Noise Enhanced Parameter Estimation Using Quantized Observations

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

AND THE INSTITUTE OF ENGINEERING AND SCIENCES OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

By

Gökce Osman Balkan

July 2010

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.

Assist. 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.

Assist. Prof. Dr. Selim Aksoy

Approved for the Institute of Engineering and Sciences:

Prof. Dr. Levent Onural
Director of Institute of Engineering and Sciences

ABSTRACT

Noise Enhanced Parameter Estimation Using Quantized Observations

Gökce Osman Balkan
M.S. in Electrical and Electronics Engineering
Supervisor: Assist. Prof. Dr. Sinan Gezici

July 2010

In this thesis, optimal additive noise is characterized for both single and multiple parameter estimation based on quantized observations. In both cases, first, optimal probability distribution of noise that should be added to observations is formulated in terms of a Cramer-Rao lower bound (CRLB) minimization problem. In the single parameter case, it is proven that optimal additive "noise" can be represented by a constant signal level, which means that randomization of additive signal levels (equivalently, quantization levels) are not needed for CRLB minimization. In addition, the results are extended to the cases in which there exists prior information about the unknown parameter and the aim is to minimize the Bayesian CRLB (BCRLB). Then, numerical examples are presented to explain the theoretical results. Moreover, performance obtained via optimal additive noise is compared to performance of the commonly used dither signals. Furthermore, mean-squared error (MSE) performances of maximum likelihood (ML) and maximum a-posteriori probability (MAP) estimates are investigated in the presence and absence of additive noise. In the multiple parameter case, the form of the optimal random additive noise is derived for CRLB minimization. Next, the theoretical result is supported with a numerical example, where

the optimum noise is calculated by using the particle swarm optimization (PSO) algorithm. Finally, the optimal constant noise in the multiple parameter estimation problem in the presence of prior information is discussed.

Keywords: Estimation, quantization, Cramer-Rao lower bound, noise enhanced estimation, mean-squared error, maximum likelihood, maximum a-posteriori probability, particle swarm optimization

ÖZET

NİCEMLENMİŞ GÖZLEMLER KULLANARAK GÜRÜLTÜ İLE GELİŞTİRİLMİŞ PARAMETRE KESTİRİMİ

Gökce Osman Balkan
Elektrik ve Elektronik Mühendisliği Bölümü Yüksek Lisans
Tez Yöneticisi: Yrd. Doç. Dr. Sinan Gezici
Temmuz 2010

Bu tezde nicemlenmiş gözleme dayalı tekli ve çoklu parametre kestiriminde eniyi ek gürültü tanımlanmıştır. Her iki durumda da ilk olarak gözleme eklenmesi gereken eniyi gürültünün olasılık dağılımı Cramer-Rao alt sınırı (CRLB) enküçültme problemi cinsinden formülleştirilmiştir. Tek parametreli durumda eniyi ek "gürültünün" sabit bir sinyal seviyesi ile gösterilebildiği kanıtlanmıştır. Bu da CRLB enküçültmesi için ek sinyal seviyelerinin rastgeleleştirilmesine gerek olmadığı anlamına gelmektedir. Ayrıca bu sonuçlar, bilinmeyen parametre hakkında ön bilginin mevcut olduğu ve Bayesian CRLB'nin (BCRLB) enküçültmesinin amaçlandığı durumlara genişletilmiştir. Sonrasında kuramsal sonuçları açıklamak için sayısal örnekler sunulmuştur. Bunun dışında, eniyi gürültü ile elde edilen performans gelişimi sıkça kullanılan kıpırtı (dither) sinyalleri ile karşılaştırılmıştır. Ayrıca enbüyük olabilirlikli ve enbüyük sonsal olasılık kestiricilerin ortalama hata kare performansları gürültü ile geliştirilmiş ve ek gürültüsüz durumlar için karşılaştırılmıştır. Çoklu parametre durumunda CRLB enküçültmesi için eniyi rastgele ek gürültünün şekli türetilmiştir. Ardından kuramsal sonuç, eniyi gürültünün parçacık sürü eniyileştirmesi ile bulunduğu sayısal

bir örnek ile desteklenmıştir. Son olarak, ön bilginin varsayıldığı çoklu parametre kestirim probleminde eniyi sabit gürültü incelenmiştir.

Anahtar Kelimeler: Kestirim, nicemleme, Cramer-Rao alt sınırı, gürültü ile geliştirilmiş kestirim, ortalama hata kare, enbüyük olabilirlik, enbüyük sonsal olasılık, parçacık sürü eniyileştirmesi

ACKNOWLEDGMENTS

My gratitude to my supervisor Assist. Prof. Sinan Gezici, who was the most inspiring and influential person during my graduate studies, is inexpressible. Providing me a great research environment with his positiveness, friendly approach and endless support, he kept my motivation at highest level all the time. It was a great honor and privilege for me to work with him.

Additionally, I would like to express my gratefulness to Prof. Orhan Arıkan, who was a great motivating advisor in my undergraduate years, for his invaluable support throughout my undergraduate and graduate studies and serving in my thesis committee. Also, I would like to thank Assist. Prof. Selim Aksoy for his service in my thesis committee.

The idea of the problems examined in this thesis is inspired from the contributions of Suat Bayram in the area of noise enhanced detection. Special thanks must go to him.

Finally, I especially would like to thank my family for their unconditional love and helping me to realize my dreams.

Contents

List of Figures

List of Tables

To my mother Yaşar Afet Balkan...

Chapter 1

INTRODUCTION

Although noise commonly degrades the performance of a system, some nonlinear systems can benefit from addition of noise to their inputs or from increased noise levels [1]-[4]. In detection theory, such noise benefits are observed for certain suboptimal detectors, which achieve improved detection performance in the presence of additive noise [5], [6]. Recent studies quantify the noise benefits for suboptimal detectors in the Bayesian, minimax, and Neyman-Pearson frameworks [5]-[11].

Noise benefits are also observed in the form of dithering in quantization systems (cf. [12] and references therein). It is shown in [13] that noise benefits can be obtained in sigma-delta quantizer in terms of improved signal-to-noise ratio (SNR). In addition, [14] reveals that the information transmitted in an array of comparators is maximized at a certain ratio between the standard deviation of the random input signal and that of the noise, where the cases of various probability distributions of the signal and the noise are considered. Furthermore, parameter estimation based on 1-bit dithered quantization is studied in [12], and an estimator that does not require any information about the dither signal and the noise distribution is proposed.

Additive noise benefits in parameter estimation problems are investigated also in [15]-[17]. The frequency estimation problem in [15] reveals that, the mean-squared error (MSE) of the optimal Bayesian estimator can decrease under certain conditions, when the noise level is increased. Likewise, [16] considers Bayesian estimation and provides examples of when raised noise levels result in improved MSE performance. In [15] and [16], 1-bit quantizers are employed and noise benefits are observed due to the nonlinear structure of the quantizers. In another noise enhanced estimation study [17], the first and the second moments of an estimator and a Bayesian cost function are used as performance criteria and the general form of the optimal noise probability density function (p.d.f.) is derived.

