OPTIMUM QUANTIZATION FOR DETECTION WITH SIDE INFORMATION

by

Nafiz Polat Ayerden

B.S., Electrical and Electronics Engineering, Boğaziçi University, 2006

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Department of Electrical and Electronics Engineering
Boğaziçi University

2009

ACKNOWLEDGEMENTS

First and foremost, I offer my sincerest gratitude for my thesis supervisor, Mehmet Kıvanç Mıhçak, who has supported me throughout my thesis with his patience, knowledge and friendship. I can't think of a better supervisor, not only because teaching me almost everything about research but also because his tenacity, enthusiasm make him one of the most important role model I had in my life.

I am thankful to Emin Anarım for the opportunity to work in BUICS that I will always feel proud for being a member, and also thankful to Süleyman Serdar Kozat for his helpful comments and attendance to my thesis presentation.

The gratitude from the bottom of my heart goes to, Yücel Altuğ, Kürşat Çoban, Mustafa Orhan Dirik, Muharrem Orkun Sağlamdemir, Cumhur Ozan Yalçın and Ömer Yetik simply for their friendship, which I hope to enjoy throughout the rest of my life.

Most importantly, I am indebted to my family and beloved Fatmagül Topçu for their true love, limitless support and endless patience.

ABSTRACT

OPTIMUM QUANTIZATION FOR DETECTION WITH SIDE INFORMATION

In this work, we considered the problem of binary detection with side information. The receiver, observes the transmitted data which is corrupted by Gaussian noise, and tries to make a decision between two hypotheses where it has the knowledge of noise statistics of the channel and partial information about the data. Here the partial information is obtained via passing the original data through a quantizer, thus the partial information is simply the reproduction value of the bin that the corresponding data is in. We derived the optimal decision rule and corresponding probability of error. We presented and illustrated the optimal quantizers for several quantization levels. Next, we compare quantizers (optimal quantizer, Lloyd-Max, Uniform,a suboptimal quantizer obtained by approximation), with respect to bin constellations and their corresponding probability of detection errors. Finally, from the comparison it has been shown empirically that Lloyd-Max quantizers are suboptimal yet good choice for detection with side information problem under proposed setup.

ÖZET

YAN BİLGİ İLE SEZİMLEME İÇİN EN İYİ NİCELEME

Bu çalışmada yan bilgi ile ikili sezimleme problemini inceledik. Alıcı, Gauss gürültüsüyle bozularak iletilen datayı inceler; kanalın gürültü istatistikleriyle data hakkında kısmi bilgiye sahip olduğu bu noktada, iki hipotez arasında karar vermeye çalışır. Burada kısmi bilgi orijinal datanın bir nicemleyici içerisinden geçirilmesiyle elde edilir. Yani kısmi bilgi basitçe, ilişkin datanın içinde bulunduğu selene atanmış çıktı değeridir. Optimal karar kuralını ve buna karşılık gelen hata olasılığını elde ettik. Farklı seviyeler için en iyi niceleme örnekleri sunduk. Belirtilen model altında sayısal olarak hesaplanmış en iyi niceleyicinin, yakınsama ile elde edilmiş idealin altında olan başka bir niceleyicinin, Lloyd-Max ve Uniform niceleyicilerinin sele dağılımlarını ve bu dağılımlara denk gelen hata miktarları karsılastırıldı. Son olarak, önerilen model altında, Lloyd-Max niceleyicisinin yan bilgi ile sezimleme için en iyi olmamakla birlikte, yeterince iyi bir niceleyici olduğu deneysel olarak gösterilmiştir.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii
ABSTRACT iv
ÖZET v
LIST OF FIGURES vii
LIST OF TABLES viii
LIST OF SYMBOLS/ABBREVIATIONS ix
1. INTRODUCTION 1
1.1. Introduction to Detection with Side Information 1
1.2. Notation 3
1.3. Organization of the Thesis 3
2. BACKGROUND 4
2.1. Quantization Theory 4
2.2. Detection Theory 5
3. LRT and PROBABILITY OF ERROR DERIVATIONS 10
3.1. Problem Formulation 10
3.2. Likelihood Ratio Derivation 11
3.3. Probability of Error 14
4. APPROXIMATIONS 17
4.1. Approximations on LRT 18
4.2. Approximations on Probability of Error 21
5. OPTIMIZATION 23
5.0.1. Numerically calculated optimal quantizer 23
5.0.2. Numerically calculated suboptimal quantizer 27
5.0.3. Performance of the Quantizer 28
6. CONCLUSIONS 32
REFERENCES 33

LIST OF FIGURES

Figure 3.1. The Block Diagram of Detection with Side Information Problem . 11

Figure 4.1. Exact and approximate values of α_i vs. b_i for $\sigma_x = 1$. 20

Figure 5.1. Optimal quantizer bin constellation for N=3 24

Figure 5.2. Optimal quantizer bin constellation for N=4 25

Figure 5.3. Optimal quantizer bin constellation for N=5 26

Figure 5.4. Optimal quantizer bin constellation for N=6 27

Figure 5.5. Probability of error behavior of Lloyd-Max, Uniform, Optimal and Approximate (Suboptimal) quantizers for the proposed problem. . 28

Figure 5.6. Bin constellation comparison of Lloyd-Max and Optimal Quantizer for $N = 3$. 29

Figure 5.7. Bin constellation comparison of Lloyd-Max and Optimal Quantizer for $N = 4$. 30

Figure 5.8. Bin constellation comparison of Lloyd-Max and Optimal Quantizer for $N = 5$. 30

Figure 5.9. Bin constellation comparison of Lloyd-Max and Optimal Quantizer for $N = 6$. 31

LIST OF TABLES

Table 5.1. Optimal Quantizer Bin Constellation and Corresponding P_{e_T} for $N = 3$ 24

Table 5.2. Optimal Quantizer Bin Constellation and Corresponding P_{e_T} for $N = 4$ 25

Table 5.3. Optimal Quantizer Bin Constellation and Corresponding P_{e_T} for $N = 5$ 26

Table 5.4. Optimal Quantizer Bin Constellation and Corresponding P_{e_T} for $N = 6$ 27

LIST OF SYMBOLS/ABBREVIATIONS

$A(y)$	Likelihood-ratio
b_i	Left boundary of i^{th} bin
b_{i+1}	Right boundary of i^{th} bin
B_i	i^{th} bin
C_{ij}	Cost of choosing hypothesis i given hypothesis j is correct
$\mathcal{C}$	Codebook
$\mathcal{D}$	Decoder
$\mathcal{E}$	Encoder
$f^{(i)}$	i^{th} derivative of function f
H	Hypothesis
$\mathcal{I}$	Index set
$L(y)$	Likelihood-ratio
L_{z_i}	Width of z_i^{th} bin
M_{z_i}	Middle point of z_i^{th} bin
n	Channel noise
N	Quantization level
$\mathcal{N}(0, \sigma^2)$	Normal distribution function with mean 0 and variance σ^2
$P[H_i\|H_j]$	Probability of error of choosing H_i given H_j is correct
$P_F(\delta)$	False-alarm probability
$P_M(\delta)$	Miss probability
$P_D(\delta)$	Detection probability
P_{e_T}	Total probability of detection error
$P_{e_{z_0,z_1}}$	Probability of detection error for bins z_0 and z_1
Q	Quantizer
$\mathcal{Q}$	Normal distribution function
$r(\gamma)$	Overall average risk
$\mathcal{R}$	Real numbers
$\mathcal{X}$	Whole realization set
$\mathcal{X}_i$	i^{th} partition of the whole realization set

X_i	Source random variable
x_{z_i}	Realization of random variable X_i which corresponds to z_i^{th} bin of a quantizer
y	Received signal
$\mathcal{Y}_i$	Decision region i
$\bigcup$	Union operator for sets
$\forall$	For all
$\sim$	Distributed with
$\approx$	Approximately equal to
α_i	Probability of i^{th} bin
γ_B	Bayesian decision rule
π_i	Priori probability of hypothesis i
τ	Decision Threshold
τ'	Decision Threshold
DRM	Digital Rights Management
GLRT	Generalized Likelihood Ratio Test
i.i.d.	Independent Identically Distributed
LRT	Likelihood Ratio Test
w.l.o.g.	Without Loss of Generality
w.r.t.	With respect to

1. INTRODUCTION

1.1. Introduction to Detection with Side Information

In this thesis, we propose and investigate a communication theoretic approach to the content tracking with side information problem. As the widespread dissemination of multimedia signals increases, it becomes an increasingly important problem to trace the signals for various purposes such as content tracking for the protection of digital rights. Considering the easiness in modification of the digital media contents without significant losses in quality, DRM (digital rights management) techniques fail to accomplish the content tracking task. By using watermarking techniques [1], [2], [3] the aforementioned problem in DRM can be solved and content tracking task can be achieved, but the necessity of preprocessing for hiding information into the content, makes watermarking meaningless for media that is already public. In order to overcome the problems in watermarking and DRM, robust-signal hashing has been proposed. In robust-signal hashing the aim is to characterize the content by extracting a *signature* (termed as hash) of the signal, that is independent of small distortions or modifications of its content; in other word robustness to attacks. Another important constraint of robust-signal hashing formed by privacy concerns. Thus, a "content-tracker" shall not be able or it should be relatively difficult to retrieve the original content from its hash. Among these two conditions (i.e., robustness and irreversibility), the latter one makes the many-to-one mapping approach as the nature of hashing. We refer the interested reader to [4], [5], [6], [7].

In many applications, the signal to be traced is not exactly known and the receiver (also termed as the detector in the rest of the paper) only possesses partial information about the signal. An important fundamental application includes anti-piracy search via aforementioned "robust signal hashing", which amounts to the automatic detection and tracing of important signals (which are aimed to be "protected") via utilizing their features or "robust hash values". In this setup, the receiver observes a "modified" version of an original signal and is required to make a decision about the identity of

this signal; the only information that the receiver can employ throughout this detection process is the set of hash values of the "protected signal set".

Hence, we modeled the content tracking via robust signal hashing problem with detection with side information in a classical point to point communication system.

In this thesis, we consider the "binary" version of this problem; i.e., there are two signals to be protected and the receiver performs detection via binary hypothesis testing in the presence of side information (hash values of the two signals to be protected). We assumed that the transmitted signals, which are realizations of two i.i.d. random variables, are "attacked" versions of the original ones, where we model this attack as noisy channel. The hash values are produced from the original signals via a mapping that achieves "dimensionality reduction" because of efficiency and privacy issues. We assume that the dimensionality reduction is achieved via scalar quantization (i.e., the hash values are the reproduction values of the original signals). Furthermore, the noise introduced by the channel is zero mean unit variance Gaussian random variable and independent from the transmitted signals.

