\{G_i(S_i)\} \\ &\text{subject to } \ \mathrm{E}\{|S_i|^2\} \le A \ , \quad \mathrm{E}\{|S_i|^4\} \le \kappa A^2 \end{aligned} \tag{2.45}$$

for $i = 0, 1, \ldots, M-1$. Since (2.45) is in the same form as (2.10), the results in Section 2.1.2 can be extended to M-ary PAM systems, as well.

2.1.5 Concluding Remarks and Extensions

In this section, the stochastic signaling problem under second and fourth moment constraints has been studied for binary communications systems. It has been shown that, under certain monotonicity and convexity conditions, the conventional signaling, which employs deterministic signals at the average power limit, is optimal. On the other hand, in some cases, a smaller average probability of error can achieved by using a signal that is obtained by a randomization of

multiple signal values. In addition, it has been shown that an optimal signal can be represented by a discrete random variable with at most three mass points, which simplifies the optimization problem for the optimal signal design considerably. Furthermore, it has been observed that the optimal signals achieve at least one of the second and fourth moment constraints in most practical scenarios. Finally, two techniques based on PSO and convex relaxation have been proposed to obtain the optimal signals, and simulation results have been presented.

In addition, the results in this section can be extended to a generic binary hypothesis-testing problem in the Bayesian framework [2], [15].[8] In that case, the average probability of error expression in (2.3) is generalized to the *Bayes risk*, defined as $\pi_0[C_{00}\mathrm{P}_0(\Gamma_0)+C_{10}\mathrm{P}_0(\Gamma_1)]+\pi_1[C_{01}\mathrm{P}_1(\Gamma_0)+C_{11}\mathrm{P}_1(\Gamma_1)]$, where $C_{ij}\geq 0$ represents the cost of deciding the ith hypothesis when the jth one is true. Then, all the results in this section are still valid when function G in (2.9) is replaced by $G(x)=C_{01}\int_{\Gamma_0}p_N(y-x)dy+C_{11}\int_{\Gamma_1}p_N(y-x)dy$. Moreover, it can be shown that the results in this section are valid in the *minimax* and *Neyman-Pearson* frameworks [2] due to the decoupling of the optimization problem discussed in Section 2.1.2.

[8]Hence, the results in this study can be applied to other systems than communications, as well.

2.2 Stochastic Signaling Under an Average Power Constraint

2.2.1 System Model and Motivation

Consider a scalar binary communications system, as in [6] and [8], in which the received signal is given by

$$Y = S_i + N \ , \qquad i \in \{0, 1\} \ , \tag{2.46}$$

where S_0 and S_1 denote the transmitted signal values for symbol 0 and symbol 1, respectively, and N is the noise component that is independent of S_i. In addition, the prior probabilities of the symbols, which are denoted by π_0 and π_1, are supposed to be known [38].

Note that the probability distribution of the noise component in (2.46) is not necessarily Gaussian. Due to interference, such as multiple-access interference, the noise component can have a probability distribution that is different from the Gaussian distribution [7], [6].

A generic decision rule is considered at the receiver to estimate the symbol in (2.46). Specifically, for a given observation $Y = y$, the decision rule $\phi(y)$ is expressed as

$$\phi(y) = \begin{cases} 0 \ , & y \in \Gamma_0 \\ 1 \ , & y \in \Gamma_1 \end{cases} \ , \tag{2.47}$$

where Γ_0 and Γ_1 are the decision regions for symbol 0 and symbol 1, respectively [2].

In this study, the aim is to design signals S_0 and S_1 in (2.46) in order to minimize the average probability of error for a given decision rule, which is given

by

$$\mathrm{P}_{\mathrm{avg}} = \pi_0 \mathrm{P}_0(\Gamma_1) + \pi_1 \mathrm{P}_1(\Gamma_0) \ , \tag{2.48}$$

where $\mathrm{P}_i(\Gamma_j)$ is the probability of selecting symbol j when symbol i is transmitted. In practical systems, the signal are commonly subject to an average power constraint, which can be expressed as

$$\pi_0 \mathrm{E}\{|S_0|^2\} + \pi_1 \mathrm{E}\{|S_1|^2\} \leq A \ , \tag{2.49}$$

where A is the average power limit. Therefore, the problem is to calculate the optimal probability density functions (PDFs) for signals S_0 and S_1 that minimize the average probability of error in (2.48) under the average power constraint in (2.49).

The main motivation for the optimal stochastic signaling problem is to enhance the error performance of a communications system by considering the signals at the transmitter as random variables and obtaining the optimal probability distributions for those signals [6],[11], [38].

In the conventional signal design, S_0 and S_1 are considered as deterministic signals and they are designed in such a way that the Euclidean distance between them is maximized under the constraint in (2.49). In fact, when the effective noise has a zero-mean Gaussian PDF and the receiver employs the MAP decision rule, the probability of error is minimized when the Euclidean distance between the signals is maximized for a given average power constraint [2]. To that aim, S_0 and S_1 can conventionally be set to

$$S_0 = -\sqrt{A}/\alpha \ \text{ and } \ S_1 = \alpha\sqrt{A} \ , \tag{2.50}$$

where $\alpha \triangleq \sqrt{\pi_0/\pi_1}$ by considering the average power constraint in (2.49) (see [3] for the derivation). Then, the average probability of error in (2.48) becomes

$$\begin{aligned} \mathrm{P}_{\mathrm{avg}}^{\mathrm{conv}} &= \pi_0 \int_{\Gamma_1} p_N \left(y + \sqrt{A}/\alpha\right) dy \\ &\quad + \pi_1 \int_{\Gamma_0} p_N \left(y - \alpha\sqrt{A}\right) dy \ , \end{aligned} \tag{2.51}$$

where $p_N(\cdot)$ is the PDF of the noise in (2.46). Although the conventional signal design is optimal for certain classes of noise PDFs and decision rules, in some cases, the use of stochastic signals instead of deterministic ones can improve the system performance, as studied in the next section.

2.2.2 Optimal Stochastic Signaling

Instead of using constant levels for S_0 and S_1 as in the conventional case, one can consider a more generic scenario in which the signals can be stochastic. Then, the aim is to calculate the optimal PDFs for S_0 and S_1 in (2.46) that minimize the average probability of error under the constraint in (2.49).

Let $p_{S_0}(\cdot)$ and $p_{S_1}(\cdot)$ denote the PDFs for S_0 and S_1, respectively. Then, from (2.48), the average probability of error for the decision rule in (2.47) is given by

$$\mathrm{P}_{\mathrm{avg}}^{\mathrm{stoc}} = \sum_{i=0}^{1} \pi_i \int_{-\infty}^{\infty} p_{S_i}(t) \int_{\Gamma_{1-i}} p_N(y-t)\, dy\, dt \ . \tag{2.52}$$

Therefore, the optimal stochastic signal design problem can be expressed as

$$\begin{aligned} &\min_{p_{S_0}, p_{S_1}} \ \mathrm{P}_{\mathrm{avg}}^{\mathrm{stoc}} \\ &\text{subject to} \ \ \pi_0 \mathrm{E}\{|S_0|^2\} + \pi_1 \mathrm{E}\{|S_1|^2\} \leq A \ . \end{aligned} \tag{2.53}$$

After some manipulation, the objective function in (2.52) can be expressed as

$$\mathrm{P}_{\mathrm{avg}}^{\mathrm{stoc}} = \pi_0 \int_{-\infty}^{\infty} p_{S_0}(x)(1-G(x))dx + \pi_1 \int_{-\infty}^{\infty} p_{S_1}(x) G(x) dx \ , \tag{2.54}$$

where $G(x)$ is defined as

$$G(x) \triangleq \int_{\Gamma_0} p_N(y-x)\, dy \ . \tag{2.55}$$

Then the expression in (2.54) can be written in terms of the expectation of $G(S_1)$ over S_1 and that of $G(S_0)$ over S_0 as

$$\mathrm{P}_{\mathrm{avg}}^{\mathrm{stoc}} = \pi_0 - \pi_0 \mathrm{E}\{G(S_0)\} + (1-\pi_0)\mathrm{E}\{G(S_1)\} \ . \tag{2.56}$$

Signals S_0 and S_1 can be expressed as the elements of a vector random variable $\mathbf{S}$ as $\mathbf{S} \triangleq [S_0 \; S_1]$. Then the final form of optimization problem in (2.53) can be formulated as

$$\min_{p_{\mathbf{S}}} \mathrm{E}\{F(\mathbf{S})\} \;\; \text{subject to} \;\; \mathrm{E}\{H(\mathbf{S})\} \leq A \; , \tag{2.57}$$

where the expectations are taken over $\mathbf{S}$, $p_{\mathbf{S}}(\cdot)$ denotes the joint PDF of S_0 and S_1,

$$F(\mathbf{S}) \triangleq (1-\pi_0)\, G(S_1) - \pi_0\, G(S_0) + \pi_0 \; , \tag{2.58}$$

and

$$H(\mathbf{S}) \triangleq (1-\pi_0)|S_1|^2 + \pi_0 |S_0|^2 \; . \tag{2.59}$$

Note that there are also implicit constraints in the optimization problem in (2.57), since $p_{\mathbf{S}}(\mathbf{s})$ is a joint PDF.

2.2.2.1 On the Optimality of Conventional Signaling

In some cases, the conventional signaling is the optimal approach; that is, setting $p_{\mathbf{S}}(\mathbf{s}) = \delta(\mathbf{s} - \mathbf{S}_{\mathrm{A}})$, where $\mathbf{S}_{\mathrm{A}} = [-\sqrt{A}/\alpha \;\; \alpha\sqrt{A}]$ with $\alpha = \sqrt{\pi_0/\pi_1}$, can solve the optimization problem in (2.57). In this section, we derive sufficient conditions that guarantee the optimality of the conventional signaling scheme.

Proposition 2.5: *Assume that $G(x)$ in (2.55) is twice continuously differentiable. Then, $p_{\mathbf{S}}(\mathbf{s}) = \delta(\mathbf{s} - \mathbf{S}_A)$ is a solution of the optimization problem in (2.57), if the following three conditions are satisfied:*

- *$G(x)$ is a strictly decreasing function.*
- *$x\, G''(x) > 0$, $\forall x \neq 0$, and $G''(0) = 0$.*
- *For every (x_0, x_1) that satisfies*

$$\pi_1[G(\alpha\sqrt{A}) - G(x_1)] > \pi_0[G(-\sqrt{A}/\alpha)) - G(x_0)], \tag{2.60}$$

$\pi_0\, x_0^2 + \pi_1\, x_1^2 > A$ *is satisfied as well.*

Proof: In this proof, it is shown by contradiction that, when the conditions in the proposition are satisfied, there exist no signal PDFs that can result in a lower probability of error than the conventional signal $\mathbf{S}_{\mathrm{A}}$ under the given average power constraint. To that aim, it is first assumed that there exists a PDF $p_{\mathbf{S}}(\mathbf{s})$ for signal $\mathbf{S} = [S_0\ S_1]$ such that $\mathrm{E}\{F(\mathbf{S})\} < F(\mathbf{S}_{\mathrm{A}})$ and $\mathrm{E}\{H(\mathbf{S})\} \leq A$. In other words, suppose that there exists a signal $\mathbf{S}$, with PDF $p_{\mathbf{S}}(\mathbf{s})$, which is better than the conventional signaling (see (2.57)). In addition, it is assumed without loss of generality that S_0 is a nonpositive and S_1 is a nonnegative random variable. [This assumption does not reduce the generality of the proof as $G(x)$ is a strictly decreasing function; hence, $F(\mathbf{S})$ in (2.58) is a strictly increasing (decreasing) function of S_0 (S_1). Since the average power depends only on the absolute value of the signals, choosing nonpositive S_0 and nonnegative S_1 always achieves the minimum average probability of error. In other words, for each positive (negative) value of S_0 (S_1), its negative (positive) can be used instead, which results in smaller average probability of error and the same average power value.]

Under the assumptions above, if it is shown that there can exist no PDF $p_{\mathbf{S}}(\mathbf{s})$ for the signal $\mathbf{S} = [S_0\ S_1]$, with S_0 being nonpositive and S_1 being nonnegative, that satisfies the three conditions in the proposition and $\mathrm{E}\{F(\mathbf{S})\} < F(\mathbf{S}_{\mathrm{A}})$ under the average power constraint, it means that there can exist no signal PDF $p_{\mathbf{S}}(\mathbf{s})$ (for any signs of S_0 and S_1) that has lower probability of error than the conventional signal under the average power constraint. For that purpose, it is shown in the following that $F(\mathbf{x})$ in (2.58) is a convex function. Since $F(\mathbf{x}) = (1-\pi_0)\, G(x_1) - \pi_0\, G(x_0) + \pi_0$, its Hessian matrix can be obtained as

$$\begin{bmatrix} \frac{\partial^2 F}{\partial x_0^2} & \frac{\partial^2 F}{\partial x_0\, \partial x_1} \\ \frac{\partial^2 F}{\partial x_1\, \partial x_0} & \frac{\partial^2 F}{\partial x_1^2} \end{bmatrix} = \begin{bmatrix} -\pi_0\, G''(x_0) & 0 \\ 0 & \pi_1\, G''(x_1) \end{bmatrix} . \tag{2.61}$$

Since S_0 is a nonpositive random variable, x_0 can take only nonpositive values and similarly since S_1 is a nonnegative random variable, x_1 can take only nonnegative values. Therefore, under the second condition in the proposition, namely, $x\,G''(x) > 0$, $\forall x \neq 0$, and $G''(0) = 0$, the Hessian matrix is always positive semidefinite; hence, $F(\mathbf{x})$ is a convex function.

Since $F(\mathbf{S})$ is a convex function, Jensen's inequality implies that $\mathrm{E}\{F(\mathbf{S})\} \geq F(\mathrm{E}\{\mathbf{S}\}) = F([\mathrm{E}\{S_0\}\ \mathrm{E}\{S_1\}])$. Then, $\mathrm{E}\{F(\mathbf{S})\} < F(\mathbf{S}_\mathrm{A})$ requires that $F([\mathrm{E}\{S_0\}\ \mathrm{E}\{S_1\}]) < F(\mathbf{S}_\mathrm{A})$, which can be expressed from (2.58) as

$$\pi_1\, G(\mathrm{E}\{S_1\}) - \pi_0\, G(\mathrm{E}\{S_0\}) < \pi_1\, G(\alpha\sqrt{A}) - \pi_0\, G(-\sqrt{A}/\alpha)\ . \tag{2.62}$$

In addition, Jensen's inequality also implies that $\mathrm{E}\{|S_0|^2\} \geq (\mathrm{E}\{S_0\})^2$ and $\mathrm{E}\{|S_1|^2\} \geq (\mathrm{E}\{S_1\})^2$. Therefore, $\pi_0\mathrm{E}\{|S_0|^2\} + \pi_1\mathrm{E}\{|S_1|^2\} \geq \pi_0(\mathrm{E}\{S_0\})^2 + \pi_1(\mathrm{E}\{S_1\})^2$ is obtained. At this point, defining $x_0 = \mathrm{E}\{S_0\}$ and $x_1 = \mathrm{E}\{S_1\}$, and plugging them into (2.62) yields $\pi_1\,[G(\alpha\sqrt{A}) - G(x_1)] > \pi_0\,[G(-\sqrt{A}/\alpha) - G(x_0)]$, which is the first inequality in the third condition of the proposition. According to the third condition, whenever this inequality is satisfied for any (x_0, x_1), $\pi_0\, x_0^2 + \pi_1\, x_1^2 > A$, equivalently, $\pi_0\mathrm{E}\{|S_0|^2\} + \pi_1\mathrm{E}\{|S_1|^2\} > A$, is also satisfied. Therefore, $\mathrm{E}\{H(\mathbf{S})\} > A$ always holds, which indicates that the average power constraint in (2.57) is violated. Hence, it is concluded that when the conditions in Proposition 2.5 are satisfied, no PDF can achieve $\mathrm{E}\{F(\mathbf{S})\} < F(\mathbf{S}_\mathrm{A})$ under the average power constraint. □

As an example application of Proposition 2.5, consider a zero mean and unit variance Gaussian noise N in (2.46) with $p_N(x) = \exp\{-x^2/2\}/\sqrt{2\pi}$, and assume equal priors ($\pi_0 = \pi_1 = 0.5$). Also, the average power constraint A in (2.57) is taken to be 1. In this case, the conventional signaling becomes the antipodal signaling with $S_0 = -1$ and $S_1 = 1$, and a decision rule of the form $\Gamma_0 = (-\infty, 0]$ and $\Gamma_1 = [0, \infty)$; that is, the sign detector, is the optimal MAP decision rule. Then, $G(x)$ in (2.55) can be calculated as $G(x) = Q(x)$, where $Q(x) = (\int_x^\infty \mathrm{e}^{-t^2/2}dt)/\sqrt{2\pi}$ defines the Q-function. Since $Q(x)$ is a monotone decreasing

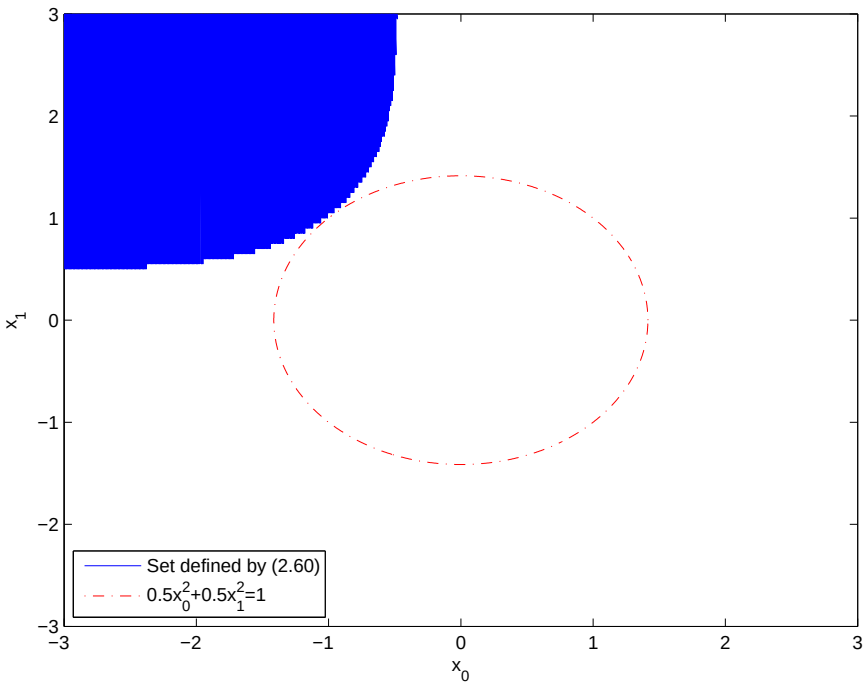

Figure 2.10: The region in which the inequality $Q(x_1) - Q(x_0) < Q(1) - Q(-1)$ is satisfied is outside of the circle $0.5\,x_0^2 + 0.5\,x_1^2 = 1$.

function and $x\,Q''(x) > 0$, $\forall x \neq 0$ with $Q''(0) = 0$, the first two conditions in Proposition 2.5 are satisfied. For the third condition, we need to check the region in which $Q(x_1) - Q(x_0) < Q(1) - Q(-1) = -0.6827$. Then, as $Q(x)$ is a decreasing function, if one can find (a, b) such that $Q(a) = Q(b) - 0.6827$, then for every $x_1 > a$ and $x_0 = b$, $Q(x_1) - Q(x_0) < -0.6827$ and $0.5x_0^2 + 0.5x_1^2 > 0.5a^2 + 0.5b^2$. Also, since the Q-function takes values only between 0 and 1, $b < -0.475$ should hold. A simple search on this region reveals that $0.5a^2 + 0.5b^2 \geq 1$, where the equality holds only at $(a, b) = (1, w - 1)$. This fact can be observed from Fig. 2.10 as well. The geometrical interpretation of the third condition in Proposition 2.5 is that the set of all (x_0, x_1) pairs that satisfy $\pi_1[G(\alpha\sqrt{A}) - G(x_1)] > \pi_0[G(-\sqrt{A}/\alpha)) - G(x_0)]$ should be completely outside of the elliptical region whose boundary is $\pi_0\,x_0^2 + \pi_1\,x_1^2 = A$.