For some noise enhanced parameter estimation problems, asymptotical behaviors of the estimators make the Cramer-Rao lower bound (CRLB), equivalently the Fisher information, an appealing metric for the quantification of performance improvements via additive noise [4]. For example, maximization of the Fisher information for parameter estimation based on quantized observations is studied in [18] by optimizing quantization intervals. In addition, the dependence of the MSE of a mean estimator on the probability distribution of observation noise is investigated in [19] and theoretical lower bounds are provided. In [20], parameter estimation based on observations from a multi-bit quantizer is considered and additive controlled perturbation of the quantizer thresholds is investigated. In particular, [20] shows that random dithering can significantly reduce the CRLB for the mean estimation problem with 1-bit precision sampling. Also, it is shown in [21] that the variance of an estimator that uses 1-bit quantizer outputs can be made quite close to the variance of a clairvoyant estimator that uses unquantized observations by an appropriate choice of the quantizer threshold. Moreover, addition of noise to quantized measurements can provide enhancement of the Fisher information for the estimation of the suprathreshold input signals [22]. Furthermore, maximization of the Fisher information by both an appropriate choice of the quantizer threshold and additive noise is studied in [23]. Finally, another related problem is the optimal quantization of random variables according to the minimum MSE criterion, which differs from the studies on noise enhanced parameter estimation that consider the CRLB as the optimization metric [22].

Although the effects of additive noise on CRLBs have been investigated in [20], [22] and [23], the optimal p.d.f. of additive noise that minimizes the CRLB for parameter estimation based on quantized observations has not been obtained before. In this thesis, a parameter estimation problem based on quantized observations is studied, where the aim to find the optimal p.d.f. of noise that should be added to the observations before the quantizer in order to minimize the CRLB for estimating the unknown parameter (see Figure 2.1). Unlike the previous studies, an explicit CRLB minimization problem is formulated in terms of the additive noise p.d.f., the quantization function, and the p.d.f. of the original observation. In addition, the quantizer is modeled by a generic multi-bit quantizer with arbitrary quantization levels.

In Chapter 2, the single parameter case of the noise enhanced estimation problem is studied [4], [24]. First, the problem is formulated as a Fisher information maximization problem, where the aim is to find the probability distribution of the optimal additive noise. In the next step, the derivation of the theoretical solution to the problem employing the convexity of the Fisher information of the estimate is given. It is shown that the optimal additive noise can be represented as a deterministic constant signal. Additionally, using similar derivations, it is also shown that this result is also valid for the random parameter case, where Bayesian CRLB (BCRLB) replaces CRLB. Then, three numerical examples are presented in order to support the theoretical results for both fixed and random parameter cases. For each example, the outcomes of theoretical results are compared with the effects of the common dithering signals. Finally, MSE performance of asymptotically efficient maximum likelihood (ML) and maximum a-posteriori probability (MAP) estimators are compared.

In Chapter 3, the multiple parameter version of the problem in Chapter 2 is investigated. The problem is formulated as an optimization problem, in which the parameters are deterministic and the p.d.f. of the additive noise maximizing the trace of the inverse Fisher information matrix is sought. By employing Carathéodory's theorem, the form of the p.d.f. of the optimal additive noise is found. As the next step, a numerical example using the theoretical results is studied. In the numerical example, the particle swarm optimization (PSO) technique is employed in order to find the characteristics of the optimal additive noise. Next, the performance improvements in terms of MSE are investigated, where the root-mean-squared errors (RMSEs) of the ML estimates for the cases with optimal random noise enhanced, optimal constant noise enhanced and noiseless observations are compared with their CRLBs. It is shown that a random additive noise can result in better estimation performance than constant additive noise, if more than one parameter in the observations is to be estimated. Finally, the optimal constant additive noise is investigated for the random parameter case of the problem.

In Chapter 4, the conclusions inferred from this noise enhanced parameter estimation study are summarized and future works are discussed.

Chapter 2

OPTIMAL ADDITIVE NOISE IN SINGLE PARAMETER ESTIMATION PROBLEMS

2.1 Problem Formulation

Consider a system in which a quantized version of observation $\mathbf{x}$ is used to estimate an underlying parameter θ [4]. Let $p_{\mathbf{X}}(\mathbf{x};\theta)$ represent the p.d.f. of the observation, and $\varphi(\cdot)$ denote the quantizer. Instead of using observation $\mathbf{x}$, a noise modified version of the observation, $\mathbf{x}+\mathbf{n}$, can be used as in Figure 2.1 in order to improve the estimation accuracy of the system, where the additive noise $\mathbf{n}$ is independent of the observation $\mathbf{x}$ [5], [6]. The aim is to obtain the p.d.f. of $\mathbf{n}$, denoted by $p_{\mathbf{N}}(\cdot)$, that maximizes the estimation accuracy of the system in Figure 2.1. It is noted that this noise enhanced parameter estimation problem can also be regarded as a dynamic bias control problem as in [20], when $\mathbf{n}$ represents the control input for the quantizer bias.

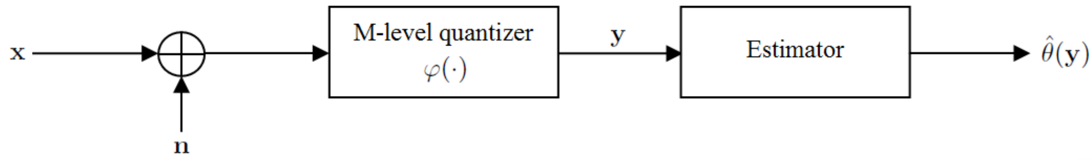

Figure 2.1: Block diagram of the system, where $\mathbf{n}$ denotes the additive noise that is independent of the original observation $\mathbf{x}$.

Suppose that quantizer $\varphi(\cdot)$ is an M-level quantizer that generates the quantized observation vector $\mathbf{y}$ based on the noise modified input observation as follows:

$$\mathbf{y} = \varphi(\mathbf{x} + \mathbf{n}) \ , \tag{2.1}$$

where $\mathbf{y} = [y_1 \ y_2 \cdots y_L]$, $\mathbf{x} = [x_1 \ x_2 \cdots x_L]$, $\mathbf{n} = [n_1 \ n_2 \cdots n_L]$, and the quantizer levels are determined by thresholds $\tau_1, \ldots, \tau_{M-1}$. Specifically, the relation between the input and the output of the quantizer is described by

$$y_j = \begin{cases} 0 \ , & \text{if } x_j + n_j \leq \tau_1 \\ 1 \ , & \text{if } \tau_1 < x_j + n_j \leq \tau_2 \\ \vdots & \vdots \\ M-1 \ , & \text{if } \tau_{M-1} < x_j + n_j \end{cases} . \tag{2.2}$$

Let $p_{\mathbf{Y}}(\cdot\,;\theta)$ represent the probability mass function (p.m.f.) of the quantizer output for a given value of θ. From (2.2), it can be obtained as

$$p_{\mathbf{Y}}(\boldsymbol{i}\,;\theta) = \tag{2.3}$$
$$\int_{\mathbb{R}^L} \mathrm{P}(\tau_{i_1} - n_1 < X_1 \leq \tau_{i_1+1} - n_1, \ldots, \tau_{i_L} - n_L < X_L \leq \tau_{i_L+1} - n_L)\, p_{\mathbf{N}}(\mathbf{n})\, d\mathbf{n}$$

for $\boldsymbol{i} \in \mathcal{I} \triangleq \{0, 1, \ldots, M-1\}^L$, where i_l represents the lth component of $\boldsymbol{i}$.

The additive noise component $\mathbf{n}$ in Figure 2.1 is optimized according to the CRLB in this study [4]. In other words, the optimal noise p.d.f. that minimizes the CRLB is sought for. The CRLB on the MSE of unbiased estimators $\hat{\theta}$ of θ

is stated as

$$\mathrm{MSE}_\theta\{\hat{\theta}\} \geq \mathrm{J}_\theta^{-1} = \left(\mathrm{E}\left\{\left(\frac{\partial \log p_{\mathbf{Y}}(\mathbf{y};\theta)}{\partial\theta}\right)^2\right\}\right)^{-1}, \tag{2.4}$$

where $\mathrm{MSE}_\theta\{\hat{\theta}\} = \mathrm{E}\{(\hat{\theta}(\mathbf{y}) - \theta)^2\}$, J_θ is defined as the Fisher information [25], and $p_{\mathbf{Y}}(\cdot\,;\theta)$ is as in (2.3). Since the CRLB imposes a lower limit on the MSE of an unbiased estimator and since certain estimators, such as the maximum likelihood estimator, can (asymptotically) achieve the CRLB under certain conditions [25], the aim in this study is to obtain the optimal p.d.f. of the additive noise that minimizes the CRLB specified by (2.4). It should be noted that this approach does not require any information about the estimator that is used after the quantizer. If the aim is to minimize the MSE of a *given* suboptimal estimator, then the approach in [17] can be employed.