A detection theory based investigation involving partial information deserves to be mentioned here. In [8], the authors focused on the distributed detection for wireless sensor networks, in which the from all sensors the partial information that is obtained via quantization, is send to a fusion center, where the decision process is performed. Another related work is the classical Wyner-Ziv problem [9] in which the aim is in presence of correlated signals about the source and subject to a distortion constraint, to find out the achievable rate of compression. In other two related works [10] and [11], Koval *et. al.* investigated the robust perceptual hashing from information theoretic point of view. Finally, in [12], optimal GLRT rule and corresponding probability of error and the optimal linear transform matrix (used for dimensionality reduction) that minimizes the worst case probability of error, was analytically derived. Considering the formulation of the problem the work in [12] is closest one except in our case the receiver also knows the probability distribution defined on the observation space of the source.

1.2. Notation

In this section, we provide the notation used throughout the thesis. Capital letters denote random variables (i.e. X_0); regular letters denote a realization (i.e., x_{z_0}) of random variable. A quantizer is denoted as $Q(.)$, where the reproduction values of a realization of a random variable is defined as $z_i = Q(x_i)$, where $i = 0, 1$. Thus, x_{z_i} refers to a realization of X_i that corresponds to the z_i'th bin of the quantizer where $z_i \in \{0, 1, \ldots, N-1\}$ for a N-level quantizer. The $\mathcal{Q}(.)$ is the well-known normal distribution function, i.e., $\mathcal{Q}(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt$. Lastly, $f^{(i)}(.)$ denotes the i^{th} derivative of real function $f(.)$.

1.3. Organization of the Thesis

In Chapter 2 we will give brief introduction about Quantization and Detection theory, namely in Section 2.1 and Section 2.2, respectively. Chapter 3 includes the problem formulation and the main derivations, the problem formulation is given in Section 3.1, LRT and probability of error derivations are given in Section 3.2 and Section 3.3, respectively. The approximate LRT is presented in Section 4.1 and approximate probability of error is presented in Section 4.2 which constitute Chapter 4. The performance results of investigated and derived quantizers are given in Chapter 5. Thesis ends with discussions and conclusions given in Chapter 6.

2. BACKGROUND

In this chapter, the basic concepts about Quantization and Detection theory that used throughout the thesis, will be presented. This chapter is intended for the reader that does not have the knowledge about these aforementioned theories.

2.1. Quantization Theory

Quantization is a non-linear mapping of a continuous range of values to a smaller and finite set of values. It can be viewed as a predetermined rule of approximation of a continuous number set. For example, N-level scalar quantizer is a mapping from real line to a codebook with size N, i.e., $Q : \mathcal{R} \rightarrow \mathcal{C}$ where the codebook $\mathcal{C} \stackrel{\triangle}{=} \{y_0, y_1, \ldots, y_{N-1}\} \subset \mathcal{R}$. Here y_i for $i = \{0, 1, \ldots, N-1\}$ is called the *output levels* or *reproduction values* [14].

A N-level quantizer (i.e., the aforementioned mapping Q) can be written as the concatenation of two successive mappings which are called *encoder* and *decoder* respectively. Encoder, $\mathcal{E}$, is a mapping defined as, $\mathcal{E} : \mathcal{R} \rightarrow \mathcal{I}$ where $\mathcal{I} = \{0, 1, 2, \ldots, N-1\}$ and decoder, $\mathcal{D}$, is a mapping defined as, $\mathcal{D} : \mathcal{I} \rightarrow \mathcal{C}$, where $\mathcal{C}$ is the codebook. Considering the quantizer level is N, a mapping from the real line to a set $\mathcal{I}$ is simply a partition of the real line into N, and again simply $\mathcal{I}$ is the set of indices of these partitions. The rest of the work, that is matching the reproduction values with corresponding partitions of real line, is left to decoder mapping. Thus A quantized value of b, i.e., $Q(b)$ can be written as $\mathcal{D}(\mathcal{E}(b))$ [14].

In quantization lingo the aforementioned partitions are referred as *cells* or *bins*. A cell or bin, B_i, is defined as $B_i \stackrel{\triangle}{=} \{x \in \mathcal{R} | Q(x) = y_i\}$, moreover $\bigcup_i B_i = \mathcal{R}$ and for $i \neq j$ $\mathcal{B}_i \bigcap \mathcal{B}_j = \emptyset$, where $i = \{0, 1, \ldots, N-1\}$. Hence, by using these let's make a more precise definition of quantizer.

Definition 2.1.1 *A quantizer is uniquely described by its bin constellation* $\{B_i\}_{i=0}^{N-1}$ *and the corresponding reproduction values* $\{y_i\}_{i=0}^{N-1}$. *Here,* $B_i = (b_i, b_{i+1})$, *and the boundary points* b_i *has the following properties,* $b_i \in \mathcal{R}$, $b_i < b_{i+1}$, $b_0 \to -\infty$ *and* $b_N \to \infty$, *for* $i = \{0, 1, \ldots, N-1\}$.

Next, we will briefly introduce some well-known quantizers.

Uniform Quantizer: A uniform quantizer is a regular quantizer in which

- The bins are equally spaced, i.e., $(b_{i+1} - b_i) = C$ for $i = \{1, 2, \ldots, N-2\}$, where C is constant.
- The reproduction values are the middle points of the bins, i.e., $y_i = \frac{(b_i + b_{i+1})}{2}$ for $i = \{1, 2, \ldots, N-2\}$.

Lloyd-Max Quantizer: Lloyd-Max Quantizer is a regular quantizer that minimizes the variance of quantization error. The conditions of Lloyd-Max quantizer are given below, where $i \in \{0, 1, \ldots, N-1\}$.

- The bins are formed according to $b_i = \frac{y_{i-1} + y_i}{2}$.
- The reproduction values are formed according to $\dfrac{\int_{b_i}^{b_{i+1}} x f_X(x) dx}{\int_{b_i}^{b_{i+1}} f_X(x) dx}$.

2.2. Detection Theory

In telecommunication, detection theory is the area of study that deals with the processing of information-bearing signals for the purpose of extracting information from them [13]. Since it involves the concepts that helps making a decision under uncertainty (i.e., unknown variables, randomness ...), it is also used in many other fields than telecommunications like business management, finance, psychology.

The idea of decision making under uncertainty makes easier to understand detection theory. Let's think about an ordinary decision that everyone needs to do at

least once in his/her life, choosing a career (or job). Depending how accurate choice one wants to make, one will try to answer questions like,

- Which jobs are available in the market?
- What is the availability (occurrence rate) of these jobs?
- How much will be my salary?
- What is my concern (money, happiness,...)?
- If I choose a wrong career, what will be the risk (or cost)?
- etc.

Obviously, the choice will depend on the answers of these questions. The more question that you are able to answer, means more accurate decision. Basically, the procedures in detection theory are formed depending on the answers of such questions.

Let's turn back to signal detection lingo. Suppose we have two hypothesis, namely H_0 and H_1, and we are looking for a decision rule. Obviously, as every decision we made, choosing one of these hypothesis has also a cost. The cost of choosing H_i when H_j is the correct choice, is defined as C_{ij} where $i, j = \{0, 1\}$. Next, we will mention some well-known criterions about making a decision.

Bayesian Hypothesis Testing: Consider that we are trying to choose the distribution of a random variable X from two possible distributions, P_0 and P_1. Thus our hypotheses will be in the form of, $H_0 : X \sim P_0$ and $H_1 : X \sim P_1$. Next, assume that the probability of occurrence of these hypotheses (i.e., *priori probabilities*) H_0, H_1 are π_0,π_1 respectively. Think that this information is obtained statistically from the past occurrences of the events (i.e., X will have the probability distribution P_0 with probability π_0 and vice versa). By using these information, our ultimate goal will be to design a decision rule that will minimize the average cost.

In order to define such a decision rule we need to partition the whole possible realization set (or observation set), let's say $\mathcal{X}$, in to two, i.e., $\mathcal{X}_0$ and $\mathcal{X}_1$. Note that

$\mathcal{X}_0 \bigcup \mathcal{X}_1 = \mathcal{X}$. Thus the decision rule will be,

$$\delta_B(x) = \begin{cases} 1 & \text{, if } x \in \mathcal{X}_0 \\ 0 & \text{, if } x \in \mathcal{X}_1 \end{cases}, \tag{2.1}$$

where x is any realization of the random variable X.

Remark 2.2.1 *The observation space partitions $\mathcal{X}_0$ and $\mathcal{X}_1$ are called decision regions. Thus if the observation is in $\mathcal{X}_i$ then H_i is chosen for $i = \{0, 1\}$.*

The overall average cost (or risk), $r(\delta)$ given by,

$$\begin{aligned} r(\delta) &= \sum_{j=0}^{1} \pi_j \left[C_{0j} Pr[H_0|H_j] + C_{1j} Pr[H_1|H_j] \right], & (2.2) \\ &= \sum_{j=0}^{1} \pi_j \left[C_{0j}(1 - Pr[H_1|H_j]) + C_{1j} Pr[H_1|H_j], \right] & (2.3) \\ &= \sum_{j=0}^{1} \pi_j C_{0j} + \sum_{j=0}^{1} \pi_j (C_{1j} - C_{0j}) Pr[H_1|H_j], & (2.4) \\ &= \sum_{j=0}^{1} \pi_j C_{0j} + \int_{\mathcal{X}_1} \left[\sum_{j=0}^{1} \pi_j (C_{1j} - C_{0j}) \right] dx, & (2.5) \end{aligned}$$

where $Pr[H_1|H_j] = \int_{\mathcal{X}_1} p_j(x)dx$ is the probability of choosing H_1 given H_j is correct and note that $Pr[H_1|H_j] + Pr[H_0|H_j] = 1$, for $j = 0, 1$.

Note that, we are after the decision rule that will minimize the average cost given in (2.5). Since the first term in (2.5) is always positive, minimizing the second term will minimize the overall cost. Thus, the cost will be at its minimum if we partition the realization space as,

$$\mathcal{X}_1 = \left\{ y \in \mathcal{X} | \sum_{j=0}^{1} \pi_j (C_{1j} - C_{0j}) \leq 0 \right\}. \tag{2.6}$$

It is quite reasonable if we make an assumption stating $C_{10} > C_{00}$ and $C_{01} > C_{11}$, since the cost of making a wrong decision should be higher than making a correct one. Hence, the decision rule given in (2.1) will minimize the average cost, if we design it as,

$$\delta_B(x) = \begin{cases} 1 & \text{, if } L(x) \geq \tau \\ 0 & \text{, if } L(x) < \tau \end{cases}, \tag{2.7}$$

where the *likelihood-ratio* is defined as $L(y) \stackrel{\triangle}{=} \frac{p_1(x)}{p_0(x)}$ and the *threshold* is defined as $\tau \stackrel{\triangle}{=} \frac{\pi_0(C_{10}-C_{00})}{\pi_1(C_{01}-C_{11})}$.