In Fig. 2.10, this is shown for this example and it is observed that every point that satisfies the inequality $Q(x_1) - Q(x_0) < Q(1) - Q(-1)$, is located outside of the circle $0.5\,x_0^2 + 0.5\,x_1^2 = 1$. Thus, the third condition in Proposition 2.5 holds as well. Therefore, it is guaranteed that the conventional signaling is optimal in this scenario.

2.2.2.2 Sufficient Conditions for Improvability

In this section, we obtain sufficient conditions under which the performance of the conventional signaling approach can be improved via stochastic signaling.

Proposition 2.6: *Assume that $G(x)$ in (2.55) is twice continuously differentiable. Then, $p_{\mathbf{S}}(\mathbf{s}) = \delta(\mathbf{s} - \mathbf{S}_A)$ is not an optimal solution of (2.57) if*

$$G''(\alpha\sqrt{A}) < \frac{G'(\alpha\sqrt{A})}{\alpha\sqrt{A}} \ , \tag{2.63}$$

or, alternatively,

$$G''(-\sqrt{A}/\alpha) > \frac{G'(-\sqrt{A}/\alpha)}{-\sqrt{A}/\alpha} \ . \tag{2.64}$$

Proof: In order to prove the suboptimality of the conventional solution $p_{\mathbf{S}}(\mathbf{s}) = \delta(\mathbf{s} - \mathbf{S}_{\mathrm{A}})$, it is shown that, under the conditions in the proposition, there exist $\lambda \in (0,1)$, Δ_1, Δ_2, Δ_3, and Δ_4 such that[9]

$$p_{\mathbf{S}_2}(\mathbf{s}) = \lambda\,\delta(\mathbf{s} - (\mathbf{S}_{\mathrm{A}} + \boldsymbol{\epsilon}_1)) + (1-\lambda)\,\delta(\mathbf{s} - (\mathbf{S}_{\mathrm{A}} + \boldsymbol{\epsilon}_2)) \ , \tag{2.65}$$

where $\boldsymbol{\epsilon}_1 = [\Delta_1\ \Delta_2]$ and $\boldsymbol{\epsilon}_2 = [\Delta_3\ \Delta_4]$, yields a lower probability of error than than $p_{\mathbf{S}}(\mathbf{s})$ and satisfies the constraint in (2.57). Specifically, proving the existence of $\lambda \in (0,1)$, Δ_1, Δ_2, Δ_3, and Δ_4 that satisfy

$$\lambda\,F(\mathbf{S}_{\mathrm{A}} + \boldsymbol{\epsilon}_1) + (1-\lambda)\,F(\mathbf{S}_{\mathrm{A}} + \boldsymbol{\epsilon}_2) < F(\mathbf{S}_{\mathrm{A}}) \tag{2.66}$$

[9] It is assumed that Δ_1, Δ_2, Δ_3, and Δ_4 are not all zeros, since that would result in the conventional signaling.

and

$$\pi_0\left[\lambda\left(-\sqrt{A}/\alpha+\Delta_1\right)^2+(1-\lambda)\left(-\sqrt{A}/\alpha+\Delta_3\right)^2\right]+$$
$$\pi_1\left[\lambda\left(\alpha\sqrt{A}+\Delta_2\right)^2+(1-\lambda)\left(\alpha\sqrt{A}+\Delta_4\right)^2\right]=A \tag{2.67}$$

is *sufficient* to prove that the conventional signaling is not optimal. From (2.67), the following equation is obtained:

$$\left[\pi_0\left(\lambda\Delta_1^2+(1-\lambda)\Delta_3^2\right)+\pi_1\left(\lambda\Delta_2^2+(1-\lambda)\Delta_4^2\right)\right]/\sqrt{A}$$
$$=-2\left[\pi_1\left(\lambda\Delta_2\alpha+(1-\lambda)\Delta_4\alpha\right)-\pi_0\left(\frac{\Delta_1\lambda}{\alpha}+\frac{(1-\lambda)\Delta_3}{\alpha}\right)\right]. \tag{2.68}$$

Since the left-hand-side of the equality in (2.68) is always positive, the term on the right-hand-side should also be positive, which leads to the following inequality since $\alpha=\sqrt{\pi_0/\pi_1}$:

$$\lambda\Delta_2+(1-\lambda)\Delta_4<\lambda\Delta_1+(1-\lambda)\Delta_3\ . \tag{2.69}$$

In addition, from (2.58) and (2.66), the following inequality is obtained:

$$\lambda\pi_1\,G(\alpha\sqrt{A}+\Delta_2)+(1-\lambda)\pi_1\,G(\alpha\sqrt{A}+\Delta_4)-\lambda\pi_0\,G(-\sqrt{A}/\alpha+\Delta_1)$$
$$-(1-\lambda)\pi_0\,G(-\sqrt{A}/\alpha+\Delta_3)<\pi_1\,G(\alpha\sqrt{A})-\pi_0\,G(-\sqrt{A}/\alpha)\ . \tag{2.70}$$

For infinitesimally small Δ_1, Δ_2, Δ_3 and Δ_4, the first three terms of the Taylor series expansion for $G(\alpha\sqrt{A}+\Delta_2)$, $G(\alpha\sqrt{A}+\Delta_4)$, $G(-\sqrt{A}/\alpha+\Delta_1)$ and $G(-\sqrt{A}/\alpha+\Delta_3)$ can be used to approximate (2.70) as

$$G^{'}(\alpha\sqrt{A})[\lambda\pi_1\Delta_2+(1-\lambda)\pi_1\Delta_4]+G^{'}(-\sqrt{A}/\alpha)[-\lambda\pi_0\Delta_1-(1-\lambda)\pi_0\Delta_3]$$
$$+\frac{G^{''}(\alpha\sqrt{A})}{2}[\lambda\pi_1{\Delta_2}^2+(1-\lambda)\pi_1{\Delta_4}^2]+\frac{G^{''}(-\sqrt{A}/\alpha)}{2}[-\lambda\pi_0{\Delta_1}^2-(1-\lambda)\pi_0{\Delta_3}^2]<0\ . \tag{2.71}$$

For $\Delta_1=\Delta_3=0$, (2.69) becomes $\lambda\Delta_2+(1-\lambda)\Delta_4<0$ and (2.68) becomes $\pi_1(\lambda\Delta_2^2+(1-\lambda)\Delta_4^2)=-2\sqrt{A\pi_1\pi_0}\,(\lambda\Delta_2+(1-\lambda)\Delta_4)$. Then, (2.71) simplifies to

$$G^{'}(\alpha\sqrt{A})[\lambda\pi_1\Delta_2+(1-\lambda)\pi_1\Delta_4]+G^{''}(\alpha\sqrt{A})[-\sqrt{A\pi_0\pi_1}(\lambda\Delta_2+(1-\lambda)\Delta_4)]<0\ . \tag{2.72}$$

Since $\lambda\Delta_2+(1-\lambda)\Delta_4 < 0$, (2.72) implies that $G^{'}(\alpha\sqrt{A})\pi_1 - G^{''}(\alpha\sqrt{A})\sqrt{A\pi_0\pi_1} > 0$, which is equivalent to $G^{'}(\alpha\sqrt{A}) - G^{''}(\alpha\sqrt{A})(\alpha\sqrt{A}) > 0$; that is, the first condition in the proposition.

Similarly, for $\Delta_2 = \Delta_4 = 0$, (2.69) becomes $\lambda\Delta_1 + (1-\lambda)\Delta_3 > 0$ and (2.68) becomes $\pi_0(\lambda\Delta_1^2 + (1-\lambda)\Delta_3^2) = 2\sqrt{A\pi_1\pi_0}\,(\lambda\Delta_1 + (1-\lambda)\Delta_3)$. Then, (2.71) can be rewritten as follows:

$$\begin{aligned} &G^{'}(-\sqrt{A}/\alpha)[-\lambda\pi_0\Delta_1 - (1-\lambda)\pi_0\Delta_3] \\ &+ G^{''}(-\sqrt{A}/\alpha)[-\sqrt{A\pi_0\pi_1}\,(\lambda\Delta_1 + (1-\lambda)\Delta_3)] < 0\,. \end{aligned} \tag{2.73}$$

Since $\lambda\Delta_1+(1-\lambda)\Delta_3 > 0$, (2.73) becomes $G^{'}(-\sqrt{A}/\alpha)\pi_0 + G^{''}(-\sqrt{A}/\alpha)\sqrt{A\pi_0\pi_1} > 0$, which is equivalent to $G^{'}(-\sqrt{A}/\alpha) + G^{''}(-\sqrt{A}/\alpha)(\sqrt{A}/\alpha) > 0$. Hence, the second condition in the proposition is obtained.

This proof indicates that that $p_{\mathbf{S}_2}(\mathbf{s})$ in (2.65) can result in a lower probability of error than the conventional signaling for infinitesimally small Δ_2 and Δ_4 values along with $\Delta_1 = \Delta_3 = 0$, or, for infinitesimally small Δ_1 and Δ_3 values along with $\Delta_2 = \Delta_4 = 0$, which satisfy (2.68). $\square$

Proposition 2.6 provides simple sufficient conditions to determine if stochastic signaling can improve the probability of error performance of a given detector. A practical example is presented in Section 2.2.3 on the use of the results in the proposition.

2.2.2.3 Statistical Characteristics of Optimal Signals

The optimization problem in (2.57) may be difficult to solve in general since the optimization needs to be performed over a space of PDFs. However, by using the following result, that optimization problem can be formulated over a set of variables instead of functions, hence can be simplified to a great extent.

Lemma 2.1: *Assume that $G(x)$ in (2.55) is a continuous function and possible signal values for S_0 and S_1 reside in $[-\gamma, \gamma]$ for some finite $\gamma > 0$. Then, the solution of the optimization problem in (2.57) is in the form of*

$$p_{\mathbf{S}}(\mathbf{s}) = \lambda\, \delta(\mathbf{s} - \mathbf{s}_1) + (1 - \lambda)\, \delta(\mathbf{s} - \mathbf{s}_2)\ , \tag{2.74}$$

where $\lambda \in [0, 1]$ and $\mathbf{s}_i$ is two-dimensional vector for $i = 1, 2$.

Proof: Optimization problems in the form of (2.57) have been investigated in various studies in the literature [11], [12], [18], [25]. Under the conditions in the lemma, the optimal solution of (2.57) can be represented by a randomization of at most two signal levels as a result of Carathéodory's theorem [28], [39]. Hence, the optimal signal PDF can be expressed as in (2.74). □

Lemma 2.1 states that the optimal signal PDF that solves the optimization problem in (2.57) can be represented by a discrete probability distribution with at most two mass points. Therefore, the optimization problem in (2.57) can be simplified as follows:

$$\begin{aligned} &\min_{\lambda, \mathbf{s}_1, \mathbf{s}_2}\ \lambda F(\mathbf{s}_1) + (1 - \lambda) F(\mathbf{s}_2) \\ &\text{subject to}\ \ \lambda H(\mathbf{s}_1) + (1 - \lambda) H(\mathbf{s}_2) \leq A\ . \end{aligned} \tag{2.75}$$

In other words, instead of optimization over functions, an optimization over a five-dimensional space (two two-dimensional mass points, $\mathbf{s}_1$ and $\mathbf{s}_2$, plus the weight, λ) can be considered for the optimal signaling problem as a result of Lemma 2.1.

Although (2.75) is significantly simpler than (2.57), it can still be a nonconvex optimization problem in general. Therefore, global optimization techniques such as particle-swarm optimization (PSO) [13], [31], [32], genetic algorithms and differential evolution [29], can be used to obtain the optimal solution [18], [19]. In the next section, the PSO algorithm is used to calculate the optimal stochastic signals in the numerical examples. For the details of the PSO algorithm, please

refer to [13] and for the PSO parameters used in PSO approach on this section, please refer to [40].

2.2.3 Numerical Results

In this section, a numerical example is presented to show the improvements over conventional signaling via optimal stochastic signaling. For this example, a binary communications system with priors $\pi_0 = 0.2$ and $\pi_1 = 0.8$ is considered [3]. Hence $\alpha = \sqrt{\pi_0/\pi_1}$ is equal to 0.5 in this case. Also, the average power constraint A is set to 1. It is assumed that the receiver employs a simple threshold detector such that $\Gamma_0 = (-\infty, \tau)$ and $\Gamma_1 = (\tau, \infty)$, where $\tau = (2\sigma^2 \ln(0.25) - 3.75)/5$. In fact, this is the optimal MAP decision rule for given the prior probabilities and the average power constraint, when the conventional signaling is performed and the noise is zero-mean Gaussian noise with variance σ^2.

In this example, the effective noise in (2.46) is modeled by Gaussian mixture noise [7], whose PDF can be expressed as

$$p_N(y) = \frac{1}{\sqrt{2\pi}\,\sigma} \sum_{l=1}^{L} \mathbf{v}_l \, \mathrm{e}^{-\frac{(y-\boldsymbol{\mu}_l)^2}{2\sigma^2}} \ . \tag{2.76}$$

By using this noise model, and the receiver structure specified above, $G(x)$ in (2.55) can be obtained as

$$G(x) = \sum_{l=1}^{L} \mathbf{v}_l \, Q\big(\frac{-\tau + x + \boldsymbol{\mu}_l}{\sigma}\big) \ . \tag{2.77}$$

In the numerical example, $\mathbf{v} = [0.035\ 0.465\ 0.465\ 0.035]$ and $\boldsymbol{\mu} = [-1.251\ -0.7\ 0.7\ 1.251]$ are used. Gaussian mixture noise is encountered in practical systems in the presence of interference [7]. Note that the variance of each component of the Gaussian mixture noise is set to σ^2 and the average power of the noise can be calculated as $\mathrm{E}\{N^2\} = \sigma^2 + 0.5653$ for the given values.

In this example, three different signaling schemes are considered:

Conventional Signaling: In this case, the transmitter selects the signals as $S_0 = -\sqrt{A}/\alpha = -2$ and $S_1 = \sqrt{A}\,\alpha = 0.5$, which are known to be optimal if the noise is zero-mean Gaussian and the receiver structure is as specified above [2].

Stochastic Signaling: In this case, the solution of the most generic optimization problem in (2.53) is obtained. Since that problem can be reduced to the optimization problem in (2.75), the optimal stochastic signals are calculated via PSO based on the formulation (2.75) in this scenario.

Deterministic Signaling: In this case, it is assumed that the signals are deterministic, and the optimization problem in (2.57) is solved under that assumption. That is, the optimal signal PDF is given by $p_{\mathbf{S}}(\mathbf{s}) = \delta(\mathbf{s} - \mathbf{s}^*)$, where $\mathbf{s}^*$ is the solution of the following optimization problem:

$$\begin{aligned} &\min_{\mathbf{s}} \; F(\mathbf{s}) \\ &\text{subject to} \;\; H(\mathbf{s}) \leq A \; . \end{aligned} \tag{2.78}$$

In other words, this solution provides a simplified version of the optimal solution in (2.57). Indeed, there are two optimization variables (two signal levels, S_0 and S_1) in this case, instead of the five optimization variables in the stochastic signaling case (see (2.75)).