As the CRLB is the inverse of the Fisher information, the optimal additive noise p.d.f. can be formulated, from (2.4), as the solution of the following optimization problem:

$$p_{\mathbf{N}}^{\mathrm{opt}}(\mathbf{n}) = \arg\max_{p_{\mathbf{N}}(\cdot)} \mathrm{E}\left\{\left(\frac{\partial \log p_{\mathbf{Y}}(\mathbf{y};\theta)}{\partial\theta}\right)^2\right\}. \tag{2.5}$$

Since $\mathbf{Y}$ is equal to $\boldsymbol{i}$ with probability $p_{\mathbf{Y}}(\boldsymbol{i}\,;\theta)$ as defined in (2.3), the problem in (2.5) can be expressed as

$$p_{\mathbf{N}}^{\mathrm{opt}}(\mathbf{n}) = \arg\max_{p_{\mathbf{N}}(\cdot)} \sum_{\boldsymbol{i}\in\mathcal{I}} \frac{1}{p_{\mathbf{Y}}(\boldsymbol{i}\,;\theta)} \left(\frac{\partial p_{\mathbf{Y}}(\boldsymbol{i}\,;\theta)}{\partial\theta}\right)^2. \tag{2.6}$$

As a special case of the generic problem formulation in (2.6), when both $\mathbf{X}$ and $\mathbf{N}$ consist of independent components, it can be shown that the components of the optimal additive noise can be calculated separately; i.e.,

$$p_{N_l}^{\mathrm{opt}}(n) = \arg\max_{p_{N_l}(\cdot)} \mathrm{E}\left\{\left(\frac{\partial \log p_{Y_l}(y_l;\theta)}{\partial\theta}\right)^2\right\}, \tag{2.7}$$

for $l = 1, \ldots, L$, where $p_{N_l}(\cdot)$ represents the marginal p.d.f. of the lth component of the additive noise. If $p_{Y_l}(i\,;\theta)$ denotes the probability that Y_l is equal to i for

$i = 0, 1, \ldots, M-1$, then (2.7) can be expressed as

$$p_{N_l}^{\text{opt}}(n) = \arg\max_{p_{N_l}(\cdot)} \sum_{i=0}^{M-1} \frac{1}{p_{Y_l}(i\,;\theta)} \left(\frac{\partial p_{Y_l}(i\,;\theta)}{\partial \theta} \right)^2 , \tag{2.8}$$

for $l = 1, \ldots, L$. In addition, if $Y_1, \ldots, Y_L$ are independent and identically distributed (i.i.d.); that is, if $p_{Y_l}(i\,;\theta) = p_Y(i\,;\theta)$ for $l = 1, \ldots, L$, the optimization problems in (2.8) become identical. In other words, in the i.i.d. case, the same optimal noise value is added to each component of the original observation $\mathbf{x}$.

2.2 Statistical Characterization of Optimal Additive Noise

In order to investigate the statistical properties of the optimal additive noise in (2.6), we first introduce the following functions:

$$H_{\boldsymbol{i}}^{\theta}(\mathbf{n}) \triangleq \mathrm{P}(\tau_{i_1} - n_1 < X_1 \leq \tau_{i_1+1} - n_1, \ldots, \tau_{i_L} - n_L < X_L \leq \tau_{i_L+1} - n_L) , \tag{2.9}$$

$$G_{\boldsymbol{i}}^{\theta}(\mathbf{n}) \triangleq \frac{\partial H_{\boldsymbol{i}}^{\theta}(\mathbf{n})}{\partial \theta} . \tag{2.10}$$

It is noted from (2.3) that $0 \leq H_{\boldsymbol{i}}^{\theta}(\mathbf{n}) \leq 1$, $\forall \mathbf{n}$, and that $\sum_{\boldsymbol{i} \in \mathcal{I}} H_{\boldsymbol{i}}^{\theta}(\mathbf{n}) = 1$. Based on the definitions in (2.9) and (2.10), the p.m.f. in (2.3) and its derivative with respect to θ can be expressed as

$$p_{\mathbf{Y}}(\boldsymbol{i}\,;\theta) = \mathrm{E}\{H_{\boldsymbol{i}}^{\theta}(\mathbf{N})\}, \qquad \frac{\partial p_{\mathbf{Y}}(\boldsymbol{i}\,;\theta)}{\partial \theta} = \mathrm{E}\{G_{\boldsymbol{i}}^{\theta}(\mathbf{N})\} . \tag{2.11}$$

Then, the optimization problem in (2.6) becomes

$$p_{\mathbf{N}}^{\text{opt}}(\mathbf{n}) = \arg\max_{p_{\mathbf{N}}(\cdot)} \sum_{\boldsymbol{i} \in \mathcal{I}} \frac{\left(\mathrm{E}\left\{G_{\boldsymbol{i}}^{\theta}(\mathbf{N})\right\}\right)^2}{\mathrm{E}\left\{H_{\boldsymbol{i}}^{\theta}(\mathbf{N})\right\}} . \tag{2.12}$$

In order to obtain the solution of (2.12), the following lemma is presented first [4].

Lemma 1: *For the real-valued functions defined in* (2.9) *and* (2.10),

$$\sum_{\boldsymbol{i}\in\mathcal{I}} \frac{\left(\mathrm{E}\{G_{\boldsymbol{i}}^{\theta}(\mathbf{N})\}\right)^2}{\mathrm{E}\{H_{\boldsymbol{i}}^{\theta}(\mathbf{N})\}} \leq \max_{\mathbf{n}} \left\{ \sum_{\boldsymbol{i}\in\mathcal{I}} \frac{\left(G_{\boldsymbol{i}}^{\theta}(\mathbf{n})\right)^2}{H_{\boldsymbol{i}}^{\theta}(\mathbf{n})} \right\} \tag{2.13}$$

is satisfied for all θ and all possible p.d.f.s $p_{\mathbf{N}}(\cdot)$ of $\mathbf{N}$.

Proof: Consider a function of two variables defined as $f(\mathbf{Z}) = Z_1^2/Z_2$, where $\mathbf{Z} = [Z_1 \ Z_2]$. The Hessian of $f(\mathbf{Z})$ is calculated as

$$\mathbf{H}_f = \begin{bmatrix} 2/Z_2 & -2Z_1/Z_2^2 \\ -2Z_1/Z_2^2 & 2Z_1^2/Z_2^3 \end{bmatrix}, \tag{2.14}$$

which results in $\boldsymbol{\alpha}^T\mathbf{H}_f\boldsymbol{\alpha} = 2(\alpha_1 Z_2 - \alpha_2 Z_1)^2/Z_2^3 \geq 0$ for all $\boldsymbol{\alpha} = [\alpha_1 \ \ \alpha_2]^T$ and $Z_2 \geq 0$, implying that $\mathbf{H}_f$ is positive semidefinite; hence, $f(\mathbf{Z})$ is convex for $Z_2 \geq 0$. Therefore, Jensen's inequality implies that

$$\frac{(\mathrm{E}\{Z_1\})^2}{\mathrm{E}\{Z_2\}} \leq \mathrm{E}\left\{\frac{Z_1^2}{Z_2}\right\} \tag{2.15}$$