Remark 2.2.2 *Note that, the average cost given in* (2.5) *can be interpreted as the average probability of error since* $P[H_i|H_j]$ *is the conditional probability of error for* $i \neq j$*. Thus, the aforementioned likelihood-ratio test approach is a minimum-probability of error decision scheme [13].*

Throughout the thesis we will use the *uniform cost assignment* given by,

$$C_{ij} = \begin{cases} 1 & \text{, if } i \neq j \\ 0 & \text{, if } i = j \end{cases} \tag{2.8}$$

As we mentioned before, there are other criterions depending of the definition of and assumptions made in problems. Since, we used the Bayesian Hypothesis Testing for our problem, the other detection criterions will be briefly presented in this chapter to just give an interpretation to the reader.

Minimax Hypothesis Testing: In Bayesian Hypothesis Testing we assumed that, we know the priori probabilities, namely π_0 and π_1. Suppose that this is not the case. Since we do not know the priori probabilities we can not expect a single decision rule will minimize the expected cost for all possible prior distributions, thus Bayesian criterion will not be meaningful [13]. Hence, in Minimax-Pearson Hypothesis Testing

we first find the maximum cost over all possible prior probabilities and then minimize this cost over all possible decision rules to find the optimal one. More precisely,

$$\min_{\delta} \quad \{\max_{\pi_j} \quad r(\pi_j, \delta)\}, \tag{2.9}$$

where $r(\pi_j, \delta)$ was defined in (2.2), except here the priori probabilities, are unknown (i.e., a variable) [13].

Neyman-Pearson Hypothesis Testing: Before we explain the Neyman-Pearson criteria we will give some definitions. *False Alarm* (or *Type I error*) is defined as choosing H_1 given that H_0 is correct and *Miss* (or *Type II error*) is defined as choosing H_0 given that H_1 is correct. The terms "miss" and "false Alarm" originated from target detection radar problems while H_0 states there is no target and H_1 states there is a target. Given a decision rule δ, $P_F(\delta)$, $P_M(\delta)$ and $P_D(\delta) \stackrel{\triangle}{=} 1 - P_M(\delta)$ are defined as *false-alarm probability* , *miss probability* and *detection probability.*

Any decision rule designed for binary detection problems will define a trade-off between $P_M(\delta)$ and $P_D(\delta)$ (i.e., Minimax and Bayesian Hypothesis testings are two examples of trade-offs) and Neyman-Pearson criterion is yet another one of these possible trade-offs. The criterion is defined as,

$$\min_{\delta} \quad P_M(\delta) \quad \text{with constraint} \quad P_F(\delta) \leq \alpha, \tag{2.10}$$

where α is known as "level" or "significance level" of the test [13].

3. LRT and PROBABILITY OF ERROR DERIVATIONS

3.1. Problem Formulation

In this problem, we are after the quantizer, which is optimal in the probability of error sense, for the binary hypothesis testing problem in which, two realizations from two i.i.d. (normal distribution) random variables are transmitted through a noisy channel (or attacked during the transmission) and tried to be detected with the additional information of quantizer bin constellation.

To be more precise, X_0 and X_1 are i.i.d random variables with distribution, $X_i \sim \mathcal{N}(0, \sigma_x^2)$ for $i \in \{0, 1\}$. We refer, x_0 and x_1 as the realizations of random variables X_0 and X_1, respectively.

The bin index is defined as $z_i \stackrel{\triangle}{=} \mathcal{E}(x_i)$ for $i = \{0, 1\}$, where $\mathcal{E}(.)$ is the encoder function defined in section 2.1. We assumed that, the bin indexes of the random variable realizations are transmitted via noiseless and secure channel to the detector side, i.e., z_i for $i = \{0, 1\}$ is known by the receiver.

Remark 3.1.1 *Note that, the decoder block, $\mathcal{D}(.)$, of a quantizer is nothing but a $1-1$ mapping from bin index set to reproduction value set, which states that the nonlinear nature of a quantizer comes from the encoder block (function). Thus, for our problem, w.l.o.g. we can refer the encoder function as the whole quantizer.*

In our model, we assumed that, the realizations of random variables are transmitted through a noisy channel (or attacked during transmission), where the noise is additive Gaussian with distribution $\mathcal{N}(0, \sigma_n^2)$. Thus, we form our hypothesis as follows,

$$H_0 : Y = x_0 + N \quad \text{and} \quad H_1 : Y = x_1 + N. \tag{3.1}$$

We assume that the prior probabilities of these hypothesis are equal, i.e. $Pr(H_0) = \pi_0 = Pr(H_1) = \pi_1 = 1/2$. This assumption can be reasoned by the idea of introducing an ideal selection mechanism (switch) at transmitter side, this mechanism will choose either one transmitter or the other with probability 1/2.

The block diagram representation of the communication schema is illustrated in Figure 3.1 given below.

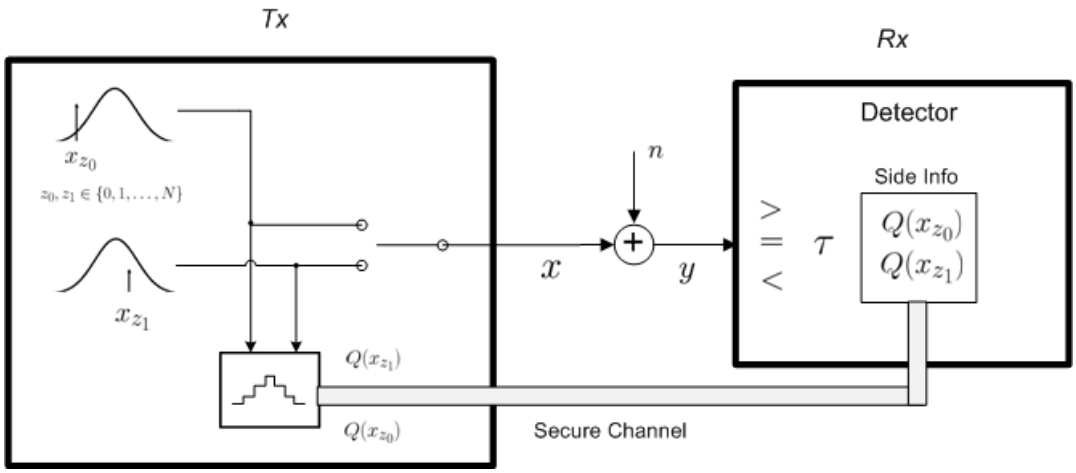

Figure 3.1. The Block Diagram of Detection with Side Information Problem

3.2. Likelihood Ratio Derivation

In this section we will derive the decision rule by well-known Bayesian Hypothesis testing approach, where the decision regions $\mathcal{Y}_0 \stackrel{\triangle}{=} \{y|L(y) < \tau\}$ and $\mathcal{Y}_1 \stackrel{\triangle}{=} \{y|L(y) > \tau\}$ are defined according to the likelihood ratio rule (LRT) which is given as follows,

$$L(y) \stackrel{\triangle}{=} \frac{p(y|H_1, z_1)}{p(y|H_0, z_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \tau \stackrel{\triangle}{=} \frac{\pi_0}{\pi_1} = 1, \tag{3.2}$$

under the uniform cost assumption.

Definition 3.2.1 *For a N-level quantizer, left and right bin (or cell) boundaries are defined as,* b_k *and* b_{k+1} *respectively, for the* k^{th} *bin from left, where* $k = \{0, 1, \ldots, N-1\}$, $\forall k \quad b_k, b_{k+1} \in \mathcal{R}$, $b_0 \to -\infty$, $b_N \to \infty$ *and* $\forall k \quad b_{k+1} > b_k$.

Remark 3.2.1 *Note that, since we refer the bin indexes as z_i for $i = \{0, 1\}$, b_{z_0} (resp. b_{z_1}) is the left boundary of z_0th (resp. z_1th) bin, that a realization of X_0 (resp. X_1) is in. Similarly, we denote the bin indexes as z_i for $i = \{0, 1\}$, b_{z_0+1} (resp. b_{z_1+1}) is the right boundary of z_0th (resp. z_1th) bin, that a realization of X_0 (resp. X_1) is in.*

Remark 3.2.2 *In Lemma 3.2.1 and in the rest of the paper, w.l.o.g. we will use x instead of x_i's since their distributions are same. One can understand which random variable realization is being mentioned in the following equations, from the index of z_i for $i = \{0, 1\}$.*

In order to derive the LRT given in equation (3.2), first we need to derive $p(y|H_i, z_i)$.

Lemma 3.2.1

$$p(y|H_i, z_i) = \frac{1}{\alpha_i 2\pi\sigma_x\sigma_n} \int_{b_{z_i}}^{bz_i+1} e^{-\frac{(y-x)^2}{2\sigma_n^2} - \frac{x^2}{2\sigma_x^2}} dx, \tag{3.3}$$

$$= \frac{1}{\alpha_i\sqrt{2\pi}\sqrt{\sigma_x^2+\sigma_n^2}} \quad e^{-\frac{y^2}{2(\sigma_x^2+\sigma_n^2)}}, \times \left[\mathcal{Q}\left(\frac{b_{z_i} - \gamma y}{\sqrt{\sigma_n^2\gamma}}\right) - \mathcal{Q}\left(\frac{b_{z_i+1} - \gamma y}{\sqrt{\sigma_n^2\gamma}}\right)\right], \tag{3.4}$$

where $\alpha_i \stackrel{\triangle}{=} \mathcal{Q}\left(\frac{b_{z_i}}{\sigma_x}\right) - \mathcal{Q}\left(\frac{b_{z_i+1}}{\sigma_x}\right)$ *and* $\gamma \stackrel{\triangle}{=} \frac{\sigma_x^2}{\sigma_x^2+\sigma_n^2}$ *for* $i = 0, 1$.

Proof: First, note that

$$p(y|H_i, z_i) = \int_{\mathcal{X}} p(y, x_i|H_i, z_i)dx. \tag{3.5}$$

Conditioned on H_i, $[y_i \leftrightarrow x_i \leftrightarrow z_i]$ forms a Markov chain. Hence conditioned on H_i we have,

$$p(y, x_i|H_i, z_i) = p(y|x_i, H_i, z_i)p(x_i|z_i) = p(y|x_i, H_i)p(x_i|z_i),$$

where $p(y|x_i, H_i) \sim \mathcal{N}(x, \sigma_n^2)$, $p(x_i|z_i) = \begin{cases} \frac{w(x)}{\alpha_k} & , \text{ if } x \in [b_{z_i}, b_{z_i+1}] \\ 0 & , \quad otherwise \end{cases}$ and $w(x) \sim \mathcal{N}(0, \sigma_x^2)$, for $i = 0, 1$.

Hence the proof follows for (3.3).