In Fig. 2.11, the average probabilities of error are plotted versus A/σ^2 for the three signaling schemes. In order to calculate both the stochastic signaling and the deterministic signaling solutions, the PSO approach is used. From Fig. 2.11, it is observed that for low values of σ, the conventional signaling performs worse than the others, and the stochastic signaling achieves the lowest probabilities of error. Specifically, after A/σ^2 exceeds 30 dB, significant improvements can be obtained via stochastic signaling over the conventional and deterministic signaling approaches. Indeed, improvements are expected based on Proposition 2.6 as well. For example, at 30 dB, $G^{''}(-2) = 0.6514$ and $G^{'}(-2) = -0.441$, and at 40 dB, $G^{''}(-2) = 13.84$ and $G^{'}(-2) = -1.389$, which results in $G^{''}(-2) > -G^{'}(-2)/2$

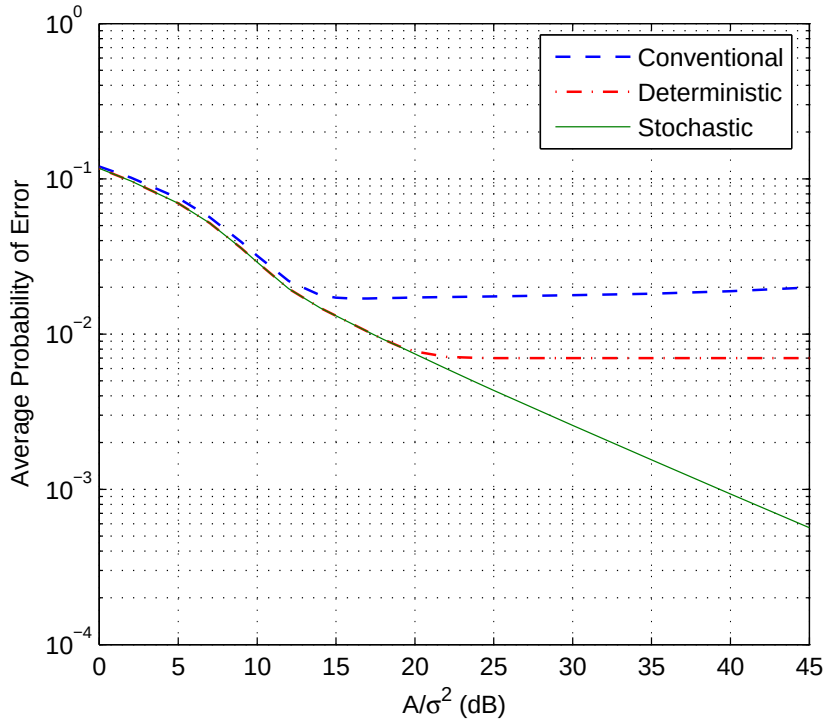

Figure 2.11: Average probability of error versus A/σ^2 for conventional, optimal deterministic, and optimal stochastic signaling.

for both of the cases. Therefore, the second sufficient condition in Proposition 2.6 (i.e., the inequality in (2.64)) is satisfied and improvements over the conventional solution are guaranteed in those scenarios.

Moreover, it should be noted that the average probability of error does not monotonically decrease for the conventional and deterministic solutions as A/σ^2 increases. This is because of the fact that average probability of error is related to the area under the two shifted noise PDFs as in (2.51). Since the noise PDF has a multimodal PDF in this example, and the amount of shifts that can be imposed on the noise PDFs is restricted by the average power constraint, that area may increase or remain same as A/σ^2 increases in some cases.

In order to provide further explanations of the results, Table 2.2 and 2.3 present the solutions of the stochastic and deterministic signaling schemes for some A/σ^2 values. In Table 2.2, the optimal $\mathbf{s}_1$ and $\mathbf{s}_2$ in (2.75) are expressed as

Table 2.2: Optimal stochastic signaling.

A/σ^2(dB)	λ	s_{11}	s_{12}	s_{21}	s_{22}
0	1	-1.8221	0.6480	N/A	N/A
15	1	-1.8424	0.6336	N/A	N/A
30	0.3149	-1.5467	0.5782	-2.0607	0.5782
45	0.0733	-1.4702	0.5185	-2.0159	0.5185

Table 2.3: Optimal deterministic signaling.

A/σ^2(dB)	S_0	S_1
0	-1.8221	0.6480
15	-1.8424	0.6336
30	-1.6911	0.7314
45	-1.6249	0.7306

$\mathbf{s}_1 = [s_{11}\ s_{12}]$ and $\mathbf{s}_2 = [s_{21}\ s_{22}]$ for each A/σ^2 value. For small A/σ^2 values, such as 0 dB and 15 dB, the deterministic solutions are the same as the stochastic ones. In fact, the performance of the deterministic and the stochastic signaling is same for A/σ^2 values less than 20 dB, as can be observed from Fig. 2.11. Also, their performance is very close to the performance of conventional signaling at high σ values. For example, at 0 dB, the average probability of error for the conventional signaling is 0.120, and it is 0.117 for the other schemes.

Furthermore, it can be observed from Table 2.2 that as A/σ^2 increases, the randomization between two signal vectors becomes more effective and this helps reduce the average probability of error as compared with the other signaling schemes. For example, at $A/\sigma^2 = 45$ dB, the average probability of error for the stochastic signaling is 5.66×10^{-4}, whereas it is 0.007 and 0.02 for the deterministic signaling and the conventional signaling schemes, respectively.

2.2.4 Concluding Remarks

The optimal stochastic signaling problem has been studied under an average power constraint. It has been shown that, under certain conditions, the conventional signaling approach, which maximizes the Euclidean distance between the signals, is the optimal signaling strategy. Also, sufficient conditions have been obtained to specify when randomization between different signal values may result in improved performance in terms of the average probability of error. In addition, the discrete structure of the optimal stochastic signals has been specified, and a global optimization technique, called PSO, has been used to solve the generic stochastic signaling problem under the average power constraint. Finally, numerical examples have been presented to illustrate some applications of the theoretical results.

Chapter 3

OPTIMAL SIGNALING AND DETECTOR DESIGN FOR POWER CONSTRAINED COMMUNICATION SYSTEMS

In this chapter, joint optimization of signal structures and detectors is studied for binary communications systems under average power constraints in the presence of additive non-Gaussian noise. First, it is observed that the optimal signal for each symbol can be characterized by a discrete random variable with at most two mass points. Then, optimization over all possible two mass point signals and corresponding maximum *a posteriori* probability (MAP) decision rules are considered. It is shown that the optimization problem can be simplified into an optimization over a number of signal parameters instead of functions, which can be solved via global optimization techniques, such as particle swarm optimization. Finally, the improvements that can be obtained via the joint design of the signaling and the detector are illustrated via an example.

3.1 Optimal Signaling and Detector Design

Consider a binary communications system, in which the receiver obtains K-dimensional observations over an additive noise channel [41]:

$$\mathbf{y} = \mathbf{s}_i + \mathbf{n} \ , \quad i \in \{0, 1\} \ , \tag{3.1}$$

where $\mathbf{y}$ is the noisy observation, $\mathbf{s}_0$ and $\mathbf{s}_1$ represent the transmitted signal values for symbol 0 and symbol 1, respectively, and $\mathbf{n}$ is the noise component that is independent of $\mathbf{s}_i$. In addition, the prior probabilities of the symbols, represented by π_0 and π_1, are assumed to be known. The signal model in (3.1) can be considered for flat-fading channels assuming perfect channel estimation; that is, the model in (3.1) can be obtained after appropriate equalization [41].

The receiver uses the observation in (3.1) in order to determine the information symbol. A generic decision rule (detector) is considered for that purpose, which estimates the transmitted symbol based on a given observation $\mathbf{y}$ as follows:

$$\phi(\mathbf{y}) = \begin{cases} 0 \ , & \mathbf{y} \in \Gamma_{\phi_0} \\ 1 \ , & \mathbf{y} \in \Gamma_{\phi_1} \end{cases} \ , \tag{3.2}$$

where Γ_{ϕ_0} and Γ_{ϕ_1} are the decision regions for symbol 0 and symbol 1, respectively [2].

The average probability of error for a decision rule ϕ can be expressed as $\mathrm{P_e} = \pi_0 \mathrm{P_{e,0}} + \pi_1 \mathrm{P_{e,1}}$, where

$$\mathrm{P_{e,}}_i = \int_{\Gamma_{\phi_{1-i}}} p_i(\mathbf{y}) \, d\mathbf{y} \ , \tag{3.3}$$

for $i = 0, 1$, represents the probability of error, with $p_i(\mathbf{y})$ denoting the conditional probability density function (PDF) of the observation, when the ith symbol is transmitted.

Unlike the conventional case, a stochastic signaling framework is adopted in this study [38], and $\mathbf{s}_0$ and $\mathbf{s}_1$ in (3.1) are modeled as random variables. Since the signals and the noise are independent, the conditional PDFs of the observation can be calculated as $p_i(\mathbf{y}) = \int_{\mathbb{R}^K} p_{\mathbf{s}_i}(\mathbf{x}) p_{\mathbf{n}}(\mathbf{y}-\mathbf{x})\, d\mathbf{x}$ for $i = 0, 1$. Then, after some manipulation, (3.3) can be expressed as

$$\mathrm{P}_{\mathrm{e},i} = \mathrm{E}\left\{ \int_{\Gamma_{\phi_{1-i}}} p_{\mathbf{n}}(\mathbf{y}-\mathbf{s}_i)\, d\mathbf{y} \right\} \triangleq \mathrm{E}\left\{ f(\phi\,;\mathbf{s}_i) \right\} \ , \tag{3.4}$$

where the expectation is taken over the PDF of $\mathbf{s}_i$.

In practical systems, there is a constraint on the average power of the signals, which can be expressed as [2]

$$\mathrm{E}\left\{ |\mathbf{s}_i|^2 \right\} \leq A \ , \quad \text{for} \ \ i = 0, 1 \ , \tag{3.5}$$

where A is the average power limit. Then, the optimal signaling and detector design problem can be stated as

$$\begin{aligned} &\min_{p_{\mathbf{s}_0}, p_{\mathbf{s}_1}, \phi} \ \pi_0 \mathrm{P}_{\mathrm{e},0} + \pi_1 \mathrm{P}_{\mathrm{e},1} \\ &\text{subject to } \mathrm{E}\left\{ |\mathbf{s}_i|^2 \right\} \leq A \ , \quad i = 0, 1 \ , \end{aligned} \tag{3.6}$$

where $\mathrm{P}_{\mathrm{e},i}$ is as in (3.4).

The problem in (3.6) is difficult to solve in general since the optimization needs to be performed over a space of PDFs and decision rules. In the following, a simpler optimization problem over a set of variables (instead of functions) is formulated in order to obtain optimal signal PDFs and the decision rule. To that aim, the following result is obtained first.

Lemma 3.1: *Assume $f(\phi\,;\mathbf{s}_i)$ in (3.4) is a continuous function of $\mathbf{s}_i$, and each component of $\mathbf{s}_i$ resides in $[-\gamma, \gamma]$ for some finite $\gamma > 0$. Then, for a given (fixed) decision rule ϕ, the solution of the optimization problem in (3.6) is in the form of*

$$p_{\mathbf{s}_i}(\mathbf{y}) = \lambda_i \delta(\mathbf{y}-\mathbf{s}_{i1}) + (1-\lambda_i)\delta(\mathbf{y}-\mathbf{s}_{i2}) \ , \tag{3.7}$$

for $i = 0, 1$, *where* $\lambda_i \in [0, 1]$.

Proof: When the decision rule ϕ is given, $f(\phi\,; \mathbf{s}_i) = \int_{\Gamma_{\phi_{1-i}}} p_{\mathbf{n}}(\mathbf{y} - \mathbf{s}_i)\, d\mathbf{y}$ in (3.4) can be considered as a function of $\mathbf{s}_i$ only. In other words, $\mathrm{P}_{\mathrm{e},i}$ in (3.4) can be expressed as $\mathrm{P}_{\mathrm{e},i} = \mathrm{E}\{f(\mathbf{s}_i)\}$ for $i = 0, 1$. Since the objective function in (3.6) is the sum of $\pi_0 \mathrm{P}_{\mathrm{e},0}$ and $\pi_1 \mathrm{P}_{\mathrm{e},1}$, and the average power constraints are individually imposed on the signals, the optimization problem in (3.6) can be decoupled into two separate optimization problems as follows:

$$\min_{p_{\mathbf{s}_i}} \mathrm{E}\{f(\mathbf{s}_i)\}\ , \quad \text{subject to} \ \ \mathrm{E}\left\{|\mathbf{s}_i|^2\right\} \leq A\ , \tag{3.8}$$

for $i = 0, 1$. Optimization problems in the form of (3.8) have been investigated in various studies in the literature [38], [11]. Under the conditions in the lemma, the optimal solution of (3.8) can be represented by a randomization of at most two signal levels as a result of Carathéodory's theorem [39]. Hence, the optimal signal PDFs can be expressed as in (3.7). □

Note that the assumption in the lemma about the continuity of f in (3.4) is quite realistic for communications systems since the noise $\mathbf{n}$ in (3.1) has a continuous PDF in practice, as it is commonly the sum of zero-mean Gaussian thermal noise and interference terms that are independent of the thermal noise.

Lemma 3.1 states that, under certain conditions, the optimal stochastic signaling involves randomization among at most four different signal levels (two for symbol "0" and two for symbol "1"). Therefore, the problem in (3.6) can be solved over the signal PDFs that are in the form of (3.7). Hence, the search space for the optimization problem is reduced significantly. To achieve further simplification, the following result is obtained.

Proposition 3.1: *Under the conditions in Lemma 3.1, the optimization problem in* (3.6) *can be expressed as follows:*

$$\begin{aligned} \min_{\{\lambda_i, \mathbf{s}_{i1}, \mathbf{s}_{i2}\}_{i=0}^{1}} \quad & \int_{\mathbb{R}^K} \min\{\pi_0 g_0(\mathbf{y})\,,\, \pi_1 g_1(\mathbf{y})\}\, d\mathbf{y} \\ \text{subject to} \quad & \lambda_i |\mathbf{s}_{i1}|^2 + (1-\lambda_i)|\mathbf{s}_{i2}|^2 \leq A \\ & \lambda_i \in [0,1]\ , \quad i = 0,1 \end{aligned} \tag{3.9}$$

where $g_i(\mathbf{y}) = \lambda_i p_{\mathbf{n}}(\mathbf{y} - \mathbf{s}_{i1}) + (1-\lambda_i) p_{\mathbf{n}}(\mathbf{y} - \mathbf{s}_{i2})$.

Proof: For a given signal PDF pair $p_{\mathbf{s}_0}$ and $p_{\mathbf{s}_1}$, the conditional probability of observation $\mathbf{y}$ in (3.1) can be expressed as $p_i(\mathbf{y}) = \int_{\mathbb{R}^K} p_{\mathbf{s}_i}(\mathbf{x}) p_{\mathbf{n}}(\mathbf{y} - \mathbf{x}) d\mathbf{x}$ for $i = 0, 1$. When deciding between two symbols based on observation $\mathbf{y}$, the MAP decision rule, which selects symbol 1 if $\pi_1 p_1(\mathbf{y}) \geq \pi_0 p_0(\mathbf{y})$ and selects symbol 0 otherwise, minimizes the average probability of error [2]. Therefore, when signal PDFs $p_{\mathbf{s}_0}$ and $p_{\mathbf{s}_1}$ are specified, it is not necessary to search over all the decision rules; only the MAP decision rule should be determined and its corresponding average probability of error should be considered.

From (3.3), the average probability of error for any decision rule ϕ can be expressed as

$$\mathrm{P_e} = \int_{\Gamma_{\phi_1}} \pi_0 p_0(\mathbf{y})\, d\mathbf{y} + \int_{\Gamma_{\phi_0}} \pi_1 p_1(\mathbf{y})\, d\mathbf{y}\ . \tag{3.10}$$

Since the MAP decision rule decides symbol 1 if $\pi_1 p_1(\mathbf{y}) \geq \pi_0 p_0(\mathbf{y})$ and decides symbol 0 otherwise, the average probability of error expression in (3.10) can be expressed for a MAP decision rule, as [6]

$$\mathrm{P_e} = \int_{\mathbb{R}^K} \min\left\{\pi_0 p_0(\mathbf{y})\,,\, \pi_1 p_1(\mathbf{y})\right\}\, d\mathbf{y}\ . \tag{3.11}$$

Since Lemma 3.1 states that the optimal signal PDFs are in the form of (3.7), the conditional PDFs $p_i(\mathbf{y}) = \int_{\mathbb{R}^K} p_{\mathbf{s}_i}(\mathbf{x}) p_{\mathbf{n}}(\mathbf{y} - \mathbf{x}) d\mathbf{x}$ can be obtained as $p_i(\mathbf{y}) = \lambda_i p_{\mathbf{n}}(\mathbf{y} - \mathbf{s}_{i1}) + (1-\lambda_i) p_{\mathbf{n}}(\mathbf{y} - \mathbf{s}_{i2})$, and the average power constraints in (3.6) become $\lambda_i |\mathbf{s}_{i1}|^2 + (1-\lambda_i)|\mathbf{s}_{i2}|^2 \leq A$, for $i = 0, 1$. Therefore, (3.11) implies

that the optimization problem in (3.6) can be implemented as the constrained minimization problem in the proposition. □

Comparison of the optimization problems in (3.6) and (3.9) reveals that the latter is much simpler than the former since it is over a set of variables instead of a set of functions. However, it is still a non-convex optimization problem in general; hence, global optimization techniques, such as PSO [13], differential evolution and genetic algorithms [29], should be employed to obtain the optimal PDF. In this chapter, the PSO approach is used in the next section to obtain the solution of (3.9).

After obtaining the solution of the optimization problem in (3.9), the optimal signals are specified as $p_{\mathbf{s}_i}^{\text{opt}}(\mathbf{y}) = \lambda_i^{\text{opt}}\delta(\mathbf{y}-\mathbf{s}_{i1}^{\text{opt}})+(1-\lambda_i^{\text{opt}})\delta(\mathbf{y}-\mathbf{s}_{i2}^{\text{opt}})$ for $i = 0, 1$, and the optimal detector is the MAP decision rule that decides symbol 1 if $\pi_1 p_1(\mathbf{y}) \geq \pi_0 p_0(\mathbf{y})$ and decides symbol 0 otherwise.