for $Z_2 \geq 0$. If we define $Z_1 \triangleq G_{\boldsymbol{i}}^{\theta}(\mathbf{N})$ and $Z_2 \triangleq H_{\boldsymbol{i}}^{\theta}(\mathbf{N})$, (2.15) becomes

$$\frac{\left(\mathrm{E}\{G_{\boldsymbol{i}}^{\theta}(\mathbf{N})\}\right)^2}{\mathrm{E}\{H_{\boldsymbol{i}}^{\theta}(\mathbf{N})\}} \leq \mathrm{E}\left\{ \frac{\left(G_{\boldsymbol{i}}^{\theta}(\mathbf{N})\right)^2}{H_{\boldsymbol{i}}^{\theta}(\mathbf{N})} \right\} \tag{2.16}$$

for all $p_{\mathbf{N}}(\cdot)$, θ and $\boldsymbol{i}$, since $H_{\boldsymbol{i}}^{\theta}(\mathbf{n}) \geq 0$, $\forall \mathbf{n}, \boldsymbol{i}, \theta$, by definition (cf. (2.9)). As the inequality in (2.16) is valid for all $\boldsymbol{i}$'s, we obtain

$$\sum_{\boldsymbol{i}\in\mathcal{I}} \frac{\left(\mathrm{E}\{G_{\boldsymbol{i}}^{\theta}(\mathbf{N})\}\right)^2}{\mathrm{E}\{H_{\boldsymbol{i}}^{\theta}(\mathbf{N})\}} \leq \mathrm{E}\left\{ \sum_{\boldsymbol{i}\in\mathcal{I}} \frac{\left(G_{\boldsymbol{i}}^{\theta}(\mathbf{N})\right)^2}{H_{\boldsymbol{i}}^{\theta}(\mathbf{N})} \right\}, \tag{2.17}$$

for all $p_{\mathbf{N}}(\cdot)$ and θ. Finally, as the expression on the right-hand-side of (2.17) is never larger than $\max_{\mathbf{n}} \left\{ \sum_{\boldsymbol{i}\in\mathcal{I}} \frac{(G_{\boldsymbol{i}}^{\theta}(\mathbf{n}))^2}{H_{\boldsymbol{i}}^{\theta}(\mathbf{n})} \right\}$, the result in the lemma is obtained. $\square$

Lemma 1 states that for each possible noise p.d.f. $p_{\mathbf{N}}(\mathbf{n})$, the Fisher information $\sum_{\boldsymbol{i}\in\mathcal{I}} \frac{\left(\mathrm{E}\{G_{\boldsymbol{i}}^{\theta}(\mathbf{N})\}\right)^2}{\mathrm{E}\{H_{\boldsymbol{i}}^{\theta}(\mathbf{N})\}}$ can never be larger than the maximum of $\sum_{\boldsymbol{i}\in\mathcal{I}} \frac{(G_{\boldsymbol{i}}^{\theta}(\mathbf{n}))^2}{H_{\boldsymbol{i}}^{\theta}(\mathbf{n})}$ over all possible noise values, $\mathbf{n}$. In other words, Lemma 1 states that randomization among different noise values cannot improve (increase) the objective function in (2.12). This result leads to the following proposition.

Proposition 1: *The optimal noise p.d.f. in* (2.12) *can be expressed as*

$$p_{\mathbf{N}}^{\text{opt}}(\mathbf{n}) = \delta(\mathbf{n} - \mathbf{n}_{\text{o}})\,, \tag{2.18}$$

where

$$\mathbf{n}_{\text{o}} = \arg\max_{\mathbf{n}} \sum_{\boldsymbol{i}\in\mathcal{I}} \frac{\left(G_{\boldsymbol{i}}^{\theta}(\mathbf{n})\right)^2}{H_{\boldsymbol{i}}^{\theta}(\mathbf{n})}\ . \tag{2.19}$$

Proof: Since the result in Lemma 1 holds for any $p_{\mathbf{N}}(\cdot)$, the following inequality can be obtained:

$$\max_{p_{\mathbf{N}}(\cdot)} \left\{ \sum_{\boldsymbol{i}\in\mathcal{I}} \frac{\left(\mathrm{E}\{G_{\boldsymbol{i}}^{\theta}(\mathbf{N})\}\right)^2}{\mathrm{E}\{H_{\boldsymbol{i}}^{\theta}(\mathbf{N})\}} \right\} \leq \max_{\mathbf{n}} \left\{ \sum_{\boldsymbol{i}\in\mathcal{I}} \frac{\left(G_{\boldsymbol{i}}^{\theta}(\mathbf{n})\right)^2}{H_{\boldsymbol{i}}^{\theta}(\mathbf{n})} \right\}\ . \tag{2.20}$$

Therefore, the maximum value of the objective function in (2.12) can never be larger than the expression on the right-hand-side of (2.20). However, this upper bound is achievable for $p_{\mathbf{N}}(\mathbf{n}) = \delta(\mathbf{n}-\mathbf{n}_{\text{o}})$, where $\mathbf{n}_{\text{o}}$ is defined as in (2.19). Hence, the optimal additive noise can be expressed as specified in the proposition. □

Proposition 1 states that for any additive noise that has a p.d.f. with multiple mass points, there always exists a corresponding constant "noise" level that provides an equal or smaller CRLB. In addition, it is noted from Lemma 1 and Proposition 1 that a constant additive "noise" component is optimal irrespective of the number of quantization levels (M) and the dimension of the observation vector (L). In addition, no assumption is imposed on the p.d.f. of the original observation, $\mathbf{x}$.

For the special case in which $\mathbf{X}$ and $\mathbf{N}$ consist of independent components, the formulation in (2.8) leads to

$$p_{N_l}^{\text{opt}}(n) = \delta(n - n_l)\ , \qquad n_l = \arg\max_{n} \sum_{i=0}^{M-1} \frac{\left(G_{l,i}^{\theta}(n)\right)^2}{H_{l,i}^{\theta}(n)}\ , \tag{2.21}$$

for $l = 1, \ldots, L$, where

$$H_{i,l}^{\theta}(n) \triangleq \mathrm{P}(\tau_i - n < X_l \leq \tau_{i+1} - n)\ , \tag{2.22}$$

$$G_{i,l}^{\theta}(n) \triangleq \partial H_{i,l}^{\theta}(n)/\partial\theta\ . \tag{2.23}$$

In other words, optimal additive noise can be calculated for each component separately in that case.

2.3 Optimal Additive Noise in the Presence of Prior Information

In Section 2.2, the optimal additive noise is calculated for a given value of θ. Although the value of θ is unknown in practice, the theoretical analysis in the previous section is useful in two aspects. First, it provides theoretical performance limits for unbiased estimators that perform parameter estimation based on quantized observations. In other words, the maximum Fisher information at the output of the quantizer in Figure 2.1 is obtained when the optimal additive noise specified by Proposition 1 is employed for each value of θ. Second, the theoretical results in the previous section form a basis for more practical results, and the ideas can be extended to the cases of unknown parameters. In the following, it is assumed that the exact value of θ is unknown, but its p.d.f., denoted by $w(\theta)$, is known *a priori*. Then, it is shown that the results in Lemma 1 and Proposition 1 can be extended to characterize the optimal additive noise.