Next, for the exponential term in (3.3), Define,

$$\begin{aligned} A &\triangleq -\frac{(y-x)^2}{2\sigma_n^2} - \frac{x^2}{2\sigma_x^2}, \\ &= -\frac{y^2}{2\sigma_n^2} + \frac{2xy}{2\sigma_n^2} - x^2\frac{\sigma_n^2+\sigma_x^2}{2\sigma_n^2\sigma_x^2}. \end{aligned}$$

$$\begin{aligned} 2A\frac{\sigma_x^2\sigma_n^2}{\sigma_n^2+\sigma_x^2} &= -y^2\frac{\sigma_x^2}{\sigma_x^2+\sigma_n^2} + 2xy\frac{\sigma_x^2}{\sigma_x^2+\sigma_n^2} - x^2, \\ &= -y^2\frac{\sigma_x^2}{\sigma_x^2+\sigma_n^2} + y^2\left(\frac{\sigma_x^2}{\sigma_x^2+\sigma_n^2}\right)^2 - \left[y^2\left(\frac{\sigma_x^2}{\sigma_x^2+\sigma_n^2}\right)^2 - 2xy\frac{\sigma_x^2}{\sigma_x^2+\sigma_n^2} + x^2\right] \\ &= -y^2\frac{\sigma_x^2\sigma_n^2}{(\sigma_x^2+\sigma_n^2)^2} - \left(x - y\frac{\sigma_x^2}{\sigma_x^2+\sigma_n^2}\right)^2. \\ 2A\sigma_n^2\gamma &= -y^2\gamma(1-\gamma) - (x-y\gamma)^2. \\ A &= -\frac{y^2}{2(\sigma_x^2+\sigma_n^2)} - \frac{(x-\gamma y)^2}{2\sigma_n^2\gamma}. \end{aligned}$$

Hence, we have,

$$\begin{aligned} p(y|H_i, z_i) &= \frac{1}{\alpha_i 2\pi\sigma_x\sigma_n}\int_{b_{z_i}}^{bz_i+1} e^{-\frac{(y-x)^2}{2\sigma_n^2} - \frac{x^2}{2\sigma_x^2}}dx \\ &= \frac{1}{\alpha_i\sqrt{2\pi}\sqrt{\sigma_x^2+\sigma_n^2}} \quad e^{-\frac{y^2}{2(\sigma_x^2+\sigma_n^2)}}\int_{b_{z_i}}^{b_{z_i+1}} \frac{1}{\sqrt{2\pi}\sqrt{\sigma_n^2\gamma}} e^{-\frac{(x-\gamma y)^2}{2\sigma_n^2\gamma}}dx, \\ &= \frac{1}{\alpha_i\sqrt{2\pi}\sqrt{\sigma_x^2+\sigma_n^2}} \quad e^{-\frac{y^2}{2(\sigma_x^2+\sigma_n^2)}}\left[\mathcal{Q}\left(\frac{b_{z_i}-\gamma y}{\sqrt{\sigma_n^2\gamma}}\right) - \mathcal{Q}\left(\frac{b_{z_i+1}-\gamma y}{\sqrt{\sigma_n^2\gamma}}\right)\right], \end{aligned}$$

where $\gamma = \frac{\sigma_x^2}{\sigma_x^2+\sigma_n^2}$. Hence the proof follows for (3.4). □

Theorem 3.2.1 *The Bayes rule is given by,*

$$\delta_B(y) = \begin{cases} 1 & , \text{ if } \mathcal{A}(y) \geq \tau' \\ 0 & , \text{ else} \end{cases}, \tag{3.6}$$

where

$$A(y) \stackrel{\triangle}{=} \frac{\mathcal{Q}\left(\frac{b_{z_1}-\gamma y}{\sqrt{\sigma_n^2\gamma}}\right) - \mathcal{Q}\left(\frac{b_{z_1+1}-\gamma y}{\sqrt{\sigma_n^2\gamma}}\right)}{\mathcal{Q}\left(\frac{b_{z_0}-\gamma y}{\sqrt{\sigma_n^2\gamma}}\right) - \mathcal{Q}\left(\frac{b_{z_0+1}-\gamma y}{\sqrt{\sigma_n^2\gamma}}\right)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\mathcal{Q}\left(\frac{b_{z_1}}{\sigma_x}\right) - \mathcal{Q}\left(\frac{b_{z_1+1}}{\sigma_x}\right)}{\mathcal{Q}\left(\frac{b_{z_0}}{\sigma_x}\right) - \mathcal{Q}\left(\frac{b_{z_0+1}}{\sigma_x}\right)} = \frac{\alpha_1}{\alpha_0} \stackrel{\triangle}{=} \tau'. \tag{3.7}$$

Proof: From (3.2) and Lemma 3.2.1, the proof follows. □

Remark 3.2.3 *After Theorem 3.2.1, the decision regions can be rewritten as,*

$$\mathcal{Y}_1 \stackrel{\triangle}{=} \{y | A(y) \geq \tau' \} and \mathcal{Y}_0 \stackrel{\triangle}{=} \{y | A(y) \leq \tau' \},$$

where $A(y)$ and τ' is defined in Theorem 3.2.1.

Next, we will continue with the derivation of the probability of detection error emanating from the Bayes rule given in Theorem 3.2.1.

3.3. Probability of Error

In this section, first we will derive the probability of detection error given any two quantizer bins. After that, we will generalize and present the total probability of error expression. Since the former one is a direct consequence of the decision rule, we will give it as a corollary.

Definition 3.3.1 *The probability of error for choosing,*

- H_1 *given that* H_0 *is correct, i.e., Type-I error or false alarm probability, is defined as,*

$$P_e(H_0) \stackrel{\triangle}{=} \int_{\mathcal{Y}_1} p(y|H_0, z_0)dy.$$

- H_0 *given that* H_1 *is correct, i.e., Type-II error or miss probability, is defined as,*

$$P_e(H_1) \stackrel{\triangle}{=} \int_{\mathcal{Y}_0} p(y|H_1, z_1)dy.$$

Corollary 3.3.1 *The probability of detection error given any two quantizer bins is given by,*

$$P_{e_{z_0,z_1}} \stackrel{\triangle}{=} \pi_0 P_e(H_0) + \pi_1 P_e(H_1) = \frac{1}{2}\left[P_e(H_0) + P_e(H_1)\right], \tag{3.8}$$

where $P_e(H_j)$ *for* $j = \{0, 1\}$ *are Type-I and Type-II errors, which are defined in Definition 3.3.1.*

Until now, we derived the Bayesian detection rule given in Theorem 3.2.1. Moreover, we also have the probability of error for any given two quantizer bins mentioned in Corollary 3.3.1. However, since optimization over boundaries of two quantizer bins can not give us the whole constellation, it is necessary to include all bin boundaries in the cost function. Hence, we will form the cost function by averaging out the probability of error for any given two quantizer bins over all possible bin pairs.

From Corollary 3.3.1 we have a closed form expression for the probability of error derived for any two quantizer bins (i.e., $P_{e_{z_0,z_1}}$), as we mentioned above next we need to average out this expression over z_0, z_1 in order to derive the cost function, which

will be referred as the total probability of error. Hence we have,

$$P_{e_T} = \sum_{z_0=0}^{N-1} \sum_{z_1=0}^{N-1} \alpha_{z_0} \alpha_{z_1} P_{e_{z_0,z_1}}, \tag{3.9}$$

where $\alpha_{z_i} = \mathcal{Q}\left(\frac{b_{z_i}}{\sigma_x}\right) - \mathcal{Q}\left(\frac{b_{z_i+1}}{\sigma_x}\right)$, $P_e(H_j)$ for $j = \{0, 1\}$ is defined in Definition 3.3.1 and $\mathcal{Y}_j$ for $j = \{0, 1\}$ is given in Remark 3.2.3.

4. APPROXIMATIONS

Since it is hard to handle the $\mathcal{Q}(.)$ function differences given in Theorem 3.2.1 for determining the regions $\mathcal{Y}_0$ and $\mathcal{Y}_1$, we will apply approximations to the $\mathcal{Q}(.)$ function differences. In the first part we will apply approximations only to the LRT. Generally, this approach is not quite logical, normally if an approximation is going to applied than it should be applied to whole derivations. However, in our case it will mean we will try to find a optimal (in the probability of error sense) quantizer for a suboptimal decision rule which will lead a suboptimal quantizer constellation, and it is worth to investigate the behavior of such solution.

In the second part we will apply approximations to all $\mathcal{Q}(.)$ function differences in the cost function P_{e_T} and LRT, in order to obtain an easy to handle cost function so that an analytical result may be found under some mild assumptions.

In order to ease the derivations, we will use zeroth order Taylor series approximations for the $\mathcal{Q}(.)$ function differences.

Lemma 4.0.1 *If the boundaries are sufficiently close, the zeroth order Taylor series approximation of a definite integral is given by*

$$\int_b^a f(t)dt \approx f(c)(b-a),$$

where c, such that $a \leq c \leq b$, is an arbitrary point that the function $f(.)$ is expanded around.

Proof: From the well known Taylor series expansion we know that,

$$f(t) = \sum_{l=0}^{\infty} \frac{f^{(l)}(c)}{l!}(t-c)^l = f(c) + f^{(1)}(c)(t-c) + \frac{f^{(2)}(c)}{2}(t-c)^2 + \dots \qquad (4.1)$$

Since the terms other than $(t-c)^l$ in equation (4.1)are constant,

$$\int_a^b (t-c)^l dt = \frac{(t-c)^{l+1}}{l+1}\Big|_{c=a}^{b}. \tag{4.2}$$

Hence we have,

$$\int_b^a f(t)dt = \sum_{l=0}^{\infty} \frac{f^{(l)}(c)}{(l+1)!}[(b-c)^{l+1} - (a-c)^{l+1}]. \tag{4.3}$$

For $l = 0$ in equation (4.3) we have the zeroth order approximation, hence the proof follows. □

Remark 4.0.1 *If the point is chosen to be the middle point of the integral boundaries that the function will be expanded (i.e., $c = \frac{a+b}{2}$), then* (4.3) *becomes,*

$$\int_b^a f(t)dt = \sum_{l=0}^{\infty} \frac{f^{(l)}(c)}{(l+1)!} \left(\frac{b-a}{2}\right)^{l+1} [1-(-1)^{l+1}].$$

, which means all the odd terms in Taylor expansion vanishes. Hence the first order approximation ($l = 1$) reduces to zeroth order approximation ($l = 0$).

4.1. Approximations on LRT

After applying zeroth order (or first order with choosing the middle point as the one that the function is expanded around) Taylor series approximation given in Lemma 4.0.1 to the Bayes rule in Theorem 3.2.1 we will obtain an approximate (suboptimal) Bayes rule.