Finally, it should be noted for symmetric signaling, that is, when $\mathbf{s}_{01} = -\mathbf{s}_{11}$, $\mathbf{s}_{02} = -\mathbf{s}_{12}$ and $\lambda_0 = \lambda_1$, the optimization in (3.9) can be performed over $\mathbf{s}_{11}$, $\mathbf{s}_{12}$ and λ_1 only.

3.2 Numerical Results and Conclusions

A numerical example is presented to illustrate the improvements that can be obtained via the joint design of the signaling structure and the decision rule for scalar observations. The noise in (3.1) is modeled by a Gaussian mixture as in [7] with its PDF being given by $p_n(y) = \frac{1}{\sqrt{2\pi}\,\sigma L}\sum_{i=1}^{L} \mathrm{e}^{-\frac{(y-\boldsymbol{\mu}_i)^2}{2\sigma^2}}$, where $L = 6$ and $\boldsymbol{\mu} = [0.27\ 0.81\ 1.08\ -1.08\ -0.81\ -0.27]$ are used. Note that the average power of the noise can be calculated as $\mathrm{E}\{n^2\} = \sigma^2 + 0.6318$. In addition, the average power limit in (3.5) is set to $A = 1$ and equally likely symbols are considered ($\pi_0 = \pi_1 = 0.5$).

In the following, three different approaches are compared.

Gaussian Solution: In this case, the transmitter is assumed to have no information about the noise PDF and selects the signals as $\mathbf{s}_0 = -\sqrt{A}$ and $\mathbf{s}_1 = \sqrt{A}$, which are known to be optimal in the presence of zero-mean Gaussian noise [2]. On the other hand, the MAP decision rule is used at the receiver.

Optimal – Stochastic: This approach refers to the solution of the most generic optimization problem in (3.6), which can also be obtained from (3.9) as studied in the previous section.

Optimal – Deterministic: This is a simplified version of the optimal solution in (3.9). It assumes that the signals are deterministic; i.e., they are not randomization of two different signal levels. Hence, the optimization problem in (3.9) becomes

$$\begin{aligned} &\min_{\mathbf{s}_0,\mathbf{s}_1} \int_{\mathbb{R}^K} \min\{\pi_0 p_{\mathbf{n}}(\mathbf{y}-\mathbf{s}_0)\,,\, \pi_1 p_{\mathbf{n}}(\mathbf{y}-\mathbf{s}_1)\}\, d\mathbf{y} \\ &\text{subject to } \ |\mathbf{s}_0|^2 \leq A \ , \ \ |\mathbf{s}_1|^2 \leq A \ . \end{aligned} \tag{3.12}$$

In other words, this approach provides the optimal solution when the signals are deterministic.

In Fig. 3.1, the average probabilities of error are plotted versus A/σ^2 for the three algorithms above by considering symmetric signaling. In obtaining the optimal stochastic solution from (3.9), the PSO algorithm is employed with 50 particles and 1000 iterations. Please refer to [13] for the details of the PSO algorithm[1]. On the other hand, the optimal deterministic solution in (3.12) can be obtained via a one-dimensional search due to symmetric signaling. From Fig. 3.1, it is observed that the Gaussian solution performs significantly worse than the optimal approaches for small σ values. In addition, the optimal approach based on stochastic signaling has the best performance. In other words, the

[1]The other parameters are set to $c_1 = c_2 = 2.05$ and $\chi = 0.72984$, and the inertia weight ω is changed from 1.2 to 0.1 linearly with the iteration number [13].

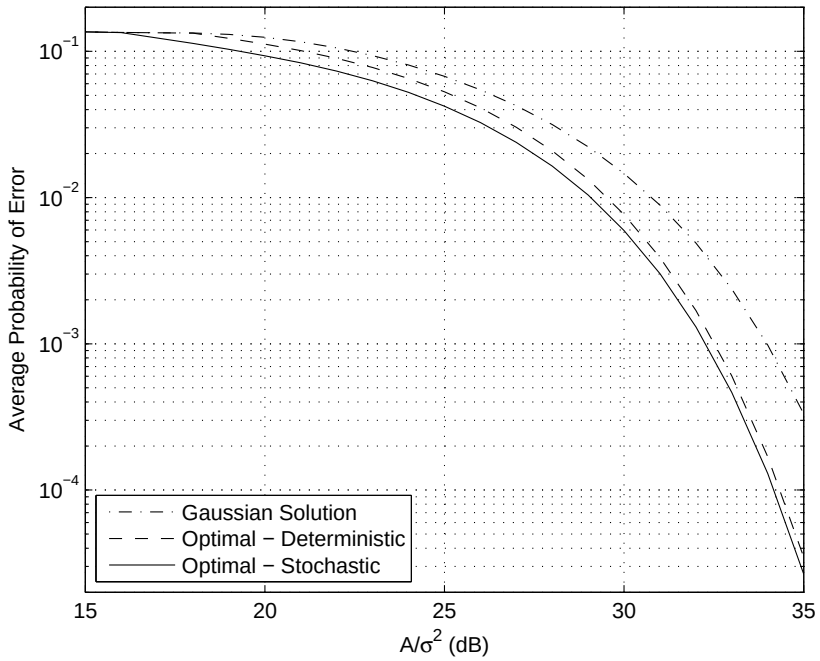

Figure 3.1: Average probability of error versus A/σ^2 for the three algorithms.

smallest average probability of error is obtained when each signal is modeled as stochastic signal that is a randomization of two signal values as in (3.7).

In order to explain the results in Fig. 3.1, Table 3.1 presents the solutions of the optimization problems in (3.6) and (3.12) for the optimal stochastic and the optimal deterministic approaches, respectively. Note that the results for symbol 1 are listed in Table 3.1, and the results for symbol 0 are the negatives of the signal values in the table since symmetric signaling is considered. For small A/σ^2 values, such as 15 dB, the optimal solutions are the same as the Gaussian solution, that is, $\mathbf{s}_{11} = \mathbf{s}_{12} = \mathbf{s}_1 = \sqrt{A} = 1$. However, for large A/σ^2's, the Gaussian solution becomes quite suboptimal and choosing the largest possible deterministic signal value, 1, results in higher average probabilities of error, as can be observed from Fig. 3.1. For example, at $A/\sigma^2 = 30$ dB, the optimal deterministic solution sets $\mathbf{s}_1 = -\mathbf{s}_0 = 0.7476$ and achieves an error rate of 7.66×10^{-3}, whereas the Gaussian one uses $\mathbf{s}_1 = -\mathbf{s}_0 = 1$, which yields an error rate of 0.0146. This seemingly counterintuitive result is obtained since the average probability of error is related to the area under the overlaps of the two shifted

Table 3.1: Optimal stochastic and deterministic signals for symbol 1.

	Stochastic			Deterministic
A/σ^2 (dB)	λ_1	$\mathbf{s}_{11}$	$\mathbf{s}_{12}$	$\mathbf{s}_1$
15	N/A	1	1	1
20	0.1836	1.648	0.7846	0.7927
25	0.2104	1.614	0.7576	0.7587
30	0.2260	1.586	0.7475	0.7476
35	0.2347	1.568	0.7441	0.8759

noise PDFs as in (3.12). Although optimal deterministic signaling uses less power than permitted, it results in a lower error probability than Gaussian signaling by avoiding the overlaps between the components of the Gaussian mixture noise more effectively. On the other hand, optimal stochastic signaling further reduces the average probability of error by using all the available power and assigning some of the power to a large signal component that results in less overlapping between the shifted noise PDFs. For example, at $A/\sigma^2 = 30$ dB, the optimal stochastic signal is a randomization of $\mathbf{s}_{11} = -\mathbf{s}_{01} = 1.586$ and $\mathbf{s}_{12} = -\mathbf{s}_{02} = 0.7475$ with $\lambda_0 = \lambda_1 = 0.226$ (cf. (3.7)), which achieves an error rate of 5.95×10^{-3}.

The results in this chapter can be extended to M-ary communications systems as well by noting that the average probability of error expression in (3.11) becomes $\mathrm{P_e} = 1 - \int \max\{\pi_0 p_0(\mathbf{y}), \ldots, \pi_{M-1} p_{M-1}(\mathbf{y})\} d\mathbf{y}$ for M-ary systems. Then, an optimization problem similar to that in Proposition 3.1 can be obtained, where the optimization is performed over $\{\lambda_i, \mathbf{s}_{i1}, \mathbf{s}_{i2}\}_{i=0}^{M-1}$.

Chapter 4

STOCHASTIC SIGNALING UNDER CHANNEL STATE INFORMATION UNCERTAINTIES

In this chapter, stochastic signaling is studied for power-constrained scalar valued binary communications systems in the presence of uncertainties in channel state information (CSI). First, stochastic signaling based on the available imperfect channel coefficient at the transmitter is discussed, and it is shown that optimal signals can be represented by randomization between at most two different signal levels for each symbol. Then, performance of stochastic signaling and conventional deterministic signaling is compared for this scenario, and sufficient conditions are derived for improvability and nonimprovability of deterministic signaling via stochastic signaling in the presence of CSI uncertainty. Furthermore, under CSI uncertainty, two different stochastic signaling strategies, namely, robust stochastic signaling and stochastic signaling with averaging, are

proposed. For robust stochastic signaling problem, sufficient conditions are derived for reducing the problem to a simpler form. It is shown that optimal signals for each symbol can be expressed as randomization between at most two signal values for stochastic signaling with averaging, as well as for robust stochastic signaling under certain conditions. Finally two numerical examples are presented to explore the theoretical results.

4.1 System Model and Motivation

Consider a binary communications system with scalar observations [6] in which the channel effect can be modeled by a multiplicative term as in flat-fading channels [1], and the received signal is given by

$$Y = \alpha\, S_i + N \ , \qquad i \in \{0, 1\} \ , \tag{4.1}$$

where S_0 and S_1 denote the transmitted signal values for symbol 0 and symbol 1 respectively, N is the noise component that is independent of S_i, and α is the channel coefficient. In addition, the prior probabilities of the symbols, which are denoted by π_0 and π_1, are supposed to be known.

In (4.1), the noise term N is modeled to have an arbitrary probability distribution considering that it can include the combined effects of thermal noise, interference, and jamming. Hence, the probability distribution of the noise component is not necessarily Gaussian [7].

A generic decision rule is considered at the receiver to determine the symbol in (4.1). For a given observation $Y = y$, the decision rule $\phi(y)$ is expressed as

$$\phi(y) = \begin{cases} 0 \ , & y \in \Gamma_0 \\ 1 \ , & y \in \Gamma_1 \end{cases} \ , \tag{4.2}$$

where Γ_0 and Γ_1 are the decision regions for symbol 0 and symbol 1, respectively [2].

The aim is to design signals S_0 and S_1 in (4.1) in order to minimize the average probability of error for a given decision rule, which is given by

$$\mathrm{P}_{\mathrm{avg}} = \pi_0 \mathrm{P}_0(\Gamma_1) + \pi_1 \mathrm{P}_1(\Gamma_0) \ , \tag{4.3}$$

with $\mathrm{P}_i(\Gamma_j)$ denoting the probability of selecting symbol j when symbol i is transmitted. In practical systems, there exists an average power constraint on the signals, which can be expressed as

$$\mathrm{E}\{|S_i|^2\} \leq A \ , \tag{4.4}$$

for $i = 0, 1$, where A is the average power limit. Therefore, in the stochastic signaling approach, the aim becomes the calculation of the optimal probability density functions (PDFs) for signals S_0 and S_1 that minimize the average probability of error in (4.3) under the average power constraint in (4.4) [17].

Unlike stochastic signaling, in the conventional signal design, S_0 and S_1 are modeled as deterministic signals and set to $S_0 = -\sqrt{A}$ and $S_1 = \sqrt{A}$ [1], [2]. Then, the average probability of error in (4.3) becomes

$$\begin{aligned} \mathrm{P}_{\mathrm{conv}} = \pi_0 \int_{\Gamma_1} p_N \left(y + \alpha\sqrt{A}\right) dy \\ + \pi_1 \int_{\Gamma_0} p_N \left(y - \alpha\sqrt{A}\right) dy \ , \end{aligned} \tag{4.5}$$

where $p_N(\cdot)$ is the PDF of the noise in (4.1).

As investigated in [17], [40], [42] stochastic signaling results in lower average probabilities of error than the conventional deterministic signaling in some cases in the presence of non-Gaussian noise. However, the common assumption in the previous studies is that the channel coefficient α in (4.1) is known perfectly at the transmitter, i.e., the CSI is available at the transmitter. In practice, the transmitter can obtain CSI via feedback from the receiver, or by utilizing the reciprocity of forward and reverse links under time division duplexing [41]. In both scenarios, it is realistic to model the CSI at the transmitter to include certain errors/uncertainties. Therefore, the main motivation behind this study

is to investigate stochastic signaling under imperfect CSI; that is, to evaluate the performance of stochastic signaling in practical scenarios and to develop different design methods for stochastic signaling under CSI uncertainty. In the next section, the effects of CSI uncertainties on the performance of stochastic signaling are examined.

4.2 Effects of Channel Uncertainties on the Stochastic Signaling

4.2.1 Stochastic Signaling with Imperfect Channel Coefficient

Let $p_{S_0}(\cdot)$ and $p_{S_1}(\cdot)$ denote the PDFs of S_0 and S_1 in (4.1), respectively. Also define $\hat{S}_0 \triangleq \alpha\, S_0$ and $\hat{S}_1 \triangleq \alpha\, S_1$, and denote their PDFs as $p_{\hat{S}_0}(\cdot)$ and $p_{\hat{S}_1}(\cdot)$, respectively. Then, from (4.3), the average probability of error for the decision rule in (4.2) is given by

$$\mathrm{P}_{\mathrm{stoc}} = \sum_{i=0}^{1} \pi_i \int_{-\infty}^{\infty} p_{\hat{S}_i}(t) \int_{\Gamma_{1-i}} p_N(y-t)\, dy\, dt \ . \tag{4.6}$$

Since $p_{\hat{S}_i}(t)$ is given by $p_{\hat{S}_i}(t) = (1/|\alpha|)\, p_{S_i}(1/\alpha)$ for $i = 0, 1$, (4.6) can also be expressed, after a change of variable ($t = \alpha\, x$), as

$$\mathrm{P}_{\mathrm{stoc}} = \sum_{i=0}^{1} \pi_i \int_{-\infty}^{\infty} p_{S_i}(x) \int_{\Gamma_{1-i}} p_N(y-\alpha\, x)\, dy\, dx \ . \tag{4.7}$$

Since imperfect CSI is considered in this study, the transmitter has a distorted version of the correct channel coefficient α. Let $\hat{\alpha}$ denote this distorted (noisy) channel coefficient at the transmitter. In this section, it is assumed that the transmitter uses $\hat{\alpha}$ in the design of stochastic signals. Then, the stochastic signal

design problem can be expressed as

$$\min_{p_{S_0}, p_{S_1}} \sum_{i=0}^{1} \pi_i \int_{-\infty}^{\infty} p_{S_i}(x) \int_{\Gamma_{1-i}} p_N(y - \hat{\alpha}\, x)\, dy\, dx$$
$$\text{subject to } \ \mathrm{E}\{|S_i|^2\} \leq A \ , \quad i = 0, 1. \tag{4.8}$$

Note that there are also implicit constraints in the optimization problem in (4.8) because $p_{S_0}(\cdot)$ and $p_{S_1}(\cdot)$ need to satisfy the conditions to be valid PDFs. As in [17], this optimization problem can be expressed as two separate optimization problems for S_0 and S_1. Namely, the optimal signal PDF for symbol 1 can be obtained from the solution of the following optimization problem:

$$\min_{p_{S_1}} \int_{-\infty}^{\infty} p_{S_1}(x) \int_{\Gamma_0} p_N(y - \hat{\alpha}\, x)\, dy\, dx$$
$$\text{subject to } \ \mathrm{E}\{|S_1|^2\} \leq A \ . \tag{4.9}$$

If $G(x, k)$ is defined as

$$G(x, k) \triangleq \int_{\Gamma_0} p_N(y - k\, x)\, dy \ , \tag{4.10}$$

(4.9) can also be written as

$$\min_{p_{S_1}} \mathrm{E}\{G(S_1, \hat{\alpha})\} \ \text{ subject to } \ \mathrm{E}\{|S_1|^2\} \leq A \ , \tag{4.11}$$

where the expectations are taken over S_1. Note that, $G(S_1, \hat{\alpha})$ is only a function of S_1 for a given fixed $\hat{\alpha}$. In the previous studies, such as [17] and [11], the optimization problems with the same structure as (4.11) have been explored thoroughly. If $G(S_1, \hat{\alpha})$ in (4.11) is a continuous function of S_1 and S_1 takes values in $[-\gamma, \gamma]$ for some finite $\gamma > 0$, then the optimal solution of (4.11) can be represented by a randomization of at most two signal levels as a result of Carathéodory's theorem [39]. Hence, the optimal signal PDF for S_1 can be expressed as

$$p_{S_1}(s) = \lambda_1\, \delta(s - s_{11}) + (1 - \lambda_1)\, \delta(s - s_{12}) \ , \tag{4.12}$$

where $\lambda_1 \in [0, 1]$.

A similar optimization problem can also be formulated for S_0. After obtaining the optimal signal PDFs for S_0 and S_1, the corresponding average probability of error can be calculated. Since the optimization problems are similar for S_0 and S_1, we focus on the design of S_1 in the remainder of this section.

4.2.2 Stochastic Signaling versus Conventional Signaling

It is known that, in the presence of perfect CSI at the transmitter, conventional signaling, which sets $S_1 = \sqrt{A}$ [that is, $p_{S_1}(x) = \delta(x - \sqrt{A})$], can or cannot be optimal under certain sufficient conditions as discussed in [17]. In this section, we explore the conditions under which the use of stochastic signaling instead of deterministic signaling can result in improved average probability of error performance in the presence of *imperfect* CSI.