In the presence of prior p.d.f. $w(\theta)$ for the unknown parameter θ, the Bayesian CRLB (BCRLB), also known as the posterior CRLB [26], imposes a lower bound on the MSE of any estimator $\hat{\theta}$, which can be a biased or unbiased estimator, as [25], [27], [28]

$$\mathrm{MSE}\{\hat{\theta}\} = \mathrm{E}\{(\hat{\theta}(\mathbf{y}) - \theta)^2\} \geq (\mathrm{J_D} + \mathrm{J_P})^{-1} \ , \tag{2.24}$$

where $\mathrm{J_D}$ and $\mathrm{J_P}$ represent the information obtained from the data (observations) and from the prior knowledge, respectively, and are given by

$$\mathrm{J_D} = \mathrm{E}\left\{\left(\frac{\partial \log p_{\mathbf{Y}}(\mathbf{y};\theta)}{\partial\theta}\right)^2\right\} \ , \quad \mathrm{J_P} = \mathrm{E}\left\{\left(\frac{\partial \log w(\theta)}{\partial\theta}\right)^2\right\} \ . \tag{2.25}$$

It is important to note that $\mathrm{J_D}$ in (2.25) differs from J_θ in (2.4) due to the fact that the expectation is over both $\mathbf{y}$ and θ in the former whereas it is only over $\mathbf{y}$ in the latter.

Since $\mathrm{J_P}$ depends only on the prior p.d.f., it is independent of the additive noise component. Therefore, the optimal additive noise p.d.f. is defined to be the one that maximizes $\mathrm{J_D}$. Then, similar to (2.5) and (2.6), the optimal additive noise p.d.f. can be formulated as

$$p_{\mathbf{N}}^{\mathrm{opt}}(\mathbf{n}) = \arg\max_{p_{\mathbf{N}}(\cdot)} \int w(\theta) \sum_{\boldsymbol{i}\in\mathcal{I}} \frac{1}{p_{\mathbf{Y}}(\boldsymbol{i}\,;\theta)} \left(\frac{\partial p_{\mathbf{Y}}(\boldsymbol{i}\,;\theta)}{\partial\theta}\right)^2 d\theta \ . \tag{2.26}$$

In other words, the aim now becomes maximizing the average of Fisher information J_θ (cf. (2.4)-(2.6)) for different parameter values. Since $p_{\mathbf{Y}}(\boldsymbol{i}\,;\theta) = \mathrm{E}\{H_{\boldsymbol{i}}^{\theta}(\mathbf{N})\}$ and $\frac{\partial p_{\mathbf{Y}}(\boldsymbol{i}\,;\theta)}{\partial\theta} = \mathrm{E}\{G_{\boldsymbol{i}}^{\theta}(\mathbf{N})\}$ as defined in Section 2.2, (2.26) can also be expressed as

$$p_{\mathbf{N}}^{\mathrm{opt}}(\mathbf{n}) = \arg\max_{p_{\mathbf{N}}(\cdot)} \int w(\theta) \sum_{\boldsymbol{i}\in\mathcal{I}} \frac{\left(\mathrm{E}\left\{G_{\boldsymbol{i}}^{\theta}(\mathbf{N})\right\}\right)^2}{\mathrm{E}\left\{H_{\boldsymbol{i}}^{\theta}(\mathbf{N})\right\}} d\theta \ . \tag{2.27}$$

Then, the following proposition presents the p.d.f. of the optimal additive noise.

Proposition 2: *The optimal noise p.d.f. in* (2.27) *can be expressed as* $p_{\mathbf{N}}^{\mathrm{opt}}(\mathbf{n}) = \delta(\mathbf{n}-\mathbf{n}_{\mathrm{o}})$, *where*

$$\mathbf{n}_{\mathrm{o}} = \arg\max_{\mathbf{n}} \int w(\theta) \sum_{\boldsymbol{i}\in\mathcal{I}} \frac{\left(G_{\boldsymbol{i}}^{\theta}(\mathbf{n})\right)^2}{H_{\boldsymbol{i}}^{\theta}(\mathbf{n})} d\theta \ . \tag{2.28}$$

Proof: Consider the inequality in (2.17), which is valid for all θ and $p_{\mathbf{N}}(\cdot)$. Since it holds for all θ values, the following inequality can be obtained:

$$\int w(\theta) \sum_{\boldsymbol{i}\in\mathcal{I}} \frac{\left(\mathrm{E}\{G_{\boldsymbol{i}}^{\theta}(\mathbf{N})\}\right)^2}{\mathrm{E}\{H_{\boldsymbol{i}}^{\theta}(\mathbf{N})\}} d\theta \le \mathrm{E}\left\{\int w(\theta) \sum_{\boldsymbol{i}\in\mathcal{I}} \frac{\left(G_{\boldsymbol{i}}^{\theta}(\mathbf{N})\right)^2}{H_{\boldsymbol{i}}^{\theta}(\mathbf{N})} d\theta\right\} \tag{2.29}$$

for all $p_{\mathbf{N}}(\cdot)$. Therefore, the maximum value of the objective function in (2.27) can be bounded from above as

$$\max_{p_{\mathbf{N}}(\cdot)} \int w(\theta) \sum_{\boldsymbol{i}\in\mathcal{I}} \frac{\left(\mathrm{E}\{G_{\boldsymbol{i}}^{\theta}(\mathbf{N})\}\right)^2}{\mathrm{E}\{H_{\boldsymbol{i}}^{\theta}(\mathbf{N})\}} d\theta \le \max_{p_{\mathbf{N}}(\cdot)} \mathrm{E}\left\{\int w(\theta) \sum_{\boldsymbol{i}\in\mathcal{I}} \frac{\left(G_{\boldsymbol{i}}^{\theta}(\mathbf{N})\right)^2}{H_{\boldsymbol{i}}^{\theta}(\mathbf{N})} d\theta\right\} . \tag{2.30}$$

Since the upper bound in (2.30) is always smaller than or equal to $\max_{\mathbf{n}} \left\{ \int w(\theta) \sum_{i \in \mathcal{I}} \frac{\left(G_i^\theta(\mathbf{n})\right)^2}{H_i^\theta(\mathbf{n})} d\theta \right\}$, the following result is obtained:

$$\max_{p_{\mathbf{N}}(\cdot)} \int w(\theta) \sum_{i \in \mathcal{I}} \frac{\left(\mathrm{E}\{G_i^\theta(\mathbf{N})\}\right)^2}{\mathrm{E}\{H_i^\theta(\mathbf{N})\}} d\theta \leq \max_{\mathbf{n}} \left\{ \int w(\theta) \sum_{i \in \mathcal{I}} \frac{\left(G_i^\theta(\mathbf{n})\right)^2}{H_i^\theta(\mathbf{n})} d\theta \right\}$$

$$= \int w(\theta) \sum_{i \in \mathcal{I}} \frac{\left(G_i^\theta(\mathbf{n}_o)\right)^2}{H_i^\theta(\mathbf{n}_o)} d\theta \, , \qquad (2.31)$$

where $\mathbf{n}_o$ is as defined in (2.28). Since the upper bound in (2.31) can be achieved for $p_{\mathbf{N}}(\mathbf{n}) = \delta(\mathbf{n} - \mathbf{n}_o)$, the result in the proposition is obtained. □

Proposition 2 states that among all possible p.d.f.s for the additive noise components, a p.d.f. with a single mass point (that is, a constant "noise" component) minimizes the BCRLB. Therefore, adding the optimum noise to the observation is equivalent to shifting the threshold levels of the quantizer, which is a simple operation since no randomization among different noise values is needed.

2.4 Numerical Results

2.4.1 CRLB Optimization for Different Parameter Types

In this section, we investigate three examples, in which different types of parameters in the scalar observations (which have symmetric Gaussian mixture probability distribution consisting of two components) are to be estimated. In addition, the additive noise taken as a constant signal as the consequence of Proposition 1.