Theorem 4.1.1 *The zeroth order Taylor series approximate of the Bayes rule given*

in Theorem 3.2.1 is given by,

$$\delta_B(y) = \begin{cases} 1 & , \text{ if } y \geq \frac{M_{z_0}+M_{z_1}}{2} \\ 0 & , \text{ else} \end{cases},$$

where $M_{z_i} \triangleq \frac{b_{z_i}+b_{z_i+1}}{2}$ *(i.e., the middle point of the* i^{th} *bin).*

Proof: By introducing lemma 4.0.1 in Theorem 3.2.1, and choose c as the middle point of the boundaries a and b, i.e. $c = \frac{a+b}{2}$, in lemma 4.0.1, then we have,

$$\mathcal{Q}\left(\frac{b_{z_i}-\gamma y}{\sqrt{\sigma_n^2\gamma}}\right) - \mathcal{Q}\left(\frac{b_{z_i+1}-\gamma y}{\sqrt{\sigma_n^2\gamma}}\right) \approx \frac{(b_{z_i+1}-b_{z_i})}{\sqrt{2\pi}\sqrt{\sigma_n^2\gamma}} e^{-\frac{(M_{z_i}-\gamma y)^2}{2\sigma_n^2\gamma}}, \tag{4.4}$$

$$\alpha_i = \mathcal{Q}\left(\frac{b_{z_i}}{\sigma_x}\right) - \mathcal{Q}\left(\frac{b_{z_i+1}}{\sigma_x}\right) \approx \frac{(b_{z_i+1}-b_{z_i})}{\sqrt{2\pi}\sigma_x} e^{-\frac{M_{z_i}^2}{2\sigma_x^2}}, \tag{4.5}$$

where $i = 0, 1$.

Hence, by introducing (4.4) and (4.5), the LRT given in Theorem 3.2.1 becomes,

$$\frac{\frac{(b_{z_1+1}-b_{z_1})}{\sqrt{2\pi}\sqrt{\sigma_n^2\gamma}} e^{-\frac{(M_{z_1}-\gamma y)^2}{2\sigma_n^2\gamma}}}{\frac{(b_{z_0+1}-b_{z_0})}{\sqrt{2\pi}\sqrt{\sigma_n^2\gamma}} e^{-\frac{(M_{z_0}-\gamma y)^2}{2\sigma_n^2\gamma}}} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\frac{(b_{z_1+1}-b_{z_1})}{\sqrt{2\pi}\sigma_x} e^{-\frac{M_{z_1}^2}{2\sigma_x^2}}}{\frac{(b_{z_0+1}-b_{z_0})}{\sqrt{2\pi}\sigma_x} e^{-\frac{M_{z_0}^2}{2\sigma_x^2}}}, \tag{4.6}$$

$$e^{\left(\frac{(M_{z_0}-\gamma y)^2-(M_{z_1}-\gamma y)^2}{2\sigma_n^2\gamma}\right)} \underset{H_0}{\overset{H_1}{\gtrless}} e^{\left(\frac{M_{z_0}^2-M_{z_1}^2}{2\sigma_x^2}\right)}, \tag{4.7}$$

$$\frac{(M_{z_0}-\gamma y)^2-(M_{z_1}-\gamma y)^2}{2\sigma_n^2\gamma} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{M_{z_0}^2-M_{z_1}^2}{2\sigma_x^2}, \tag{4.8}$$

$$(M_{z_0}+M_{z_1}-2\gamma y)(\sigma_x^2+\sigma_n^2) \underset{H_1}{\overset{H_0}{\gtrless}} (M_{z_0}+M_{z_1})\sigma_n^2, \tag{4.9}$$

$$y \underset{H_0}{\overset{H_1}{\gtrless}} \frac{M_{z_0}+M_{z_1}}{2}. \tag{4.10}$$

where the first simplification (applied to (4.6)) is valid and obvious since $b_{z_i+1} - b_{z_i} > 0$ for $i = 0, 1$. Next, we took the logarithm of both sides of inequality to obtain (4.8) from (4.7). Then, we used $\gamma = \frac{\sigma_x^2}{\sigma_x^2+\sigma_n^2}$ and the assumption of $M_{z_1} > M_{z_0}$ to obtain (4.9) from (4.8). Hence the proof. □

Obviously, the zeroth order Taylor series approximation is a loose approximation. However, for higher levels of quantizers, since the bin widths will be close enough, the approximation will be better. The behavior of zeroth order Taylor series approximation for is given below in Figure 4.1. The Figure 4.1 illustrates the exact and approximate values of $\mathcal{Q}(.)$ function difference given in (4.5), for several N-level uniform quantizers, where we assumed $\sigma_x = 1$.

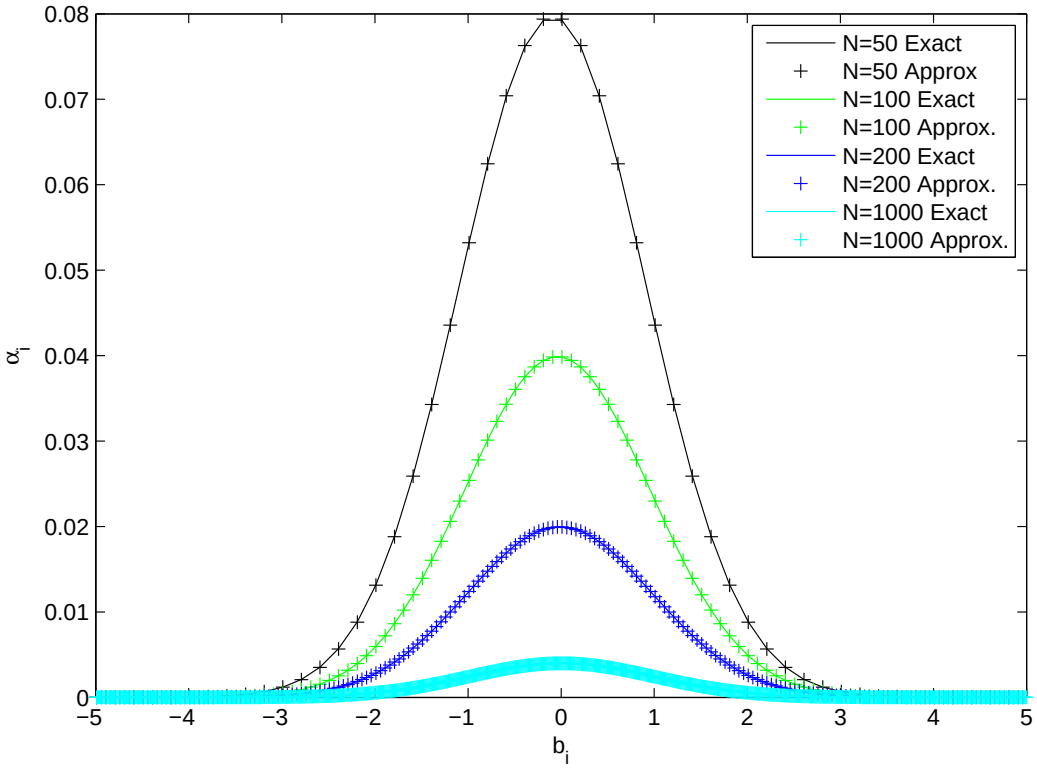

Figure 4.1. Exact and approximate values of α_i vs. b_i for $\sigma_x = 1$.

4.2. Approximations on Probability of Error

In this section we will apply the approximation on $\mathcal{Q}(.)$ function differences in probability of error expression given in equation 3.9 to obtain an easy to handle cost function for optimization part.

Theorem 4.2.1

$$P_{e_{z_0,z_1}} = \mathcal{Q}\left(\frac{M_{z_1} - M_{z_0}}{2\sigma_n}\right), \tag{4.11}$$

where $M_{z_i} = \frac{b_{z_i}+b_{z_{i+1}}}{2}$.

Proof: Introducing (4.4) and (4.5) in $p(y|H_i, z_i)$ given in Lemma 3.2.1, i.e., (3.4), we have,

$$p(y|H_i, z_i) \approx \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{y^2}{2(\sigma_x^2+\sigma_n^2)} + \frac{M_{z_i}^2}{2\sigma_x^2} - \frac{(M_{z_1}-\gamma y)^2}{2\sigma_n^2\gamma}}, \tag{4.12}$$

for $i = 0, 1$. Define,

$$\begin{aligned}
B &\triangleq \frac{-1}{2}\left[\frac{y^2}{\sigma_x^2+\sigma_n^2} - \frac{M_{z_i}^2}{\sigma_x^2} + \frac{(M_{z_1}-\gamma y)^2}{\sigma_n^2\gamma}\right], \\
&= \frac{-1}{2}\left[\frac{y^2}{\sigma_x^2+\sigma_n^2} - \frac{M_{z_i}^2}{\sigma_x^2} + \frac{(M_1-\gamma y)^2(\sigma_x^2+\sigma_n^2)}{\sigma_n^2\sigma_x^2}\right], \\
&= \frac{-1}{2}\left[\frac{y^2}{\sigma_x^2+\sigma_n^2} + \frac{1}{\sigma_n^2\sigma_x^2}\left[(M_{z_i}^2 - 2\gamma y M_{z_i} + \gamma^2 y^2)(\sigma_x^2+\sigma_n^2) - \sigma_n^2 M_{z_i}^2\right]\right], \\
&= \frac{-1}{2}\left[\frac{y^2}{\sigma_x^2+\sigma_n^2} + \frac{1}{\sigma_n^2\sigma_x^2}\left[M_{z_i}^2\sigma_x^2 - 2\gamma y M_{z_i}(\sigma_x^2+\sigma_n^2) + \gamma^2 y^2(\sigma_x^2+\sigma_n^2)\right]\right], \\
&= \frac{-1}{2}\left[\frac{y^2\gamma}{\sigma_x^2} + \frac{1}{\sigma_n^2}\left[M_{z_i}^2 - 2yM_{z_i} + y^2\gamma\right]\right], \\
&= \frac{-1}{2}\left[\frac{y^2}{\sigma_n^2} + \frac{M_{z_i}^2 - 2yM_{z_i}}{\sigma_n^2}\right], \\
&= \frac{-1}{2\sigma_n^2}(y - M_{z_i})^2.
\end{aligned} \tag{4.13}$$

Hence, introducing (4.13) in (4.12), we have,

$$p(y|H_i, z_i) \approx \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{(y-M_{z_i})^2}{2\sigma_n^2}} \tag{4.14}$$

for $i = 0, 1$.