In the presence of imperfect CSI, let the transmitter have the channel coefficient information as $\hat{\alpha}$. Then, the transmitter obtains the optimal stochastic signal S_1 from (4.11). Let $p_{S_1}^{\hat{\alpha}}(\cdot)$ denote the solution of (4.11) for a given value of $\hat{\alpha}$. Then, the corresponding conditional probability of error for symbol 1 can be expressed as

$$\mathrm{P}_{\mathrm{e}}^{\hat{\alpha}} = \int_{-\infty}^{\infty} p_{S_1}^{\hat{\alpha}}(x)\, G(x, \alpha)\, dx \ , \tag{4.13}$$

where $G(x, \alpha)$ is as defined in (4.10). Note that $G(x, \alpha)$ specifies the probability of choosing symbol 0 for a given signal value x for symbol 1 when the channel coefficient is equal to α. Therefore, when the stochastic signal for symbol 1 is specified by the PDF $p_{S_1}^{\hat{\alpha}}(x)$, the corresponding conditional probability of error for symbol 1 is obtained as in (4.13).

Suppose that $\hat{\alpha}$ can be modeled as a random variable with a generic PDF $p_{\hat{\alpha}}(\cdot)$. In order to improve the performance of conventional signaling for symbol 1 via stochastic signaling, we need to have $\mathrm{P}_{\mathrm{e}} < G(\sqrt{A}, \alpha)$, where $G(\sqrt{A}, \alpha)$ is the conditional probability of error for conventional signaling, i.e., for $S_1 = \sqrt{A}$

(see (4.5) and (4.10)), and $\mathrm{P_e}$ is the average conditional probability of error for stochastic signaling based on imperfect CSI, which can be calculated as

$$\mathrm{P_e} = \int_{-\infty}^{\infty} p_{\hat{\alpha}}(a)\, \mathrm{P_e^a}\, da \ , \tag{4.14}$$

with $\mathrm{P_e^a}$ being given by (4.13).

In order to derive sufficient conditions for the improvability and nonimprovability of conventional signaling via stochastic signaling, assume that the channel coefficient information at the transmitter is specified as $\hat{\alpha} = \alpha + \eta$, where η is a zero-mean Gaussian noise with standard deviation ε; that is, $\eta \sim \mathcal{N}(0, \varepsilon^2)$. Although the Gaussian error model is employed for the convenience of the analysis, *the results are valid also for non-Gaussian error models*, as will be discussed at the end of this section. In addition, it is assumed that α is a positive number without loss of generality.[1]. Then, the following proposition presents sufficient conditions on the improvability and nonimprovability of conventional signaling via stochastic signaling.

Proposition 4.1: *Stochastic signaling performs worse than conventional signaling if the standard deviation of the channel coefficient error ε is greater than or equal to a threshold ε^* and it performs better than conventional signaling if ε is less than or equal to another threshold $\hat{\varepsilon}$ when $G(x,k)$ and $\mathrm{P}_e^{\hat{\alpha}}$ have the following properties:*

- *$G(x,k)$ is a strictly decreasing function of x for any fixed positive k, and $G(x,k) = 1 - G(-x,k)$.*
- *$\mathrm{P}_e^{\hat{\alpha}} < \kappa_1$ when $\hat{\alpha} > \gamma_{th} > 0$, $\mathrm{P}_e^{\hat{\alpha}} < \kappa_2 < \kappa_1$ when $\alpha > \hat{\alpha} > \theta_{th} > \gamma_{th}$, and $\mathrm{P}_e^{\hat{\alpha}} = G(\sqrt{A}, \alpha)$ when $\hat{\alpha} > \beta_{th} > \alpha$.*

[1]If it is negative, one can redefine function G (4.10) by using $p_N(y+kx)$ instead of $p_N(y-kx)$

In addition, ε^* and $\hat{\varepsilon}$ can be obtained by solving[2]

$$\left(\frac{1}{2}-\kappa_1\right)Q\left(\frac{\alpha+\gamma_{th}}{\varepsilon^*}\right)+\frac{1}{2}Q\left(\frac{\alpha}{\varepsilon^*}\right)+(\kappa_1-\kappa_2)\left(Q\left(\frac{2\alpha}{\varepsilon^*}\right)-Q\left(\frac{\alpha+\theta_{th}}{\varepsilon^*}\right)\right)$$
$$=\left(1-Q\left(\frac{\beta_{th}-\alpha}{\varepsilon^*}\right)\right)G(\sqrt{A},\alpha) \tag{4.15}$$

and

$$\frac{1}{2}\left(\kappa_1+\kappa_2+Q\left(\frac{\alpha}{\hat{\varepsilon}}\right)\right)+\left(\frac{1}{2}-\kappa_1\right)Q\left(\frac{\alpha-\gamma_{th}}{\hat{\varepsilon}}\right)-\kappa_1 Q\left(\frac{\beta_{th}-\alpha}{\hat{\varepsilon}}\right)$$
$$+(\kappa_1-\kappa_2)Q\left(\frac{\alpha-\theta_{th}}{\hat{\varepsilon}}\right)=\left(1-Q\left(\frac{\beta_{th}-\alpha}{\hat{\varepsilon}}\right)+Q\left(\frac{\alpha+\beta_{th}}{\hat{\varepsilon}}\right)\right)G(\sqrt{A},\alpha)\ , \tag{4.16}$$

respectively.

Proof: In the following, lower and upper bounds for the expression in (4.14) are derived in order to prove the statements in the proposition. We start by noticing the fact that the sign of the channel coefficient knowledge at the transmitter is important. Suppose that $p_{S_1}^{\hat{\alpha}}$ is the optimal PDF obtained from (4.11) for a given $\hat{\alpha}$. Therefore, if $-\hat{\alpha}$ is used instead of $\hat{\alpha}$, then $p_{S_1}^{-\hat{\alpha}}$ will be the optimal solution of (4.11) and the value of $p_{S_1}^{-\hat{\alpha}}(x)$ will be equal to $p_{S_1}^{\hat{\alpha}}(-x)$. This observation can be utilized in (4.13), and also using the fact that $G(x,k)=1-G(-x,k)$, $\mathrm{P}_{\mathrm{e}}^{\hat{\alpha}}=1-\mathrm{P}_{\mathrm{e}}^{-\hat{\alpha}}$ can be obtained as follows:

$$\int_{-\infty}^{\infty}p_{S_1}^{\hat{\alpha}}(x)G(x,k)dx=\int_{-\infty}^{\infty}p_{S_1}^{-\hat{\alpha}}(-x)(1-G(-x,k))dx=\int_{-\infty}^{\infty}p_{S_1}^{-\hat{\alpha}}(t)(1-G(t,k))dt$$
$$=1-\int_{-\infty}^{\infty}p_{S_1}^{-\hat{\alpha}}(t)G(t,k)dt=1-\mathrm{P}_{\mathrm{e}}^{-\hat{\alpha}}. \tag{4.17}$$

It is stated in the second condition of the proposition that $\mathrm{P}_{\mathrm{e}}^{\hat{\alpha}}<\kappa_1$ when $\hat{\alpha}>\gamma_{th}$, and $\mathrm{P}_{\mathrm{e}}^{\hat{\alpha}}<\kappa_2<\kappa_1$ when $\alpha>\hat{\alpha}>\theta_{th}$. Therefore, if we insert $-\hat{\alpha}$ instead of $\hat{\alpha}$ in these conditions, we get $\mathrm{P}_{\mathrm{e}}^{-\hat{\alpha}}<\kappa_1$ when $-\hat{\alpha}>\gamma_{th}$ and $\mathrm{P}_{\mathrm{e}}^{-\hat{\alpha}}<\kappa_2<\kappa_1$ when $\alpha>-\hat{\alpha}>\theta_{th}$. Using the result in (4.17) and rearranging the terms yield

[2]Note that the choice of parameters in the conditions of Proposition 4.1 is important to ensure the existence of solutions to (4.15) and (4.16). Also, the Q-function is defined as $Q(x)=(\int_x^{\infty}\mathrm{e}^{-t^2/2}dt)/\sqrt{2\pi}$.

$\mathrm{P}_{\mathrm{e}}^{\hat{\alpha}} > 1-\kappa_1$ when $\hat{\alpha} < -\gamma_{th}$ and $\mathrm{P}_{\mathrm{e}}^{\hat{\alpha}} > 1-\kappa_2 > 1-\kappa_1$ when $-\alpha < \hat{\alpha} < -\theta_{th}$. Also, since $G(x,k)$ is a strictly decreasing function of x when k is positive, then $G(x,\hat{\alpha})$ is a strictly increasing function of x if $\hat{\alpha}$ is negative. Therefore, for a given $\hat{\alpha} < 0$, the optimal signal PDF $p_{S_1}^{\hat{\alpha}}$ assigns the weights on negative numbers instead of positive ones since for each positive value of S_1, its negative can be used instead, which results in the same average power value and a smaller $\mathrm{E}\{G(S_1,\hat{\alpha})\}$. Furthermore, since $G(x,\alpha)$ is a strictly decreasing function, and $G(x,\alpha) = 1 - G(-x,\alpha)$, we have $G(x,\alpha) > G(0,\alpha) = 0.5$ for $x < 0$. Thus, by using these two facts and the expression in (4.13), we conclude that if $\hat{\alpha} < 0$, then $\mathrm{P}_{\mathrm{e}}^{\hat{\alpha}} > 0.5$ [and $\mathrm{P}_{\mathrm{e}}^{\hat{\alpha}} < 0.5$, if $\hat{\alpha} > 0$]. Now, one can find a lower bound on P_{e} in (4.14) as follows:

$$
\begin{aligned}
&\mathrm{P}_{\mathrm{e}} = \int_{-\infty}^{\infty} p_{\hat{\alpha}}(a)\mathrm{P}_{\mathrm{e}}^{a} da \geq \int_{-\infty}^{-\gamma_{th}} p_{\hat{\alpha}}(a)\mathrm{P}_{\mathrm{e}}^{a} da + \int_{-\gamma_{th}}^{0} p_{\hat{\alpha}}(a)\mathrm{P}_{\mathrm{e}}^{a} da + \int_{\beta_{th}}^{\infty} p_{\hat{\alpha}}(a)\mathrm{P}_{\mathrm{e}}^{a} da \\
&> (1-\kappa_1)\mathrm{P}(\hat{\alpha} < -\gamma_{th}) + (\kappa_1-\kappa_2)\mathrm{P}(-\alpha < \hat{\alpha} < -\theta_{th}) + \frac{1}{2}\mathrm{P}(-\gamma_{th} < \hat{\alpha} < 0) \\
&+ \mathrm{P}(\beta_{th} < \hat{\alpha})G(\sqrt{A},\alpha) = (1-\kappa_1)\mathrm{P}\left(\frac{\eta}{\varepsilon} > \frac{\alpha+\gamma_{th}}{\varepsilon}\right) \\
&+ (\kappa_1-\kappa_2)\mathrm{P}\left(\frac{-2\alpha}{\varepsilon} < \frac{\eta}{\varepsilon} < \frac{-\alpha-\theta_{th}}{\varepsilon}\right) + \frac{1}{2}\mathrm{P}\left(\frac{-\alpha}{\varepsilon} < \frac{\eta}{\varepsilon} < \frac{-\alpha-\gamma_{th}}{\varepsilon}\right) \\
&+ \mathrm{P}\left(\frac{\eta}{\varepsilon} > \frac{\beta_{th}-\alpha}{\varepsilon}\right) G(\sqrt{A},\alpha) = (1-\kappa_1)Q\left(\frac{\alpha+\gamma_{th}}{\varepsilon}\right) + \\
&(\kappa_1-\kappa_2)\left(Q\left(\frac{2\alpha}{\varepsilon}\right) - Q\left(\frac{\alpha+\theta_{th}}{\varepsilon}\right)\right) + \frac{1}{2}\left(Q\left(\frac{\alpha}{\varepsilon}\right) - Q\left(\frac{\alpha+\gamma_{th}}{\varepsilon}\right)\right) \\
&+ Q\left(\frac{\beta_{th}-\alpha}{\varepsilon}\right) G(\sqrt{A},\alpha) = \left(\frac{1}{2}-\kappa_1\right) Q\left(\frac{\alpha+\gamma_{th}}{\varepsilon}\right) \\
&+ (\kappa_1-\kappa_2)\left(Q\left(\frac{2\alpha}{\varepsilon}\right) - Q\left(\frac{\alpha+\theta_{th}}{\varepsilon}\right)\right) + \frac{1}{2}Q\left(\frac{\alpha}{\varepsilon}\right) + Q\left(\frac{\beta_{th}-\alpha}{\varepsilon}\right) G(\sqrt{A},\alpha)\ .
\end{aligned}
\tag{4.18}
$$

If we equate this bound to $G(\sqrt{A},\alpha)$ and solve for ε, we obtain ε^*. Therefore, if $\varepsilon = \varepsilon^*$, we have $\mathrm{P}_{\mathrm{e}} > G(\sqrt{A},\alpha)$. Notice that the Q-function is strictly decreasing, hence the derived lower bound is an increasing function of ε. Thus, for $\varepsilon > \varepsilon^*$, we still have $\mathrm{P}_{\mathrm{e}} > G(\sqrt{A},\alpha)$. Overall, under the conditions given in the proposition, having the standard deviation of the channel coefficient error being larger than

or equal to a certain value ε^* is sufficient to conclude that conventional signaling performs better than stochastic signaling.

Next, one can find an upper bound on $\mathrm{P_e}$ in (4.14) as follows:

$$
\begin{aligned}
\mathrm{P_e} &= \int_{-\infty}^{\infty} p_{\hat{\alpha}}(a)\mathrm{P}_\mathrm{e}^a da = \int_{-\infty}^{-\beta_{th}} p_{\hat{\alpha}}(a)\mathrm{P}_\mathrm{e}^a da + \int_{-\beta_{th}}^{0} p_{\hat{\alpha}}(a)\mathrm{P}_\mathrm{e}^a da + \int_{0}^{\gamma_{th}} p_{\hat{\alpha}}(a)\mathrm{P}_\mathrm{e}^a da \\
&+ \int_{\gamma_{th}}^{\beta_{th}} p_{\hat{\alpha}}(a)\mathrm{P}_\mathrm{e}^a da + \int_{\beta_{th}}^{\infty} p_{\hat{\alpha}}(a)\mathrm{P}_\mathrm{e}^a da \leq (1 - G(\sqrt{A}, \alpha))\mathrm{P}(\hat{\alpha} < -\beta_{th}) \\
&+ \mathrm{P}(-\beta_{th} < \hat{\alpha} < 0) + \frac{1}{2}\mathrm{P}(0 < \hat{\alpha} < \gamma_{th}) + \kappa_1 \mathrm{P}(\gamma_{th} < \hat{\alpha} < \beta_{th}) \\
&+ (\kappa_2 - \kappa_1)\mathrm{P}(\theta_{th} < \hat{\alpha} < \alpha) + \mathrm{P}(\hat{\alpha} > \beta_{th}) G(\sqrt{A}, \alpha) \\
&= (1 - G(\sqrt{A}, \alpha))\mathrm{P}\left(\frac{\eta}{\varepsilon} > \frac{\alpha + \beta_{th}}{\varepsilon}\right) + \mathrm{P}\left(\frac{-\beta_{th} - \alpha}{\varepsilon} < \frac{\eta}{\varepsilon} < \frac{-\alpha}{\varepsilon}\right) \\
&+ \frac{1}{2}\mathrm{P}\left(\frac{-\alpha}{\varepsilon} < \frac{\eta}{\varepsilon} < \frac{-\alpha + \gamma_{th}}{\varepsilon}\right) + \kappa_1 \mathrm{P}\Big(\frac{-\alpha + \gamma_{th}}{\varepsilon} < \frac{\eta}{\varepsilon} < \frac{-\alpha + \beta_{th}}{\varepsilon}\Big) \\
&+ (\kappa_2 - \kappa_1)\mathrm{P}\left(\frac{-\alpha + \theta_{th}}{\varepsilon} < \frac{\eta}{\varepsilon} < 0\right) + \mathrm{P}\left(\frac{\eta}{\varepsilon} > \frac{\beta_{th} - \alpha}{\varepsilon}\right) G(\sqrt{A}, \alpha) \\
&= (1 - G(\sqrt{A}, \alpha))Q\left(\frac{\alpha + \beta_{th}}{\varepsilon}\right) + \frac{1}{2}\Big(Q\left(\frac{\alpha - \gamma_{th}}{\varepsilon}\right) - Q\left(\frac{\alpha}{\varepsilon}\right)\Big) + Q\left(\frac{\alpha}{\varepsilon}\right) \\
&- Q\left(\frac{\alpha + \beta_{th}}{\varepsilon}\right) + \kappa_1\left(1 - Q\left(\frac{\alpha - \gamma_{th}}{\varepsilon}\right) - Q\left(\frac{\beta_{th} - \alpha}{\varepsilon}\right)\right) \\
&+ (\kappa_2 - \kappa_1)\left(\frac{1}{2} - Q\left(\frac{\alpha - \theta_{th}}{\varepsilon}\right)\right) + Q\left(\frac{\beta_{th} - \alpha}{\varepsilon}\right) G(\sqrt{A}, \alpha) = \frac{1}{2}\left(\kappa_1 + \kappa_2 + Q\left(\frac{\alpha}{\varepsilon}\right)\right) \\
&+ \left(\frac{1}{2} - \kappa_1\right) Q\left(\frac{\alpha - \gamma_{th}}{\varepsilon}\right) - \kappa_1 Q\left(\frac{\beta_{th} - \alpha}{\varepsilon}\right) + (\kappa_1 - \kappa_2) Q\left(\frac{\alpha - \theta_{th}}{\varepsilon}\right) \\
&+ \left(Q\left(\frac{\beta_{th} - \alpha}{\varepsilon}\right) - Q\left(\frac{\alpha + \beta_{th}}{\varepsilon}\right)\right) G(\sqrt{A}, \alpha)\ . \qquad (4.19)
\end{aligned}
$$

Therefore, if we equate this bound to $G(\sqrt{A}, \alpha)$ and solve for ε, we obtain $\hat{\varepsilon}$. Therefore, if $\varepsilon = \hat{\varepsilon}$, we have $\mathrm{P_e} < G(\sqrt{A}, \alpha)$. Since the derived upper bound decreases as ε decreases, for $\varepsilon < \hat{\varepsilon}$, we still have $\mathrm{P_e} < G(\sqrt{A}, \alpha)$. Overall, under the conditions given in the proposition, having $\varepsilon \leq \hat{\varepsilon}$ is sufficient to conclude that stochastic signaling performs better than conventional signaling. □

Although the results in Proposition 4.1 are presented for channel coefficient errors with a zero-mean Gaussian distribution, they can easily be extended for any type of probability distribution as well. For example, consider a generic

PDF for the channel coefficient error, which is denoted by $p_\eta(\cdot)$. The corresponding cumulative distribution function (CDF) $F_\eta(\cdot)$ can be expressed as $F_\eta(x) = \int_{-\infty}^{x} p_\eta(t)\, dt$. Then, the results in Proposition 4.1 are valid when $Q(x/\varepsilon^*)$ in (4.15) and $Q(x/\hat{\varepsilon})$ in (4.16) are replaced by $1 - F_\eta(x)$. Hence, ε^* and $\hat{\varepsilon}$ can still be obtained by solving the updated equations.