Example 1. Mean of Symmetric Gaussian Mixture Components

Consider a scalar observation x in Figure 2.1 with a Gaussian mixture p.d.f. given by

$$p_X(x;\theta) = 0.5\gamma(x;-\theta,\sigma^2) + 0.5\gamma(x;\theta,\sigma^2)\ , \tag{2.32}$$

where

$$\gamma(x;\theta,\sigma^2) \triangleq \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left\{-\frac{(x-\theta)^2}{2\sigma^2}\right\}\ . \tag{2.33}$$

Then, $H_i^\theta(n)$ in (2.9) can be expressed as

$$H_i^\theta(n) = F_X(\tau_{i+1}-n;\theta) - F_X(\tau_i - n;\theta) \tag{2.34}$$

for $i = 0, 1, \ldots, M-1$, where the cumulative distribution function (c.d.f.) of X for a given value of θ is calculated as

$$F_X(x;\theta) = 0.5\,Q\left(\frac{-x+\theta}{\sigma}\right) + 0.5\,Q\left(\frac{-x-\theta}{\sigma}\right), \tag{2.35}$$

with $Q(a) = \frac{1}{\sqrt{2\pi}}\int_a^\infty \mathrm{e}^{-0.5t^2}dt$ denoting the Q-function. Also, $G_i^\theta(n)$ in (2.10) can be calculated as the derivative of $H_i^\theta(n)$ with respect to θ. In addition, the quantizer in (2.2) is modeled as a 4-level quantizer (i.e., $M = 4$) specified by thresholds $\tau_1 = -3$, $\tau_2 = 0$ and $\tau_3 = 3$.

First, optimal additive noise is investigated for given values of θ. The plot in Figure 2.2 investigates the CRLB versus constant "noise" levels for $\theta = 1$ and $\theta = 3$, where $\sigma = 1$ is used. Specifically, the inverse of the objective function in (2.12) is plotted against the additive "noise" level, n. It is observed for $\theta = 3$ that the optimal additive "noise" value is equal to zero, which means that the additive "noise" cannot reduce the CRLB of the system in that case. However, for $\theta = 1$, the minimum CRLB is achieved for $n = \pm 1.496$, which shows that additive "noise" n can result in a smaller CRLB. In addition, Figure 2.3 plots the CRLB versus θ for various values of the additive "noise", n. It is observed that the minimum CRLB is achieved by different n values over different ranges

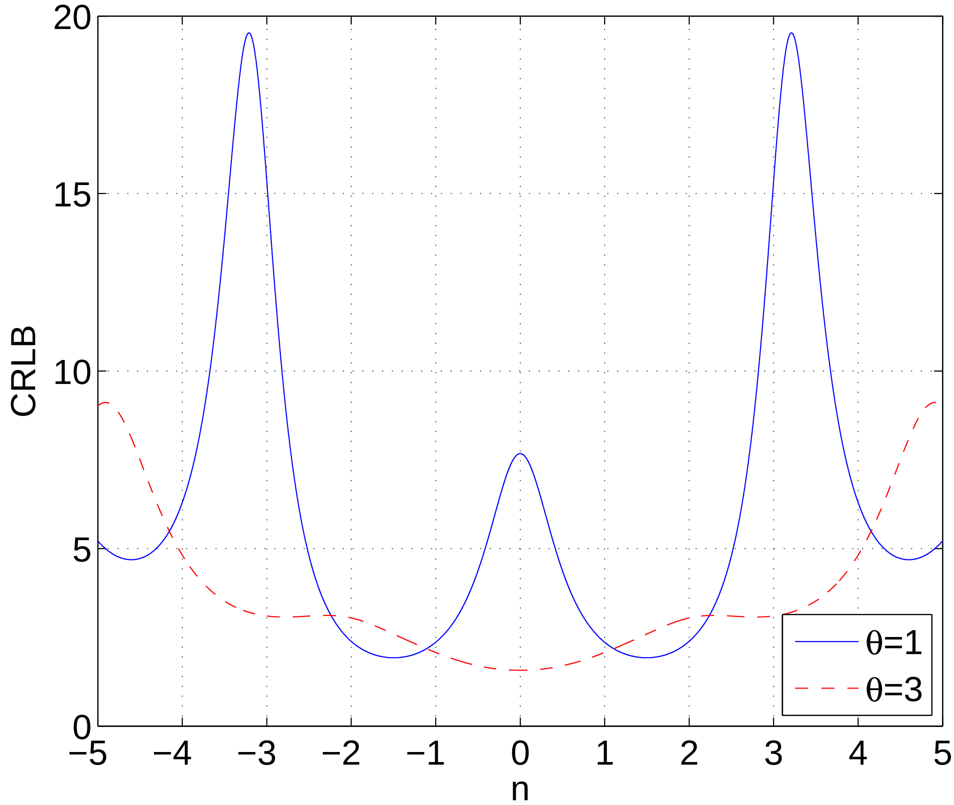

Figure 2.2: Example 1: CRLB versus additive "noise" n for various values of the mean parameter θ.

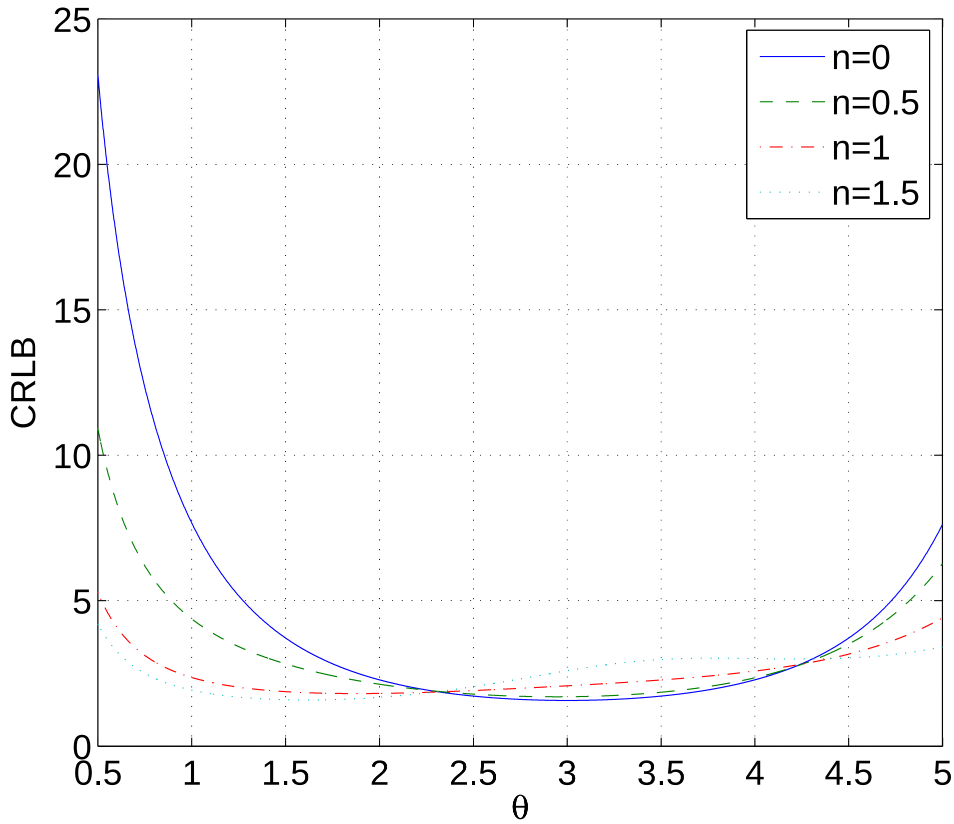

Figure 2.3: Example 1: CRLB versus θ for various values of additive "noise" n.

of parameter θ. It can also be concluded that if a rough estimate of θ is available beforehand, an n value that is optimal around that estimate can be selected as a (close-to) optimal additive "noise" component for the given estimation problem.