Thus, conditioned on H_i and z_i, $y \sim \mathcal{N}(M_{z_i}, \sigma_n^2)$ for $i = 0, 1$. Hence, the probability of error conditioned on H_0 becomes,

$$\begin{aligned} P_e(H_0) &= \int_{\mathcal{Y}_1} p(y|H_0, z_0)dy, \\ &= Pr\left[y > \frac{M_{z_0} + M_{z_1}}{2} \middle| H_0\right], \\ &= Pr\left[y > \frac{M_{z_0} + M_{z_1}}{2} \middle| y \sim \mathcal{N}(M_{z_0}, \sigma_n^2)\right], \\ &= \mathcal{Q}\left(\frac{\frac{M_{z_0}+M_{z_1}}{2} - M_{z_0}}{\sigma_n}\right), \\ &= \mathcal{Q}\left(\frac{M_{z_1} - M_{z_0}}{2\sigma_n}\right). \end{aligned} \tag{4.15}$$

Similarly, the probability of error conditioned on H_1 becomes,

$$\begin{aligned} P_e(H_1) &= \int_{\mathcal{Y}_0} p(y|H_1, z_1)dy, \\ &= Pr\left[y < \frac{M_{z_0} + M_{z_1}}{2} \middle| y \sim \mathcal{N}(M_{z_1}, \sigma_n^2)\right], \\ &= \mathcal{Q}\left(\frac{M_{z_1} - M_{z_0}}{2\sigma_n}\right)... \end{aligned} \tag{4.16}$$

Introducing (4.15) and (4.16) in $P_{e_{z_0,z_1}}$ which is defined in corollary 3.3.1, the proof follows. □

5. OPTIMIZATION

Until this section we derived the probability of error given any two quantizer bins (i.e., our cost function, P_{e_T}, is the averaged out version of this probability of error given any two quantizer bins, $P_{e_{z_0,z_1}}$), the decision rule and their approximations. In this section we will apply numeric optimization techniques to our cost function under different scenarios, which are investigated in detail in the following subsections.

5.0.1. Numerically calculated optimal quantizer

In this part, we will present the optimal quantizer that minimizes the probability of detection error (P_{e_T}) with the simulation results for N-level quantizer, where $N = \{3, 5, 6\}$. Note that the decision rule and the cost function, that are used during the simulations, are the original expressions (i.e., without approximation) given in Theorem 3.2.1 and (3.9), respectively.

Since we used a brute force search algorithm, the complexity of the algorithm did not let us to give the optimal quantizer constellation for higher levels of quantization (for instance, for $N = 6$ case it took more than 3 weeks for the simulations to be finished for a step size of 0.01).

N=3 Case: The optimal quantizer constellation for $N = 3$, $\sigma_x^2 = 1$, $\sigma_n^2 = 0.1$ is given in Table 5.1 and illustrated in Figure 5.1.

Remark 5.0.1 *Note that, since our target random variable X has a Gaussian distribution with mean* 0 *and variance* 1*, for all practical purposes we assumed* -5 *as* $-\infty$ *and* 5 *as* ∞ *during the simulations.*

Table 5.1. Optimal Quantizer Bin Constellation and Corresponding P_{e_T} for $N = 3$

Bin 1	Bin 2	Bin 3	P_e
(-5,-0.45)	(-0.45,0.45)	(0.45,5)	0.2285

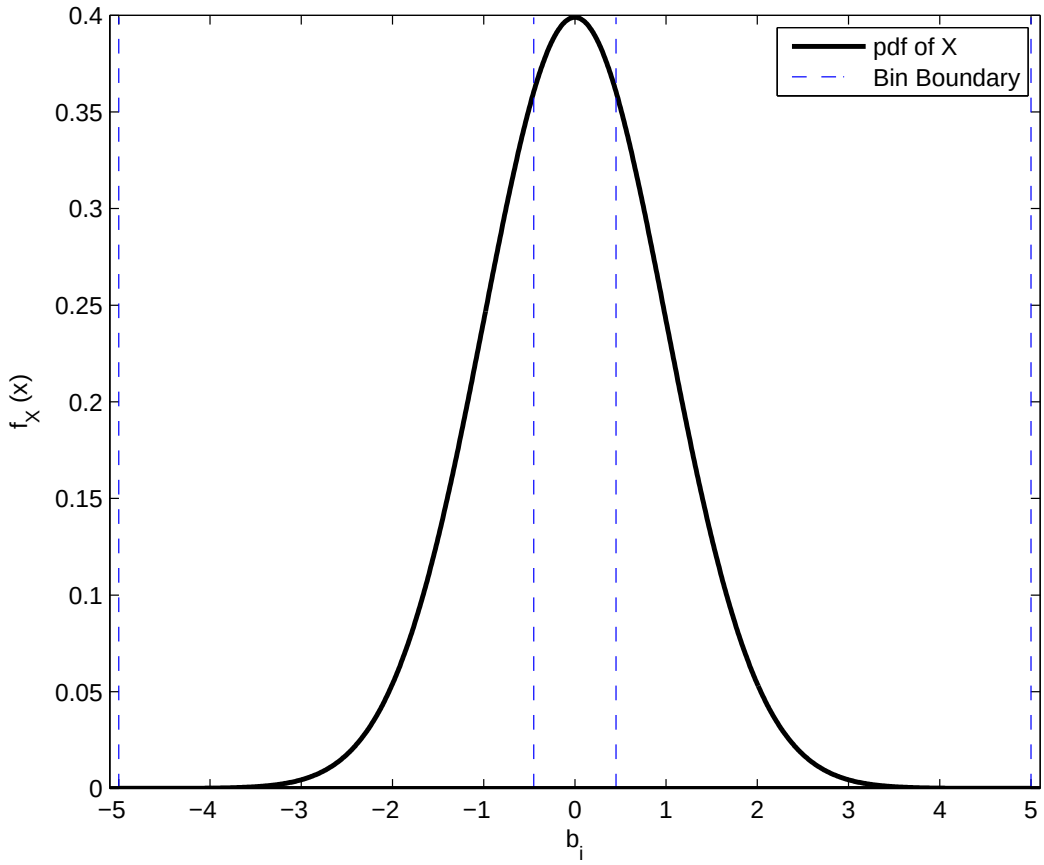

Figure 5.1. Optimal quantizer bin constellation for N=3

N=4 Case: The optimal quantizer constellation for $N = 4$, $\sigma_x^2 = 1$, $\sigma_n^2 = 0.1$ is given in Table 5.2 and illustrated in Figure 5.2.

Table 5.2. Optimal Quantizer Bin Constellation and Corresponding P_{e_T} for $N = 4$

Bin 1	Bin 2	Bin 3	Bin 4	P_e
(-5,-0.72)	(-0.72,0)	(0,0.72)	(0.72,5)	0.1951

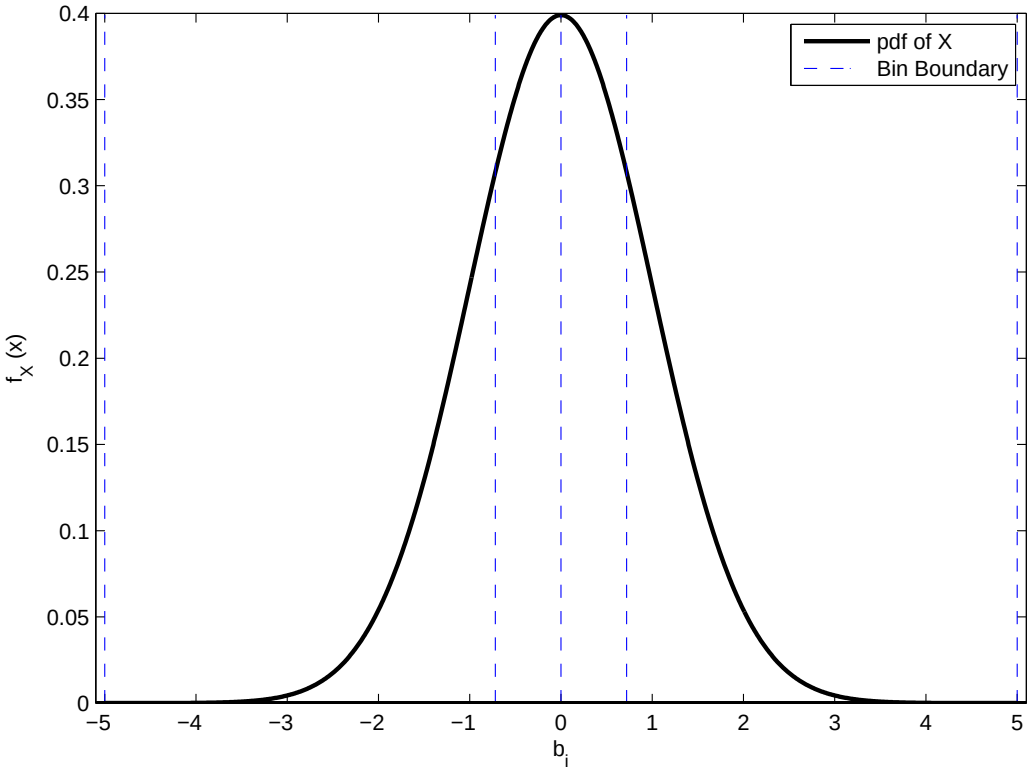

Figure 5.2. Optimal quantizer bin constellation for N=4

N=5 Case: The optimal quantizer constellation for $N = 5$, $\sigma_x^2 = 1$, $\sigma_n^2 = 0.1$ is given in Table 5.3 and illustrated in Figure 5.3.

Table 5.3. Optimal Quantizer Bin Constellation and Corresponding P_{e_T} for $N = 5$

Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	P_e
(-5,-0.92)	(-0.92,-0.28)	(-0.28,0.28)	(0.28,0.92)	(0.92,5)	0.1766

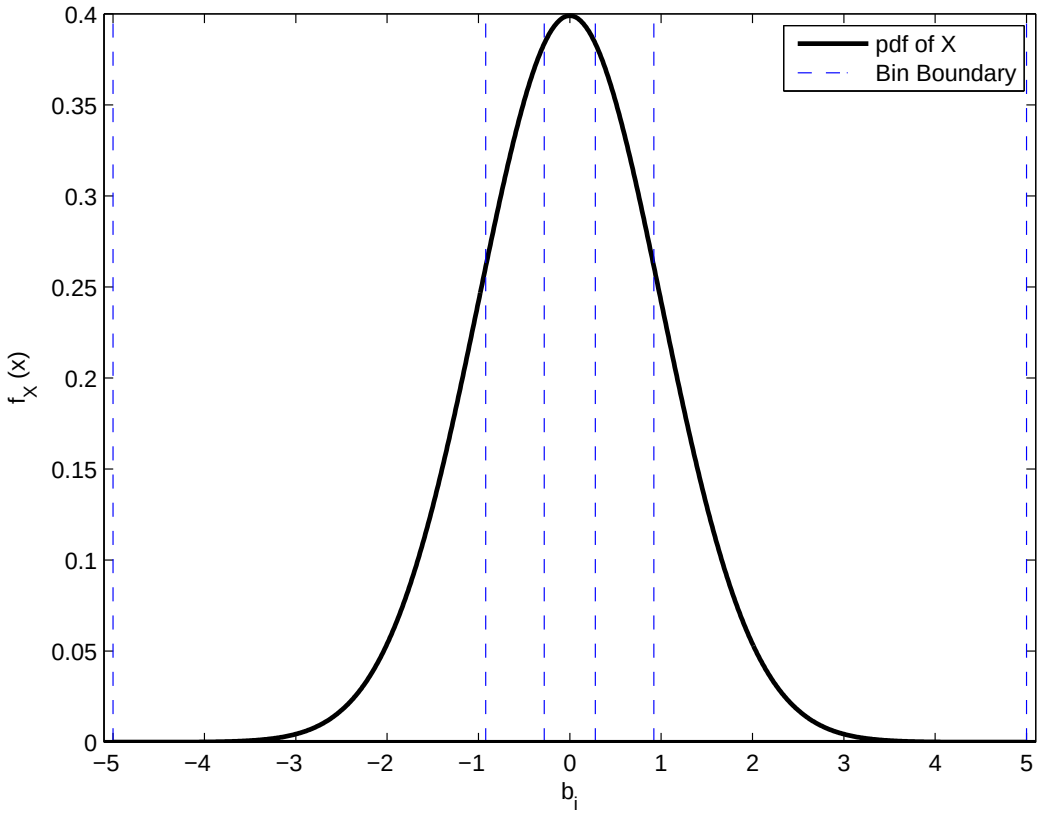

Figure 5.3. Optimal quantizer bin constellation for N=5

N=6 Case: The optimal quantizer constellation for $N = 6$, $\sigma_x^2 = 1$, $\sigma_n^2 = 0.1$ is given in Table 5.4 and illustrated in Figure 5.4.