As discussed before, $G(x, k)$ can be inferred as the probability of deciding symbol 0 instead of symbol 1, when the value of the channel coefficient is k, and $S_1 = x$. In general, for a specific channel coefficient, when a larger signal value is employed, a lower probability of error can be obtained; hence, $G(x, k)$ is usually a decreasing function of x in practice. Moreover, $G(x, k) = 1 - G(-x, k)$ can be satisfied when the channel noise has a symmetric PDF (i.e. $p_N(x) = p_N(-x)$) and the decision regions of the detector at the receiver are symmetric ($\Gamma_0 = -\Gamma_1$). In fact, the channel noise is symmetric in most practical scenarios (for example, zero-mean additive white Gaussian noise or Gaussian mixture noise with symmetric components [7]), and some receivers such as the sign detector or the optimal MAP detector for symmetric signaling under symmetric channel noise will have symmetric decision regions in fact. All in all, the first condition in the proposition is expected to hold in many practical scenarios. The details of how the second condition is satisfied and how the parameters are selected will be investigated in the Section 4.4.

4.3 Design of Stochastic Signals Under CSI Uncertainty

First, suppose that $p_\alpha(\cdot)$ denotes the PDF of the actual channel coefficient α, where each instance of the channel coefficient resides in a certain set Ω. In this section, we propose two different methods for designing the stochastic signals

under CSI uncertainty in the transmitter, and evaluate the performance of each method in Section 4.4.

4.3.1 Robust Stochastic Signaling

In this part, robust design of optimal stochastic signals is presented under CSI uncertainty at the transmitter. Suppose that Ω is given by $\Omega = [\alpha_0, \alpha_1]$, that is, the channel coefficient α takes values in the interval of $[\alpha_0, \alpha_1]$, where $\alpha_0 < \alpha_1$. It is assumed that the transmitter has the knowledge of set Ω. Note that this can be realized via feedback from the receiver to the transmitter. In robust stochastic signaling, signals are designed in such a way that they minimize the average probability of error for the worst-case channel coefficient, that is, the one which maximizes the average probability of error for the transmitted signals. For this design criterion, the optimal stochastic signaling problem in (4.8) can be expressed as a minimax problem as follows:

$$\begin{aligned} &\min_{p_{S_0}, p_{S_1}} \max_{\alpha \in [\alpha_0, \alpha_1]} \sum_{i=0}^{1} \pi_i \int_{-\infty}^{\infty} p_{S_i}(x) \int_{\Gamma_{1-i}} p_N(y - \alpha\, x)\, dy\, dx \\ &\text{subject to } \ \mathrm{E}\{|S_i|^2\} \leq A \ . \end{aligned} \tag{4.20}$$

The problem in (4.20) might be difficult to solve in general. In the following, it is shown that in most practical scenarios, this problem can be reduced to a simpler form and the optimal signal PDFs can be obtained by solving a simpler optimization problem:

Proposition 4.2: *The minimax problem in* (4.20) *is equivalent to the stochastic signaling problem for channel coefficient* α_0*, that is,*

$$\begin{aligned} &\min_{p_{S_0}, p_{S_1}} \sum_{i=0}^{1} \pi_i \int_{-\infty}^{\infty} p_{S_i}(x) \int_{\Gamma_{1-i}} p_N(y - \alpha_0\, x)\, dy\, dx \\ &\textit{subject to } \ \mathrm{E}\{|S_i|^2\} \leq A \end{aligned} \tag{4.21}$$

when the following conditions are satisfied:

- *$G(x,\alpha)$ is a strictly decreasing function of x for any $\alpha \in [\alpha_0\ \alpha_1]$.*

- *$G(x,\alpha)$ is a strictly decreasing (increasing) function of α for all $x > 0$ ($x < 0$).*

Proof: The minimax problem in (4.20) can be expressed as follows:

$$\min_{p_{S_0},p_{S_1}} \max_{\alpha\in[\alpha_0,\alpha_1]} \pi_1 \int_{-\infty}^{\infty} p_{S_1}(x)\, G(x,\alpha)dx + \pi_0 \int_{-\infty}^{\infty} p_{S_0}(x)\,(1-G(x,\alpha))dx$$
$$\text{subject to } \ \mathrm{E}\{|S_i|^2\} \leq A\ . \tag{4.22}$$

Assume that S_1 is a nonnegative and S_0 is a nonpositive random variable. First, it is shown that this assumption does not reduce the generality of the proof. Suppose that $p^*_{S_1}$ is the PDF of S_1 which is a nonnnegative random variable, and $p^*_{S_0}$ is the PDF of S_0 which is any random variable (that is, its instances can take both positive or negative values). Therefore, in the minimax problem, for given $p^*_{S_0}$ and $p^*_{S_1}$, we maximize $\pi_1 \int_{-\infty}^{\infty} p^*_{S_1}(x)\, G(x,\alpha)dx + \pi_0 \int_{-\infty}^{\infty} p^*_{S_0}(x)\,(1-G(x,\alpha))dx$ over $\alpha \in [\alpha_0,\alpha_1]$. Now assume that $p^\dagger_{S_1}$ is symmetric with $p^*_{S_1}$, that is, $p^\dagger_{S_1}$ will be a PDF for a nonpositive random variable such that $p^*_{S_1}(-x) = p^\dagger_{S_1}(x)$. Similarly, for a given $p^*_{S_0}$ and $p^\dagger_{S_1}$, we maximize $\pi_1 \int_{-\infty}^{\infty} p^\dagger_{S_1}(x)\, G(x,\alpha)dx + \pi_0 \int_{-\infty}^{\infty} p^*_{S_0}(x)\,(1-G(x,\alpha))dx$ over $\alpha \in [\alpha_0,\alpha_1]$. Because of the first condition in the proposition, for every $\alpha \in [\alpha_0,\alpha_1]$, $\int_{-\infty}^{\infty} p^*_{S_1}(x)\, G(x,\alpha)dx \leq \int_{-\infty}^{\infty} p^\dagger_{S_1}(x)\, G(x,\alpha)dx$, since $G(x,\alpha)$ is strictly decreasing function of x; hence, the value of the maximum for $p^*_{S_1}$ will be less than or equal to that for $p^\dagger_{S_1}$, and both PDFs will yield the same average power value because of the symmetry. Since it is a minimax problem, we look for the optimal signal PDFs p_{S_0} and p_{S_1} which minimize the value of the maximum. Thus, by using a nonnegative S_1 we achieve a lower maximum value as compared to a nonpositive S_1. Similarly, a nonpositive S_0 will yield a smaller maximum value as compared to a nonnegative S_0. Therefore, instead of considering all PDFs, one can just consider the PDFs of a nonpositive S_0 and a nonnegative S_1 without loss of generality under the first condition in the proposition.

By using this fact, for any given p_{S_0} and p_{S_1}, which are the PDFs of a nonpositive S_0 and a nonnegative S_1 respectively, we maximize $V(\alpha) = \pi_1 \int_0^\infty p_{S_1}(x)\, G(x,\alpha)dx + \pi_0 \int_{-\infty}^0 p_{S_0}(x)\,(1 - G(x,\alpha))dx$ over $\alpha \in [\alpha_0, \alpha_1]$. Define

$$V_1(\alpha) = \int_0^\infty p_{S_1}(x)\, G(x,\alpha)dx$$

and

$$V_0(\alpha) = \int_{-\infty}^0 p_{S_0}(x)\, G(x,\alpha)dx \ .$$

Then, we maximize $V(\alpha) = \pi_1\, V_1(\alpha) - \pi_0\, V_0(\alpha) + \pi_0$ over $\alpha \in [\alpha_0, \alpha_1]$. Under the second condition in the proposition, $\frac{\partial G(x,\alpha)}{\partial \alpha} < 0$, $\forall x > 0$ and $\frac{\partial G(x,\alpha)}{\partial \alpha} > 0$, $\forall x < 0$ [3]. First, assume that $p_{S_i}(x) \neq \delta(x)$ for $i = 0, 1$. Then,

$$\frac{dV_1(\alpha)}{d\alpha} = \int_0^\infty p_{S_1}(x)\, \frac{\partial G(x,\alpha)}{\partial \alpha} dx < 0.$$

Similarly,

$$\frac{dV_0(\alpha)}{d\alpha} = \int_{-\infty}^0 p_{S_0}(x)\, \frac{\partial G(x,\alpha)}{\partial \alpha} dx > 0.$$

Therefore, we can write that $\frac{dV(\alpha)}{d\alpha} = \pi_1 \frac{dV_1(\alpha)}{d\alpha} - \pi_0 \frac{dV_0(\alpha)}{d\alpha} < 0$. This shows that $V(\alpha)$ is a strictly decreasing function of α. Hence, for p_{S_0} and p_{S_1}, under the conditions in the proposition, $\max\limits_{\alpha \in [\alpha_0\ \alpha_1]} V(\alpha) = V(\alpha_0)$, meaning that the minimax problem can be reduced to the form in (4.21). Note that, when $p_{S_i}(x) = \delta(x)$, then $\frac{dV_i(\alpha)}{d\alpha} = 0$. If $p_{S_1}(x) = p_{S_0}(x) = \delta(x)$, then $V(\alpha)$ becomes a constant function. Also, if one of $p_{S_1}(x)$ or $p_{S_0}(x)$ is not equal to $\delta(x)$, $V(\alpha)$ is still strictly decreasing function of α. Therefore, $\max\limits_{\alpha \in [\alpha_0\ \alpha_1]} V(\alpha) = V(\alpha_0)$ holds for all possible p_{S_0} and p_{S_1} in fact. □

Proposition 4.2 states that, under certain sufficient conditions, the robust design of stochastic signals becomes equivalent to the stochastic signal design for the smallest magnitude of the channel coefficient in set Ω. The simplified problem

[3] When $x = 0$, $G(x,\alpha)$ is independent of α and just a constant as it can be seen from (4.10).

in (4.21) has a well-known structure, which was investigated for example in [17]. The problem can be solved separately for S_0 and S_1 by expressing the problem as two decoupled optimization problems. Then it can be shown that if $G(S_i, \alpha_0)$ is a continuous function of S_i and S_i takes values in $[-\gamma, \gamma]$ for some finite $\gamma > 0$, then each optimal signal PDF p_{S_i} can be represented by a randomization of at most two signal levels [17, 39].

It is also noted that if $[\alpha_0, \alpha_1]$ is a positive interval, then the two conditions in Proposition 4.2 can be reduced to a single condition. Suppose that $u = \alpha x$, then $G(x, \alpha)$ can be written as $G(u) = \int_{\Gamma_0} p_N(y - u)\, dy$. Therefore, if α is positive, then the conditions in Proposition 4.2 are equivalent to that $G(u)$ is a decreasing function of u, that is, $\frac{dG(u)}{du} < 0$.

After obtaining the optimal signal PDFs p_{S_0} and p_{S_1} by solving (4.21), the conditional average probability of error for a given $\alpha \in \Omega$ can be calculated as

$$\mathrm{P}^{\alpha}_{\mathrm{rob}} = \sum_{i=0}^{1} \pi_i \int_{-\infty}^{\infty} p_{S_i}(x) \int_{\Gamma_{1-i}} p_N(y - \alpha\, x)\, dy\, dx \ . \tag{4.23}$$

Finally, the average probability of error for robust stochastic signaling can be calculated as

$$\mathrm{P}_{\mathrm{rob}} = \int_{\Omega} p_{\alpha}(a)\, \mathrm{P}^{a}_{\mathrm{Rob}}\, da \ . \tag{4.24}$$

Note that, while calculating the conditional average probability of error for a given α, the same signal PDF is used for all α values, since the optimal signal PDFs do not depend on the value of the actual channel coefficient α, but only depend on the lower boundary point of the set Ω in robust stochastic signaling.

4.3.2 Stochastic Signaling with Averaging

In robust stochastic signaling, signal PDFs are designed for the worst-case channel coefficient, which belongs to a certain set Ω. In this section, an alternative

way of designing stochastic signals under CSI uncertainty is discussed. In this method, the transmitter assumes that the channel coefficient is distributed according to a PDF $\hat{p}_\alpha(\cdot)$.[4] Then, optimal stochastic signal PDFs are designed in such a way that the average probability of error is minimized for this assumed CSI statistics under the average power constraints. This can be formulated as follows:

$$\min_{p_{S_0}, p_{S_1}} \int_{-\infty}^{\infty} \hat{p}_\alpha(a) \sum_{i=0}^{1} \pi_i \int_{-\infty}^{\infty} p_{S_i}(x) \int_{\Gamma_{1-i}} p_N(y - ax) dy dx da$$
$$\text{subject to } \ \mathrm{E}\{|S_i|^2\} \leq A \ . \tag{4.25}$$

Note that this problem is separable over S_0 and S_1 as well. Therefore, one can consider the the optimal signals for symbol 0 and symbol 1 separately. Specifically, the optimal signal PDF for symbol 1 can be obtained by solving the following problem:

$$\min_{p_{S_1}} \int_{-\infty}^{\infty} \hat{p}_\alpha(a) \int_{-\infty}^{\infty} p_{S_1}(x) \int_{\Gamma_0} p_N(y - ax) \, dy \, dx \, da$$
$$\text{subject to } \ \mathrm{E}\{|S_1|^2\} \leq A \ . \tag{4.26}$$

Changing the order of the first and the second integrals in (4.26), the following formulation can be obtained:

$$\min_{p_{S_1}} \int_{-\infty}^{\infty} p_{S_1}(x) \int_{-\infty}^{\infty} \hat{p}_\alpha(a) \, G(x, a) \, da \, dx$$
$$\text{subject to } \ \mathrm{E}\{|S_1|^2\} \leq A \tag{4.27}$$

where $G(x, a)$ is as defined in (4.10). In addition, if $H(x)$ is defined as

$$H(x) \triangleq \int_{-\infty}^{\infty} \hat{p}_\alpha(a) G(x, a) \, da = \mathrm{E}\{G(x, a)\} \tag{4.28}$$

where the expectation is taken over the assumed PDF of the channel coefficient, then (4.27) can be written as

$$\min_{p_{S_1}} \mathrm{E}\{H(S_1)\} \quad \text{subject to } \mathrm{E}\{|S_1|^2\} \leq A \ . \tag{4.29}$$

[4]Note that this will not be the actual PDF of the channel coefficient in general due to CSI uncertainty at the transmitter.

For this problem, it can be concluded that, under most practical scenarios, the optimal signal PDF can be characterized by a randomization between at most two signal levels similarly to the previous results. The optimal signal PDF for symbol 0 can be obtained similarly.

In the stochastic signaling with averaging approach, the transmitter assigns different weights to different values of the channel coefficient and designs signals based on this averaging operation over possible channel coefficient values. For example, instead of directly using the distorted channel coefficient $\hat{\alpha}$ in the signal design as in Section 4.2.1, the transmitter may assume a legitimate PDF around $\hat{\alpha}$ for the channel coefficient and design the stochastic signals. The performance of this approach and the other approaches is compared in the following section.

4.4 Performance Evaluation

In this section, two numerical examples are presented in order to investigate the theoretical results in the previous sections. In the first numerical example, we compare the performance of conventional signaling and stochastic signaling in the presence of channel coefficient errors and observe the effects of CSI uncertainty on stochastic signaling. In the second example, we evaluate the performance of the proposed design methods in Section 4.3. In both of the examples, a binary communications system with equally likely symbols are considered ($\pi_0 = \pi_1 = 0.5$), the average power limit in (4.4) is set to $A = 1$, and the decision rule at the receiver is specified by $\Gamma_0 = (-\infty, 0]$ and $\Gamma_1 = [0, \infty)$ (i.e., the sign detector). Also the noise in (4.1) is modeled by a Gaussian mixture noise [7] with its PDF being given by

$$p_N(n) = \frac{1}{\sqrt{2\pi}\,\sigma} \sum_{l=1}^{L} v_l \, \mathrm{e}^{-\frac{(n-\mu_l)^2}{2\sigma^2}} \quad . \tag{4.30}$$

Gaussian mixture noise is encountered in practical systems in the presence of interference [7]. For the channel noise and the detector structure as described

above, $G(x,k)$ in (4.10) can be calculated as

$$G(x,k) = \sum_{l=1}^{L} v_l \, Q\left(\frac{k\,x + \mu_l}{\sigma}\right) \ . \tag{4.31}$$

In the first example, the mass points μ_l are located at $\boldsymbol{\mu} = [-1.013 \ -0.275 \ -0.105 \ 0.105 \ 0.275 \ 1.013]$ with corresponding weights $\mathbf{v} = [0.043 \ 0.328 \ 0.129 \ 0.129 \ 0.328 \ 0.043]$. Also each component of the Gaussian mixture noise has the same variance σ^2 and the average power of the noise can be calculated as $\mathrm{E}\{n^2\} = \sigma^2 + 0.1407$.