In addition, Figure 2.4 illustrates CRLB versus σ for $n = 0$ and $n = n_{\text{opt}}$, where $\theta = 1$ is used. It is observed that no additive noise is required to minimize the CRLB for $1.9 \leq \sigma \leq 4.7$. Otherwise, the CRLB is improvable. It can be concluded that the improvability of the CRLB for a given value of a parameter depends on the probability distribution of the observation. As shown in [20, 22, 23], it is possible to improve the estimation accuracy by increasing the variance of the observation, which can be achieved via Gaussian dithering in this example as explained in Section 2.4.2. However, increasing the variance after adding the optimal constant signal (noise) degrades the estimation performance.

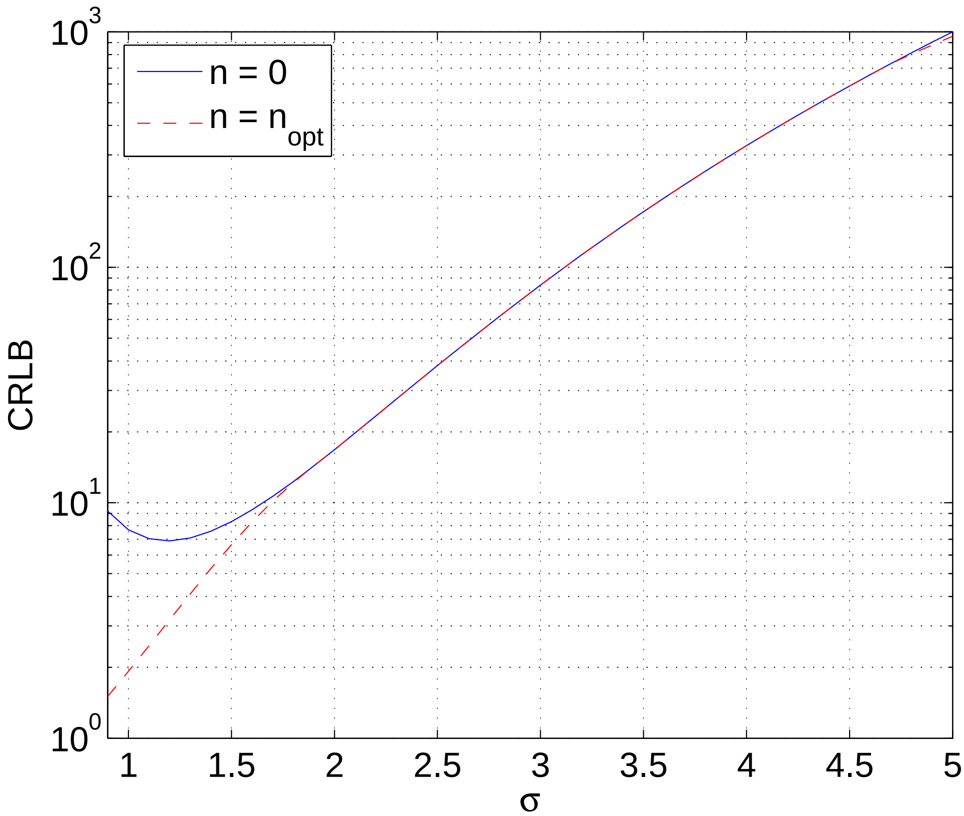

Figure 2.4: Example 1: CRLB versus σ for $n = 0$ and $n = n_{\rm opt}$.

Next, for the problem setting described above, it is assumed that the prior p.d.f. of θ is specified as

$$w(\theta) = \lambda \exp\{-\lambda\theta\} \tag{2.36}$$

for $\theta \in [0, \infty)$, where $\lambda = 1$. From (2.25), the Fisher information obtained from the prior information is calculated as

$$\begin{aligned} \mathrm{J_P} &= \lambda^2 & (2.37) \\ &= 1. & (2.38) \end{aligned}$$

In Figure 2.5, the BCRLB is plotted versus n, where the BCRLB is calculated as $(\mathrm{J_P} + \mathrm{J_D})^{-1}$, with $\mathrm{J_D}$ denoting the value of the objective function in (2.28) for various values of n. It is observed from the figure that the minimum BCRLB is achieved at $n = \pm 1.463$. In addition, since there exists prior information in this scenario, the theoretical limits are lower than those in the previous scenario in which no prior information on θ exists.

Example 2. Mean of Symmetric Gaussian Distributed Observation

In the second example, we use the same problem setting as the previous one except that the scalar observation x has the following probability distribution:

$$p_X(x;\theta) = 0.5\gamma(x; -\mu - \theta, \sigma^2) + 0.5\gamma(x; \mu - \theta, \sigma^2)\ . \tag{2.39}$$

In this case, the c.d.f. of X for a given value of θ in (2.9) can be expressed as

$$F_X(x;\theta) = 0.5\, Q\left(\frac{-x - \mu - \theta}{\sigma}\right) + 0.5\, Q\left(\frac{-x + \mu - \theta}{\sigma}\right)\ . \tag{2.40}$$

Here, θ is a location parameter, which implies

$$p_X(x;\theta) = p_X(x - \theta)\ . \tag{2.41}$$

In addition, $G_i^\theta(n)$ and $H_i^\theta(n)$ become

$$G_i^\theta(n) = p_X(\tau_{i+1} - n - \theta) - p_X(\tau_i - n - \theta) \tag{2.42}$$

$$H_i^\theta(n) = F_X(\tau_{i+1} - n - \theta) - F_X(\tau_i - n - \theta) \tag{2.43}$$

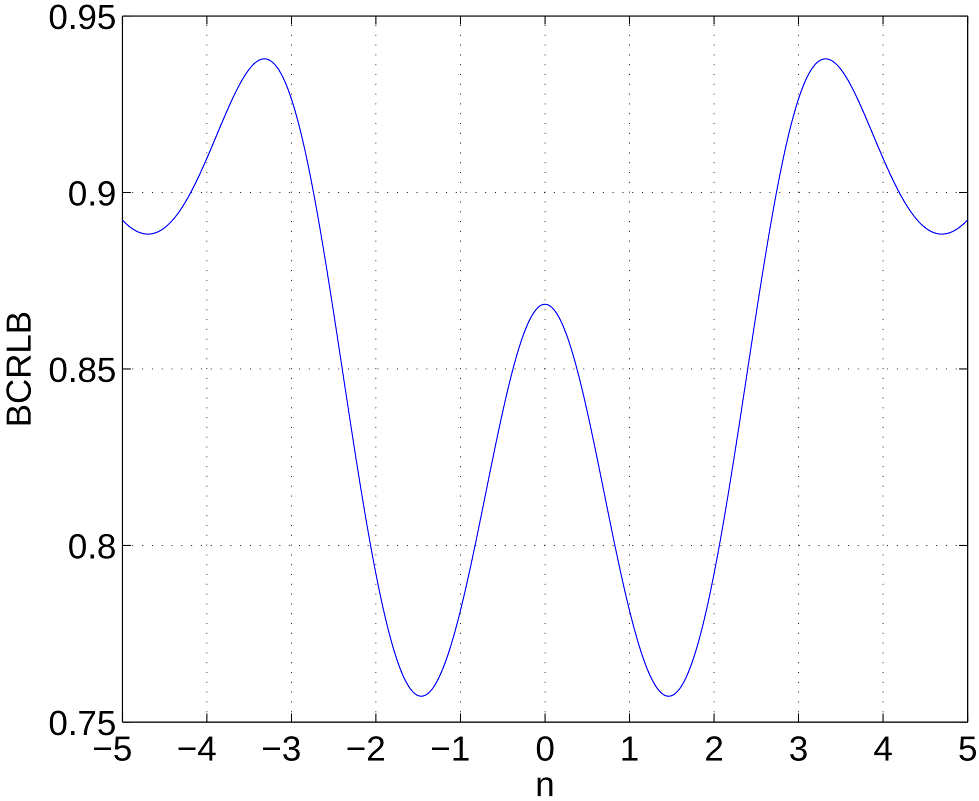

Figure 2.5: Example 1: BCRLB versus n when θ is Gaussian distributed with unit mean and variance.