Table 5.4. Optimal Quantizer Bin Constellation and Corresponding P_{e_T} for $N = 6$

Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	P_e
(-5,-1.08)	(-1.08,-0.49)	(-0.49,0)	(0,0.49)	(0.49,1.08)	(1.08,5)	0.1654

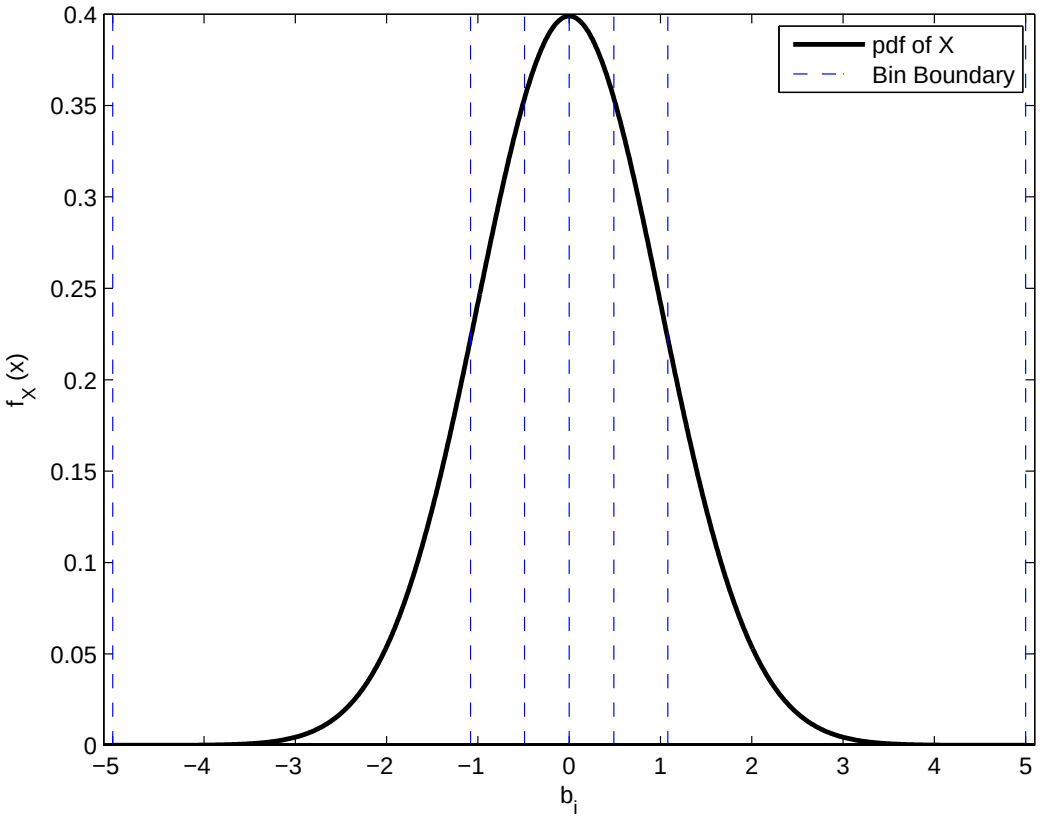

Figure 5.4. Optimal quantizer bin constellation for N=6

5.0.2. Numerically calculated suboptimal quantizer

We will use zeroth order Taylor approximation applied cost function and decision rule in our calculations to derive the suboptimal quantizer.

$$\underset{\substack{\{M_{z_i}\}_{i=1}^{N-2} \\ \forall i, M_{z_i+1} > M_{z_i}}}{\arg\min} \{P_{e_T} = \sum_{\forall z_0, z_1} \alpha_{z_0} \alpha_{z_1} P_{e_{z_0,z_1}}\},$$

where $\alpha_{z_i} \approx \frac{(b_{z_i+1}-b_{z_i})}{\sqrt{2\pi}\sigma_x} e^{-\frac{M_{z_i}^2}{2\sigma_x^2}}$ and $P_{e_{z_0,z_1}} = \mathcal{Q}\left(\frac{M_{z_1}-M_{z_0}}{2\sigma_n}\right)$, $M_{z_i} = \frac{b_{z_i}+b_{z_i+1}}{2}$, for $i = 0, 1$. Here the constraint is $\forall k, M_{k+1} > M_k$ for $k = 1, 2, \ldots, N-2$, and the initial condition is b_1.

5.0.3. Performance of the Quantizer

In this part, the performance (in probability of error sense) of numerically calculated quantizer given in Section 5.0.1, is compared to well-known Uniform and Lloyd-Max quantizers for detection with side information problem.

Note that, for all the quantizers that are compared, the probability of error is calculated according to optimal decision rule. The probability of detection errors of optimal , approximate (i.e., suboptimal), Lloyd-Max and Uniform quantizers (P_{e_T}) versus quantization level N are illustrated in the following Figure 5.5.

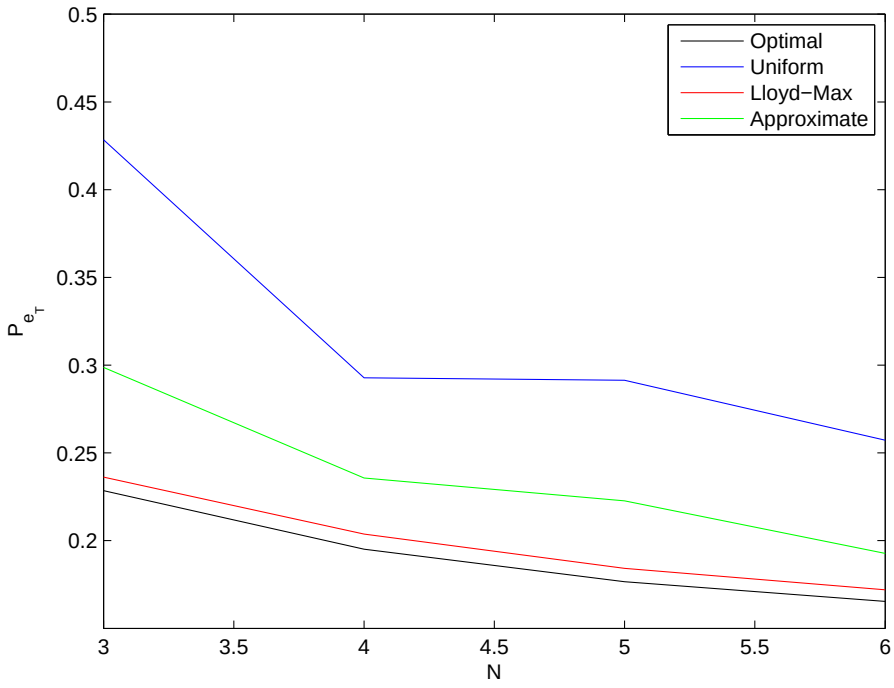

Figure 5.5. Probability of error behavior of Lloyd-Max, Uniform, Optimal and Approximate (Suboptimal) quantizers for the proposed problem.

Corollary 5.0.1 *As it can be seen from Figure 5.5, the approximate and the uniform quantizer is not successful in the probability of error sense. The reason of the bad performance of the approximate one is obviously the loose approximation we used. However, as we mentioned in Chapter 4, the approximation should be valid for fine quantization (i.e., when the quantization level N is big), so we can state that the performance will be much more better if we assume N is in the order of* 1000 *or more.*

Since, the Lloyd-Max quantization is much more closer to the optimal one in the probability of error sense, we will focus on these two, and compare their bin constellations. The constellation comparisons are illustrated in Figure 5.6, Figure 5.7, Figure 5.8, Figure 5.9 for $N = 3, 4, 5, 6$ respectively.

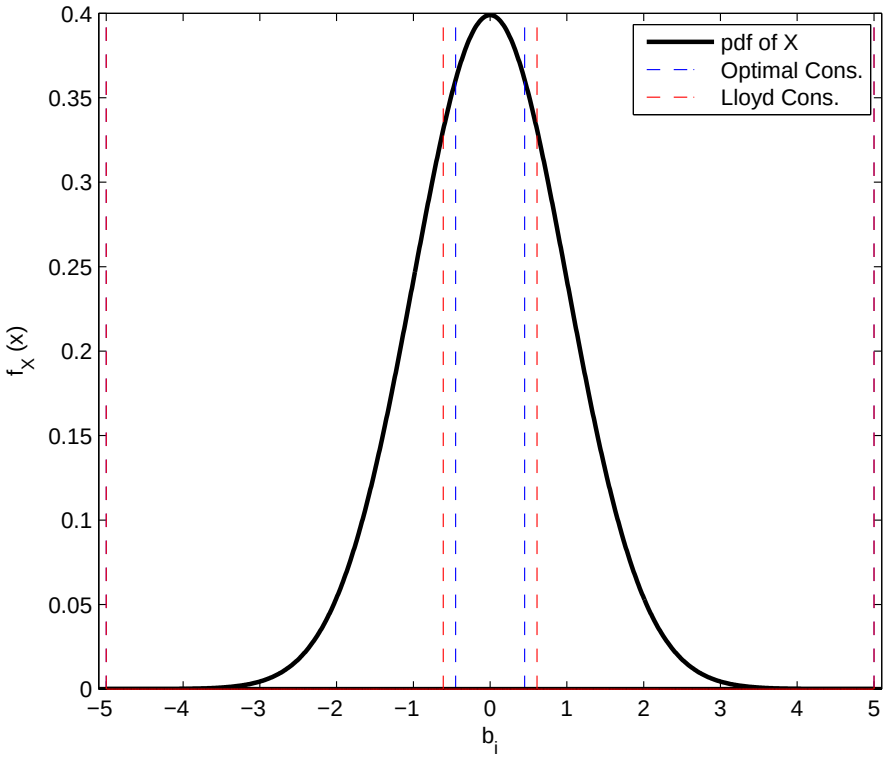

Figure 5.6. Bin constellation comparison of Lloyd-Max and Optimal Quantizer for $N = 3$.