The channel coefficient information at the transmitter is modeled as $\hat{\alpha} = \alpha + \eta$, where $\alpha = 1$ and η is a zero-mean Gaussian random variable with variance ε^2. Due to the symmetry of the problem, the conditional probability of error expression in (4.14) also provides the average probability of error in this scenario. In order to evaluate that expression, 100 realizations are obtained for $\hat{\alpha}$. Then, the optimization problem in (4.11) is solved for each realization and the optimal signal PDFs that are in the form of (4.12) are obtained by using the PSO algorithm [30]. For the details of the PSO parameters employed in this study, please refer to [40].

In Fig. 4.1, the average probabilities of error are plotted versus A/σ^2 for conventional signaling, stochastic signaling with no channel coefficient errors ($\varepsilon = 0$), and stochastic signaling with various levels of channel coefficient errors. It is observed that, for high A/σ^2 values, the best performance is obtained by stochastic signaling with perfect CSI and the performance of stochastic signaling gets worse as the variance of the channel coefficient error increases. For example, when $\varepsilon = 0.5$ and $\varepsilon = 0.6$, stochastic signaling performs worse than conventional signaling for all A/σ^2 values. Another observation is that for low values of ε, stochastic signaling still performs better than conventional signaling for high A/σ^2 values and their performance is similar for high σ^2, i.e. when A/σ^2 is smaller than 15 dB. In fact, one can calculate the average probability of error

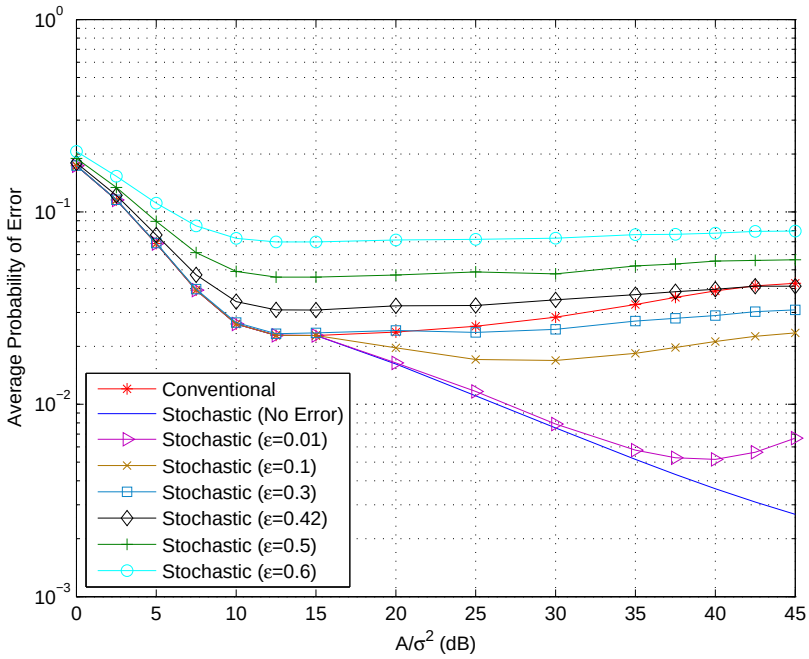

Figure 4.1: Average probability of error versus A/σ^2 for conventional signaling and stochastic signaling with various ε values.

analytically for low A/σ^2 values for each ε. At low A/σ^2 values, $\mathrm{P}_{\mathrm{e}}^{\hat{\alpha}}$ in (4.13) can be expressed as

$$\mathrm{P}_{\mathrm{e}}^{\hat{\alpha}} = \frac{1-\mathrm{sgn}(\hat{\alpha})}{2} + \mathrm{sgn}(\hat{\alpha})\, G(\sqrt{A}, \alpha)) \tag{4.32}$$

where sgn denotes the sign operator. Then, from (4.14), P_{e} can be calculated as $Q(\alpha/\varepsilon) + G(\sqrt{A}, \alpha) - 2\, G(\sqrt{A}, \alpha)\, Q(\alpha/\varepsilon)$. For instance, when $A/\sigma^2 = 10\,\mathrm{dB}$, $G(\sqrt{A}, \alpha) = 0.02613$ in this example. Then, for $\varepsilon = 0.6$, P_{e} is calculated as $0.9477\, Q(5/3) + 0.0261 = 0.0714$, which is very close to the result shown in Fig. 4.1. For this example, we can apply the conditions given in Proposition 4.1 and calculate $\hat{\varepsilon}$ and ε^*. Firstly, we check the first condition in the proposition. $G(x, k)$ is calculated above for this example and it is a linear combination of Q functions. Therefore, $G(x, k)$ is a strictly decreasing function of x as $Q(x)$ is a monotone decreasing function. Also, since $Q(x) = 1 - Q(-x)$ and the components of Gaussian mixture noise are symmetric, we have $G(x, k) = 1 - G(-x, k)$ as well. Hence, the first condition in Proposition 4.1 is satisfied. In order to check the second condition, the plot of $\mathrm{P}_{\mathrm{e}}^{\hat{\alpha}}$ versus $\hat{\alpha}$ is presented in Fig. 4.2. It is observed

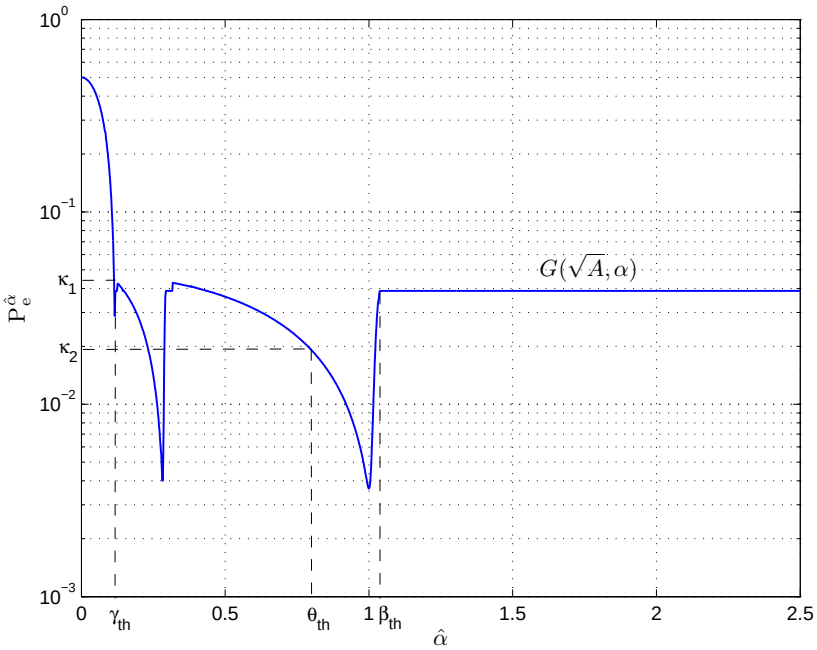

Figure 4.2: $P_e^{\hat{\alpha}}$ versus $\hat{\alpha}$ for $A/\sigma^2 = 40\,\text{dB}$. The second condition in Proposition 4.1 is satisfied for $\kappa_1 = 0.04354$, $\kappa_2 = 0.01913$, $\gamma_{th} = 0.1135$, $\theta_{th} = 0.8$, $\beta_{th} = 1.038$, and $G(\sqrt{A}, \alpha)) = 0.03884$.

that $P_e^{\hat{\alpha}}$ does not have a monotonic structure; that is, it increases, decreases or remains the same as $\hat{\alpha}$ increases. However, it obeys the structure specified in the second condition of Proposition 4.1. Specifically, when $\hat{\alpha} > \gamma_{th} = 0.1135$, $P_e^{\hat{\alpha}}$ is less than $\kappa_1 = 0.04354$, and when $\theta_{th} = 0.8 < \hat{\alpha} < \alpha = 1$, $P_e^{\hat{\alpha}}$ becomes less than $\kappa_2 = 0.01913$, which is even smaller than κ_1. Also, when $\hat{\alpha} > \beta_{th} = 1.038$, $P_e^{\hat{\alpha}}$ becomes equal to $G(\sqrt{A}, \alpha)) = 0.03884$, which is the average probability of error for conventional signaling. The values of κ_1, κ_2, γ_{th}, θ_{th}, and β_{th} are illustrated in Fig. 4.2. Now, by using the above parameters and solving (4.15), which becomes

$$0.45646\,Q\left(\frac{1.1135}{\varepsilon^*}\right) + 0.5\,Q\left(\frac{1}{\varepsilon^*}\right) + 0.02441\,Q\left(\frac{2}{\varepsilon^*}\right) - Q\left(\frac{1.8}{\varepsilon^*}\right)$$
$$= 0.03884\left(1 - Q\left(\frac{0.038}{\varepsilon^*}\right)\right) \;,$$

one can obtain $\varepsilon^* = 0.5394$. This means that when $A/\sigma^2 = 40\,\text{dB}$, if the standard deviation of the channel coefficient error is larger than 0.5394, we can conclude that stochastic signaling is outperformed by conventional signaling. In fact, it can be observed from Fig.4.1 that for $A/\sigma^2 = 40\,\text{dB}$ and $\varepsilon = 0.6 > \varepsilon^*$, the

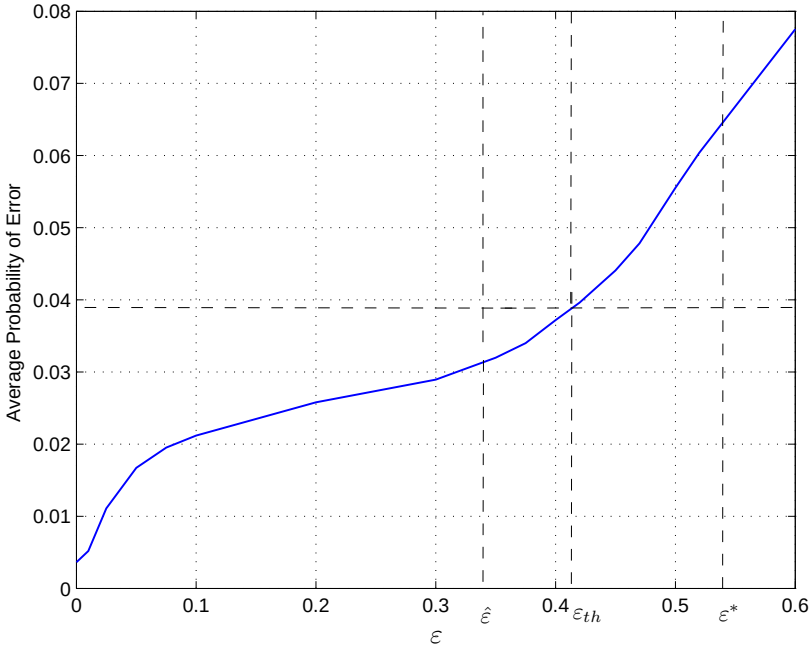

Figure 4.3: Average probability of error versus ε for stochastic signaling. At $\varepsilon_{th} = 0.413$, stochastic signaling has the same average probability of error as conventional signaling.

performance of stochastic signaling is quite worse than conventional signaling as Proposition 4.1 asserts. Similarly, by solving (4.16), which becomes

$$0.5\left(0.06267 + Q\left(\frac{1}{\hat{\varepsilon}}\right)\right) + 0.45646\,Q\left(\frac{0.8865}{\hat{\varepsilon}}\right) - 0.04354\,Q\left(\frac{0.038}{\hat{\varepsilon}}\right)$$
$$+\,0.02441\,Q\left(\frac{0.2}{\hat{\varepsilon}}\right) = 0.03884\left(1 - Q\left(\frac{0.038}{\hat{\varepsilon}}\right) + Q\left(\frac{2.038}{\hat{\varepsilon}}\right)\right)\ ,$$

one can calculate $\hat{\varepsilon} = 0.3395$. This means that, at $A/\sigma^2 = 40\,\text{dB}$, if the standard deviation of the channel coefficient error is smaller than 0.3395, we can conclude that conventional signaling is outperformed by stochastic signaling. From Fig. 4.1, it is seen that for $A/\sigma^2 = 40\,\text{dB}$ and $\varepsilon = 0.3,\ 0.1\,, 0.01 < \hat{\varepsilon}$, stochastic signaling performs better than conventional signaling.

In order to explore performance variations of stochastic signaling with respect to ε, Fig. 4.3 is presented. It is observed that as the variance of the channel coefficient error increases, the average probability of error for stochastic signaling increases. This is expected since the transmitter designs the stochastic signals in the presence of channel coefficient errors (imperfect CSI) and these errors get more significant as ε increases. Therefore, it can be concluded that in the

presence of large channel coefficient errors (i.e., large ε), using conventional deterministic signaling instead of stochastic signaling would be more preferable, whereas for small channel coefficient errors, stochastic signaling can be employed to achieve smaller average probabilities of error than conventional signaling. In Fig. 4.3, the upper bound ε^* and the lower bound $\hat{\varepsilon}$ obtained from Proposition 4.1 are also illustrated, together with the point ε_{th} at which the performance of stochastic signaling and conventional signaling becomes the same. It is observed that Proposition 4.1 provides sufficient conditions for the improvability and nonimprovability of conventional signal via stochastic signaling. However, the conditions are not necessary as illustrated in Fig. 4.3.

In the second example, the mass points μ_l of the Gaussian mixture noise are located at $\boldsymbol{\mu} = [-1.31\ -0.275\ -0.125\ 0.125\ 0.275\ 1.31]$ with corresponding weights $\mathbf{v} = [0.002\ 0.319\ 0.179\ 0.179\ 0.319\ 0.002]$. Each component of the Gaussian mixture noise has the same variance σ^2 and the average power of the noise can be calculated as $\mathrm{E}\{n^2\} = \sigma^2 + 0.0607$. For this example, $\hat{\alpha}$ is again modeled as $\hat{\alpha} = \alpha + \eta$ where η is a zero-mean Gaussian random variable with variance ε^2. We assume that the actual channel coefficient α has a uniform distribution over set $\Omega = [0.8, 1.2]$; i.e., α is distributed as $\mathcal{U}[0.8, 1.2]$.

First, we compare the average probability of error performance of different signaling strategies:

Stochastic-Perfect: It is assumed that the transmitter has the knowledge of the actual channel coefficient, which is used in the signal design. In the simulations, 100 realizations are generated for uniformly distributed α. The optimal signal PDFs and the corresponding probabilities of error are calculated for each realization. Then, by averaging over the PDF of α, the average probabilities of error are obtained.

Conventional: The transmitter selects the signals as $S_1 = -S_0 = \sqrt{A} = 1$. For each realization of α, the corresponding probabilities of error are calculated and then their average is taken over the PDF of α.

Stochastic-Distorted: The transmitter has imperfect CSI and it uses a distorted (imperfect) channel coefficient $\hat{\alpha}$ directly in the design of signals, as discussed in Section 4.2.1. In Fig. 4.4, average probabilities of error are plotted for $\varepsilon = 0.05$ and $\varepsilon = 0.1$.

Stochastic-Average: The transmitter assumes that the PDF of the channel coefficient is $\hat{p}_\alpha(a)$ is specified by $\mathcal{N}(\hat{\alpha}, \Delta^2)$. Then, by solving (4.29), the optimal signal PDF $p_{S_1}^{\hat{\alpha}}$ for signal 1 can be obtained for each $\hat{\alpha}$. Next, the conditional probability of error for symbol 1 can be expressed as

$$\mathrm{P}_{\text{aver}} = \int_{-\infty}^{\infty} p_\alpha(a) \int_{-\infty}^{\infty} p_{\hat{\alpha}|\alpha}(\hat{a}) \int_{-\infty}^{\infty} p_{S_1}^{\hat{a}}(x)\, G(x,a)\, dx d\hat{a} da \tag{4.33}$$

where $p_{\hat{\alpha}|\alpha}(\cdot)$ is the conditional PDF of $\hat{\alpha}$ for a given α. Note that, due to the symmetry, the conditional error probability is equal to the average probability of error in this example as well. In Fig. 4.4, the average probabilities of error are plotted for $\Delta = 0.01$, $\Delta = 0.05$, and $\Delta = 0.2$, where $\varepsilon = 0.05$ in each case.

Stochastic-Robust: First, one can show that this example satisfies the conditions in Proposition 4.2. In this example, $G(x, \alpha)$ can be calculated by using (4.31). Note that $G(x, \alpha)$ is a linear combination of Q functions, i.e. $Q\left(\frac{\alpha\, x + \mu_l}{\sigma}\right)$. Then, since α is always be positive ($\alpha \in [0.8, 1.2]$), $Q\left(\frac{\alpha\, x + \mu_l}{\sigma}\right)$ is a decreasing function of x. Also, it is a decreasing function of α if x is positive, and it increases with α when x is negative. In fact, since $[0.8, 1.2]$ is a positive interval, we can write $u = \alpha\, x$ and $G(u)$ will be a decreasing function of u as $Q\left(\frac{u + \mu_l}{\sigma}\right)$ decreases with u. Therefore, we can apply the result in Proposition 4.2 in this example. That is, the optimal signal PDFs are obtained by solving (4.20) with $\alpha_0 = 0.8$ as $\Omega = [0.8, 1.2]$. Then, the average probabilities of error are calculated via (4.23) and (4.24).

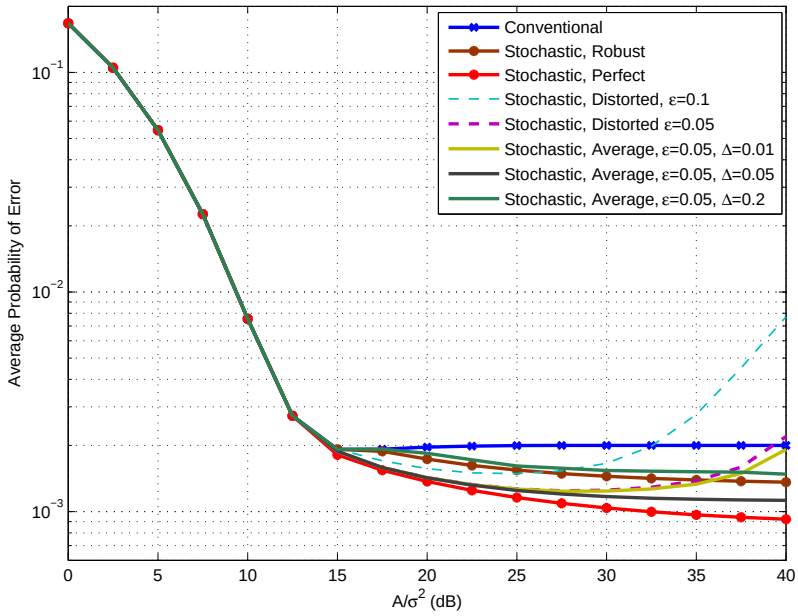

Figure 4.4: Average probability of error versus A/σ^2 for various signaling strategies.

In Fig. 4.4, the average probabilities of error are plotted versus A/σ^2 for conventional signaling, stochastic signaling with perfect CSI, distorted channel coefficient, averaging and robust stochastic signaling. It is observed that, for high σ^2, that is, specifically when A/σ^2 is smaller than 15 dB, all signaling strategies perform similarly. For high A/σ^2 values, it is observed that stochastic signaling with perfect CSI achieves the best performance. The second best performance is obtained by the stochastic signaling with averaging method when the parameters are $\epsilon = \Delta = 0.05$. Although conventional signaling gives the worst performance for medium A/σ^2 values, the worst performance is observed for stochastic signaling with distorted channel coefficient for high A/σ^2 values. Robust stochastic signaling performs somewhere between stochastic signaling with perfect CSI and conventional signaling. Robust signaling performs better (worse) than stochastic signaling with averaging for $\Delta = 0.2$ ($\Delta = 0.05$) at high or medium A/σ^2 values. For $\epsilon = 0.05$, stochastic signaling with averaging when $\Delta = 0.01$ and stochastic signaling with distorted channel coefficient performs very similarly and they achieve better performance than robust signaling for medium A/σ^2 values; however, their performance is worse than robust signaling for high A/σ^2 values.

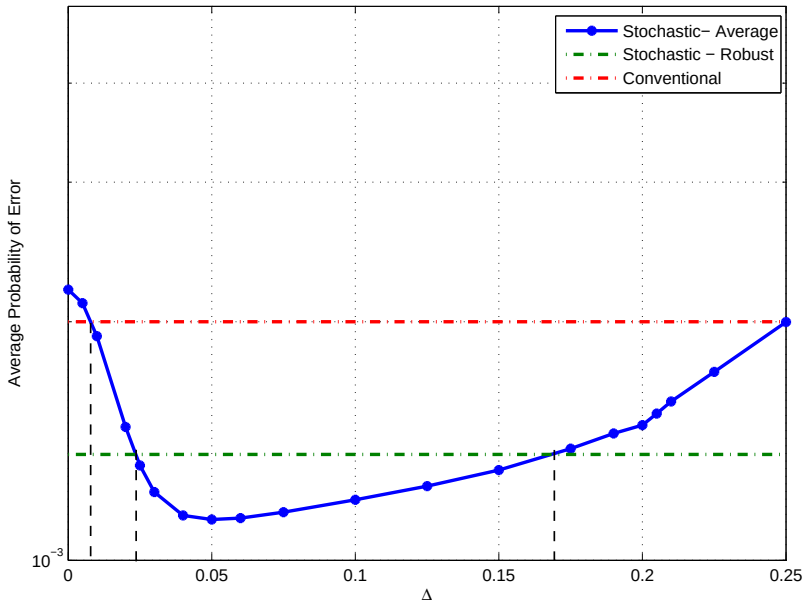

Figure 4.5: Average probability of error versus Δ for stochastic signaling with averaging when $A/\sigma^2 = 40dB$ and $\epsilon = 0.05$. Stochastic signaling with averaging performs same with conventional signaling when $\Delta = 0.0078$. It has the same average probability of error as robust stochastic signaling at $\Delta = 0.0236$ and $\Delta = 0.1684$.

In order to investigate the effects of value of Δ on the average probability of error performance of the stochastic signaling with averaging method, Fig. 4.5 is presented. It can be observed that setting Δ to 0.05 provides the best performance. This means that the average probability of error performance is smaller when the standard deviation of the assumed PDF of the channel coefficient Δ gets closer to the standard deviation of the channel coefficient error ϵ. As we increase or decrease the value of Δ from 0.05, the average probability of error increases. Therefore, choosing very small or very large Δ values degrades the performance of the stochastic signaling with averaging strategy. Note that $\Delta = 0$ corresponds to the stochastic signaling with distorted channel coefficient in fact. It can be observed from Fig. 4.5 that if Δ is less than 0.0078, conventional signaling which has an average probability of error of 0.002 is better than this averaging strategy. Also, if Δ is less than 0.0236 or it is larger than 0.1684, robust stochastic signaling which has an average probability of error of 0.00136 achieves a better performance than stochastic signaling with averaging,

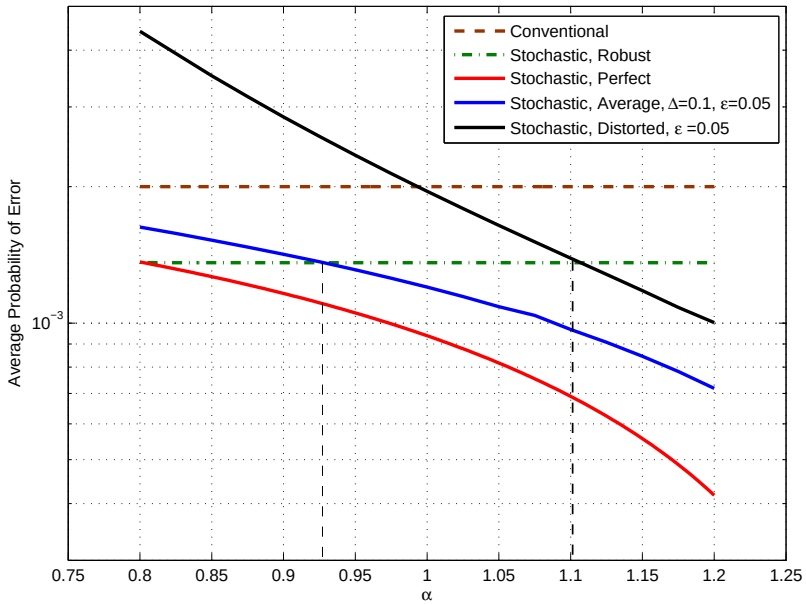

Figure 4.6: Average probability of error versus α for various signaling strategies when $A/\sigma^2 = 40dB$.

whereas the performance of stochastic signaling with averaging is better than robust signaling if $0.0236 < \Delta < 0.1684$.

Furthermore, we investigate in Fig. 4.6 the average probability of error performance of conventional signaling, stochastic signaling with perfect CSI, robust stochastic signaling, stochastic signaling with averaging when $\epsilon = 0.05$ and $\Delta = 0.1$, and stochastic signaling with distorted channel coefficient when $\epsilon = 0.05$ versus the actual value of the channel coefficient α when $A/\sigma^2 = 40dB$. We observe that the average probability of error decreases as α increases for all strategies [5]. For each value of the channel coefficient, the lower bound for the probability of error is obtained by stochastic signaling with perfect CSI. For small values of α, i.e., when $\alpha < 0.9276$, robust stochastic signaling is better than stochastic signaling with averaging. However, for larger α values such as when $\alpha > 1.107$, robust signaling performs worse than stochastic signaling with averaging and with distorted channel coefficient. This shows that since the signals are designed for $\alpha_0 = 0.8$ in robust stochastic signaling, when the actual α

[5] Although it is not very clear in Fig. 4.6, the average probabilities of error for conventional signaling and robust signaling also slightly decrease as α increases.

is close to that value, robust signaling gives a better performance. Performance of stochastic signaling with averaging is better than conventional signaling and stochastic signaling with distorted channel coefficient for every α value. Although conventional signaling gives larger average probabilities of error than stochastic signaling with distorted channel coefficient for $\alpha > 0.9935$, using noisy a channel coefficient in the signal design directly results in the worst average probability of error performance when α has a smaller value.

Finally, in order to provide additional explanations of the results, Table 4.1 and Table 4.2 are presented. In Table 4.1, the optimal signals for robust stochastic signaling and stochastic signaling for the given channel coefficient value α are presented for various A/σ^2 values. Note that in robust signaling the actual value of α is not important since the signals are designed for $\alpha = 0.8$. It is observed that when $A/\sigma^2 = 10dB$ both strategies have the same solution as the conventional signaling. However, as A/σ^2 increases, the randomization between two signal values becomes more effective and this may help reduce the average probability of error. For example, when $A/\sigma^2 = 25dB$, the average probability of error for robust signaling is 0.00155, whereas it is 0.00199 for conventional signaling. In Table 4.2, the optimal signals for stochastic signaling with averaging when $A/\sigma^2 = 40dB$ are presented. Note that the assumed PDF of the channel coefficient in that strategy is $\mathcal{N}(\hat{\alpha}, \Delta^2)$. It is observed that when Δ is very small, i.e., $\Delta = 0.01$, the optimal signal PDFs are close to the optimal signal PDFs of the stochastic signaling case given in Table 4.1. Also, when $\hat{\alpha} = 0.9$ and $\Delta = 0.2$, the optimal signal PDF is close to that for conventional signaling since the optimal PDF has a mass point at 0.9684 with a weight of 0.9302.

Table 4.1: Optimal signals for stochastic signaling for various α and robust design for symbol 1.

		Stochastic		
A/σ^2 (dB)	α	λ_1	$\mathbf{s}_{11}$	$\mathbf{s}_{12}$
10	0.9	N/A	1	1
10	1.1	N/A	1	1
25	0.9	0.3254	1.5642	0.5496
25	1.1	0.5557	1.2798	0.4497
40	0.9	0.4211	1.4838	0.3546
40	1.1	0.6590	1.214	0.2901
		Robust		
A/σ^2 (dB)	α	λ_1	$\mathbf{s}_{11}$	$\mathbf{s}_{12}$
10	N/A	N/A	1	1
25	N/A	0.2276	1.7597	0.6183
40	N/A	0.3200	1.6693	0.3989

Table 4.2: Optimal signals for stochastic signaling with averaging for symbol 1 when $A/\sigma^2 = 40dB$.

		Averaging		
$\hat{\alpha}$	Δ	λ_1	$\mathbf{s}_{11}$	$\mathbf{s}_{12}$
0.9	0.01	0.41	1.5016	0.3575
0.9	0.05	0.351	1.5922	0.4114
0.9	0.2	0.0698	1.3519	0.9684
1.1	0.01	0.6466	1.2247	0.2917
1.1	0.05	0.575	1.2892	0.323
1.1	0.2	0.476	1.2815	0.6453

4.5 Conclusions

The effects of imperfect CSI on stochastic signaling and the design of stochastic signals in the presence of CSI uncertainty have been investigated. First, a problem formulation has been presented to explore the effects of errors in the channel coefficient, and the two mass point structure of an optimal signal PDF has been observed when the signals are designed based on noisy channel coefficients at the transmitter. Then, sufficient conditions have been presented to specify when the performance of conventional deterministic signaling can or cannot be improved via stochastic signaling. Upper and lower bounds on the variance of the channel estimation error have been derived and improvability and nonimprovability conditions have been presented. Then, two different signaling strategies, called

robust stochastic signaling and stochastic signaling with averaging, have been discussed. Sufficient conditions are derived to obtain an equivalent but simpler form for the robust stochastic signaling design problem. Finally, the theoretical results have been presented over two examples.

Bibliography

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

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

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

[4] F. Cabarcas, R. D. Souza, and J. Garcia-Frias, "Turbo coding of strongly nonuniform memoryless sources with unequal energy allocation and PAM signaling," *IEEE Trans. Sig. Processing*, vol. 54, no. 5, pp. 1942–1946, May 2006.

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

[6] M. Azizoglu, "Convexity properties in binary detection problems," *IEEE Trans. Inform. Theory*, vol. 42, no. 4, pp. 1316–1321, July 1996.

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

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

[9] M. A. Klimesh and W. E. Stark, "Worst-case power-constrained noise for binary-input channels with varying amplitude signals," in *Proc. IEEE Int. Symp. on Inform. Theory (ISIT)*, July 1994, p. 381.

[10] S. Loyka, V. Kostina, and F. Gagnon, "Error rates of the maximum-likelihood detector for arbitrary constellations: Convex/concave behavior and applications," *IEEE Trans. on Inform. Theory.*, vol. 56, no. 4, pp. 1948–1960, Apr. 2010.

[11] A. Patel and B. Kosko, "Optimal noise benefits in Neyman-Pearson and inequality-constrained signal detection," *IEEE Trans. Sig. Processing*, vol. 57, no. 5, pp. 1655–1669, May 2009.

[12] L. Huang and M. J. Neely, "The optimality of two prices: Maximizing revenue in a stochastic network," in *Proc. 45th Annual Allerton Conference on Commun., Control, and Computing*, Monticello, IL, Sep. 2007.

[13] K. E. Parsopoulos and M. N. Vrahatis, *Particle swarm optimization method for constrained optimization problems.* IOS Press, 2002, pp. 214–220, in Intelligent Technologies–Theory and Applications: New Trends in Intelligent Technologies.

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

[15] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory.* Prentice Hall, Inc., 1998.

[16] E. L. Lehmann, *Testing Statistical Hypotheses*, 2nd ed. Springer, 1997.

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

[18] S. Bayram, S. Gezici, and H. V. Poor, "Noise enhanced hypothesis-testing in the restricted Bayesian framework," *IEEE Trans. Sig. Processing*, vol. 58, no. 8, Aug. 2010.

[19] 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, Sep. 2009, pp. 31–36.

[20] A. L. McKellips and S. Verdu, "Worst case additive noise for binary-input channels and zero-threshold detection under constraints of power and divergence," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1256–1264, July 1997.

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

[22] M. C. Gursoy, H. V. Poor, and S. Verdu, "Efficient signaling for low-power Rician fading channels," in *Proc. Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Oct. 2002.

[23] M. McGuire, "Location of mobile terminals with quantized measurements," in *Proc. IEEE Int. Symp. Personal, Indoor, Mobile Commun. (PIMRC)*, vol. 3, Berlin, Germany, Sep. 2005, pp. 2045–2049.

[24] T. Erseghe and S. Tomasin, "Optimized demodulation for MAI resilient UWB W-PAN receivers," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Beijing, China, May 2008, pp. 4867–4871.

[25] 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, no. 7, pp. 3172–3184, July 2007.

[26] C. C. Pugh, *Real Mathematical Analysis.* New York: Springer-Verlag, 2002.

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

[28] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar, *Convex Analysis and Optimization.* Boston, MA: Athena Specific, 2003.

[29] K. V. Price, R. M. Storn, and J. A. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization.* New York: Springer, 2005.

[30] A. I. F. Vaz and E. M. G. P. Fernandes, "Optimization of nonlinear constrained particle swarm," *Baltic Journal on Sustainability*, vol. 12, no. 1, pp. 30–36, 2006.

[31] S. Koziel and Z. Michalewicz, "Evolutionary algorithms, homomorphous mappings, and constrained parameter optimization," *Evolutionary Computation*, vol. 7, no. 1, pp. 19–44, 1999.

[32] X. Hu and R. Eberhart, "Solving constrained nonlinear optimization problems with particle swarm optimization," in *Proc. Sixth World Multiconference on Systemics, Cybernetics and Informatics 2002 (SCI 2002)*, Orlando, FL, 2002.

[33] Y. Chen and V. K. Dubey, "Ultrawideband source localization using a particle-swarm-optimized Capon estimator," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 4, Seoul, Korea, May 2005, pp. 2825–2829.

[34] Y. Rahmat-Samii, D. Gies, and J. Robinson, "Particle swarm optimization (PSO): A novel paradigm for antenna designs," *The Radio Science Bulletin*, vol. 305, pp. 14–22, Sep. 2003.

[35] Z. Yangyang, J. Chunlin, Y. Ping, L. Manlin, W. Chaojin, and W. Guangxing, "Particle swarm optimization for base station placement in mobile communication," in *Proc. IEEE International Conference on Networking, Sensing and Control*, vol. 1, May 2004, pp. 428–432.

[36] W. Jatmiko, K. Sekiyama, and T. Fukuda, "A PSO-based mobile sensor network for odor source localization in dynamic environment: Theory, simulation and measurement," in *Proc. IEEE Congress on Evolutionary Computation*, Vancouver, BC, July 2006, pp. 1036–1043.

[37] J. Pugh, A. Martinoli, and Y. Zhang, "Particle swarm optimization for unsupervised robotic learning," in *Proc. Swarm Intelligence Symposium (SIS)*, Pasadena, California, June 2005, pp. 92–99.

[38] C. Goken, S. Gezici, and O. Arikan, "Stochastic signaling under second and fourth moment constraints," in *Proc. IEEE International Workshop on Signal Processing Advances for Wireless Communications (SPAWC)*, Marrakech, Morocco, June 2010.

[39] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis.* Berlin: Springer-Verlag, 2004.

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

[41] A. Goldsmith, *Wireless Communications.* Cambridge, UK: Cambridge University Press, 2005.

[42] C. Goken, S. Gezici, and O. Arikan, "On the optimality of stochastic signaling under an average power constraint," in *Proc. 48th Annual Allerton Conf. on Commun., Control, and Computing*, Illinois, Sep. 29-Oct. 1 2010, pp. 1158–1164.