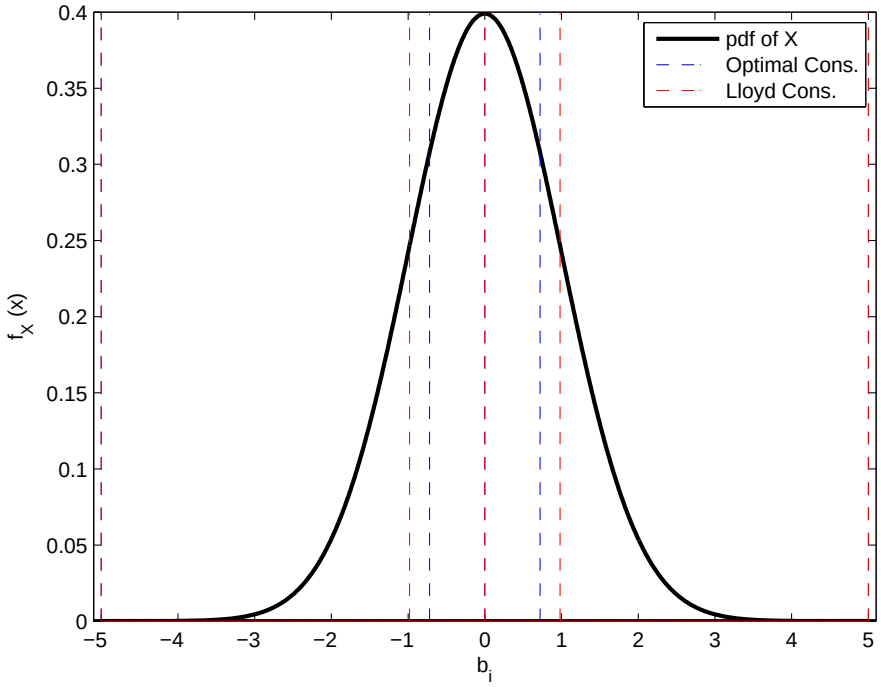

Figure 5.7. Bin constellation comparison of Lloyd-Max and Optimal Quantizer for $N = 4$.

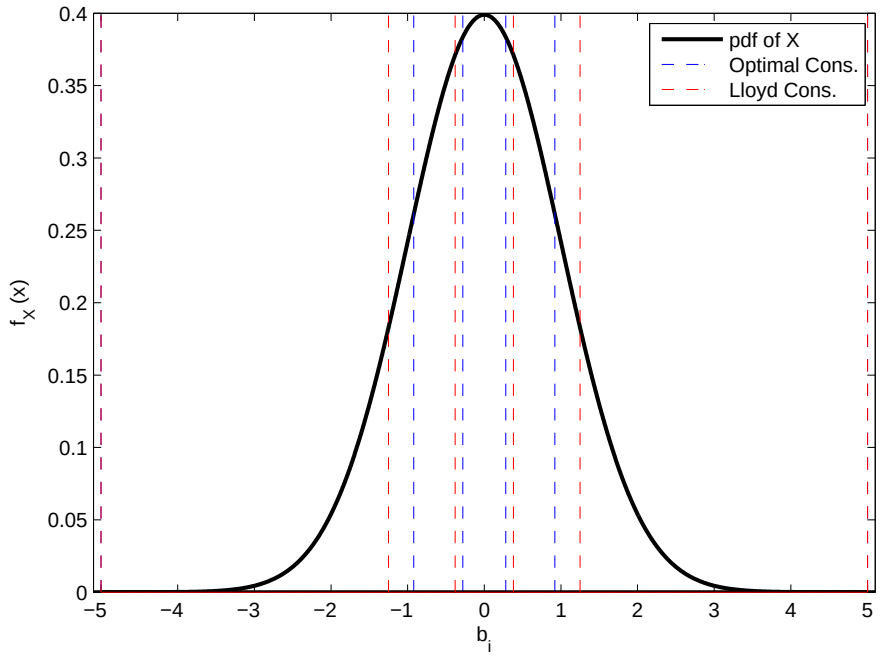

Figure 5.8. Bin constellation comparison of Lloyd-Max and Optimal Quantizer for $N = 5$.

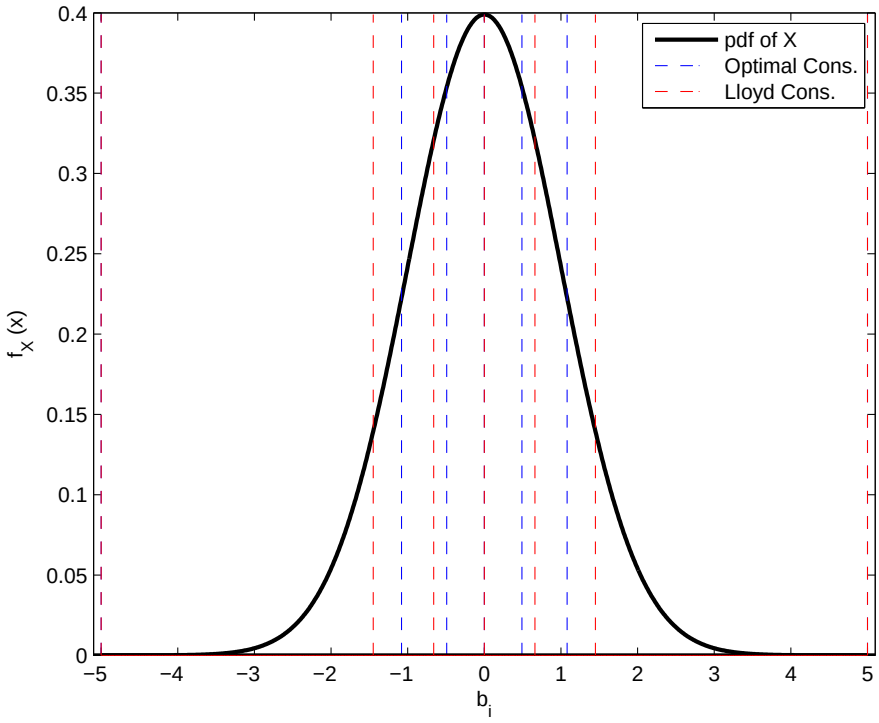

Figure 5.9. Bin constellation comparison of Lloyd-Max and Optimal Quantizer for $N = 6$.

6. CONCLUSIONS

In this thesis, we present a variant of detection with side information problem. The receiver (detector) receives a noise added (attacked) version of the target signal and tries to make a decision in a binary hypothesis schema. In our model, only the channel statistics and the hash values of the transmitted signals are known to the detector (i.e., unlike the classical communication setup, the message signals are not known by the detector). Here, we analyzed the case where hash values of the signals are obtained via scalar quantization. The channel is modeled as additive Gaussian. The transmitted signals are assumed to be realizations of a Gaussian random variable.

As a detection theoretic approach to robust signal hashing (where the hash values are reproduction values of the quantizer), we analyzed the binary version of the problem. We derived the optimal decision rule (LRT) in probability of error sense and corresponding probability of error. Since the derived decision rule and the corresponding probability of error expressions consist $\mathcal{Q}(.)$ differences, which can not be easily handled or bounded, we applied approximation to both decision threshold and probability of error expressions. Hence we obtain a nearest neighbor like decision threshold expression.

Furthermore, we presented and illustrated the optimal quantizers for several quantization levels. The *optimal quantizer*, *Lloyd-Max*, *Uniform* and *a suboptimal quantizer obtained by approximation* are compared w.r.t. bin constellations and their corresponding probability of detection errors. Finally, from the comparison it has been shown empirically that Lloyd-Max quantizers are suboptimal yet a good quantizer choice for detection with side information problem under proposed setup.

As a future work, bounds may be applied to the probability of error expression under approximated decision threshold schema, in order to obtain a easy to handle cost function for the optimization. However, this approach will only lead to a set of nonlinear equations where the solution will not be valid for coarse quantization.

REFERENCES

1. I. J. Cox, M. L. Miller and A. L. McKellips, "Watermarking as communications with side information," *Proc. IEEE,* vol. 87, No. 7, pp. 1127–1141, July 1999.

2. F. A. P. Petitcolas, R. J. Anderson and M. G. Kuhn, "Information hiding - a survey," *Proc. IEEE,* vol. 87, No. 7, pp. 1062–1078, July 1999.

3. P. Moulin and R. Koetter, "Data-Hiding Codes," (tutorial paper), *Proc. IEEE,* vol. 93, No. 12, pp. 2083–2127, Dec. 2005.

4. R. Venkatesan, S. M. Koon, M. H. Jakubowski and P. Moulin, "Robust image hashing," in *Proc. IEEE Int. Conf. Image Processing,* vol. 3, pp. 664–666, 2005.

5. M. K. Mihcak and R. Venkatesan, "A Perceptual Audio Hashing Algorithm: A Tool For Robust Audio Identification and Information Hiding," in *Proceedings of* 4^{th} *International Information Hiding Workshop,* 2001.

6. S. S. Kozat, R. Venkatesan and M. K. Mihcak, "Robust Hashing via Matrix Invariances," in *Procedings of IEEE International Conference on Image Processing (ICIP),* 2004.

7. H. Ozer, B. Sankur, N. Memon and E. Anarim, "Robust Hashing via Matrix Invariances,"Perceptual audio hashing fuctions," *EURASIP J. Appl. Signal Process.,* pp. 1780–1793, 2005.

8. R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors. Part I: Fundementals," *Proc. IEEE,* vol. 85, pp. 54–63, January 1997.

9. Wyner, A. D. and J. Ziv, "The Rate Distortion Function for Source Coding with Side Information at the Receiver," *IEEE Trans. Inf. Theory,* vol. IT–22, no. 1, pp. 1–11, 1976.

10. O. Koval, S. Voloshynovskiy, F. Beekhof and T. Pun, "DecisionTheoretic Consideration of Robust Perceptual Hashing: Link to Practical Algorithms," in *Proceedings of Third WAVILA Challenge,* 2007.

11. S. Voloshynovskiy, O. Koval, F. Beekhof and T. Pun, "Conception and limits of robust perceptual hashing: toward side information assisted hash functions," in *Proceedings of SPIE Photonics West, Electronic Imaging / Media Forensics and Security XI*, San Jose, USA, 2009.

12. M. K. Mihcak, Y. Altug and N. P. Ayerden, "On Minimax Optimal Linear Transforms for Detection with Side Information in Gaussian Setup," *IEEE Communications Letters,* vol. 12, no. 3, Mar. 2008.

13. H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed., Springer, 1994.

14. A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, 1st ed., Kluwer Academic Publishers, 1991.