using (2.41). As a result, the Fisher information for a given θ can be expressed as

$$\mathrm{J}_\theta(n) = \sum_{i=0}^{3} \frac{\left(p_X(\tau_{i+1} - n - \theta) - p_X(\tau_i - n - \theta)\right)^2}{F_X(\tau_{i+1} - n - \theta) - F_X(\tau_i - n - \theta)} \,, \tag{2.44}$$

and

$$\mathrm{J}_\theta(n) = \mathrm{J}(n + \theta) \tag{2.45}$$

is valid. Hence, the optimal noise minimizing the CRLB for a given θ depends on the value of θ such that $\mathrm{J}(n_{\mathrm{opt}} + \theta)^{-1}$ gives the minimum CRLB. Plotting CRLB versus n for $\theta = 0$ and $\theta = 0.5$, where $\mu = 1$ and $\sigma = 1$ are used, we observe that optimum additive "noise" values are found as $n = \pm 1.49$ and $n = 0.5 \pm 1.49$ respectively, as expected. The result of using a location parameter to be estimated is clearly illustrated in Figure 2.7 for different additive "noise" levels and θ. It can be concluded that the sum of the additive "noise" and θ determines the CRLB, if θ is a location parameter. Therefore, the amount of change in the optimal additive "noise" is the same as the parameter. Additionally, the variation of the optimal additive "noise" with respect to the standard deviation of the Gaussian mixture components can be seen in Figure 2.8. It is seen that no additive "noise" is needed for $\sigma \geq 1.59$. The conclusions for the Figure 2.4 are also valid for Figure 2.8.

Next, we assume that θ is random and has the p.d.f.

$$w(\theta) = \exp\left\{-(\theta - \mu_\theta)^2/(2\sigma_\theta^2)\right\} / (\sqrt{2\pi}\,\sigma_\theta) \,, \tag{2.46}$$

where $\mu_\theta = 0$ and $\sigma_\theta = 0.2$. From (2.25), it can be shown that $\mathrm{J}_{\mathrm{P}} = \sigma_\theta^{-2} = 25$. The behavior of the BCRLB with respect to the additive "noise" is plotted in Figure 2.9. It is observed from the figure that the minimum BCRLB is achieved at $n = \pm 1.487$.

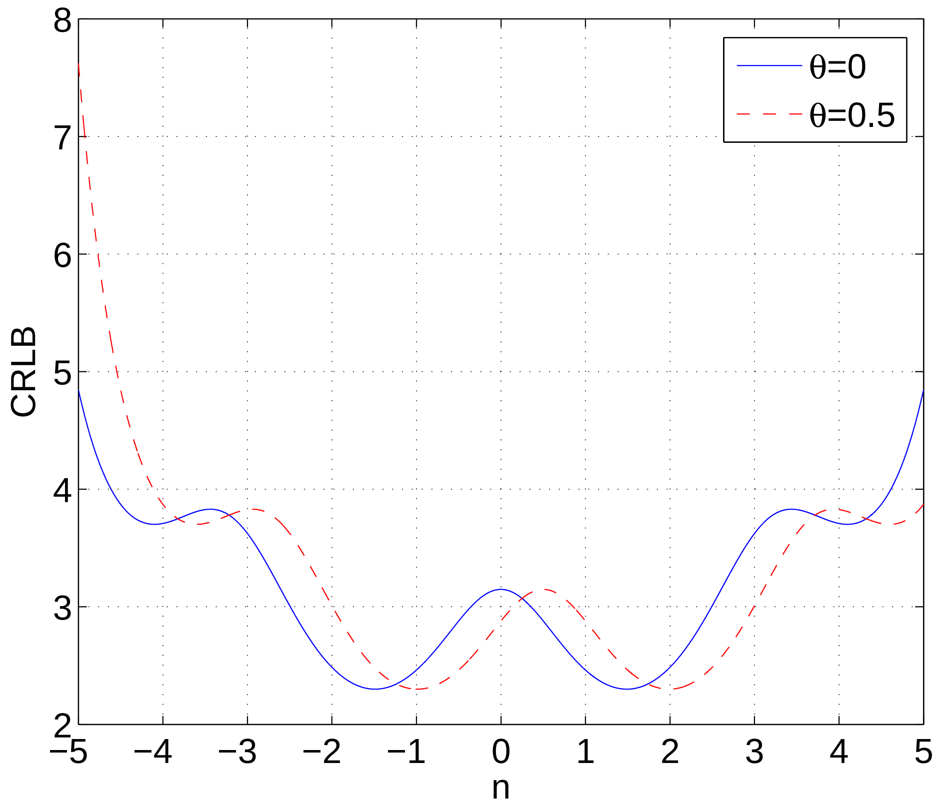

Figure 2.6: Example 2: CRLB versus additive "noise" n for various values of the mean-shift parameter θ.

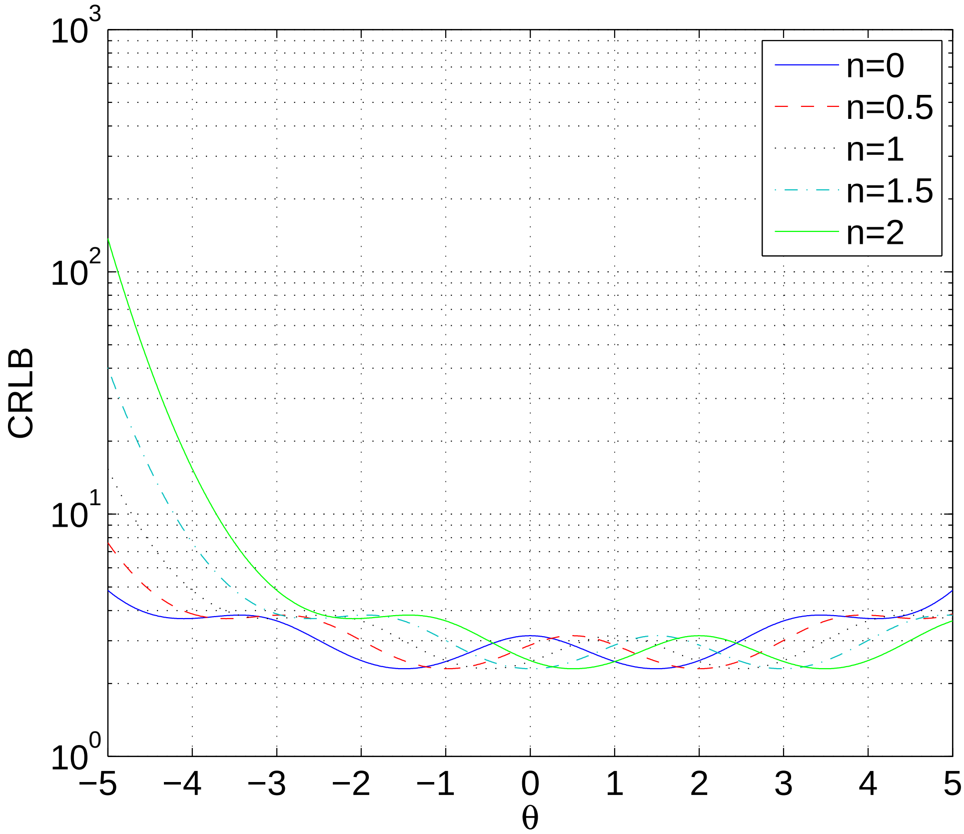

Figure 2.7: Example 2: CRLB versus θ for various values of additive "noise" n.

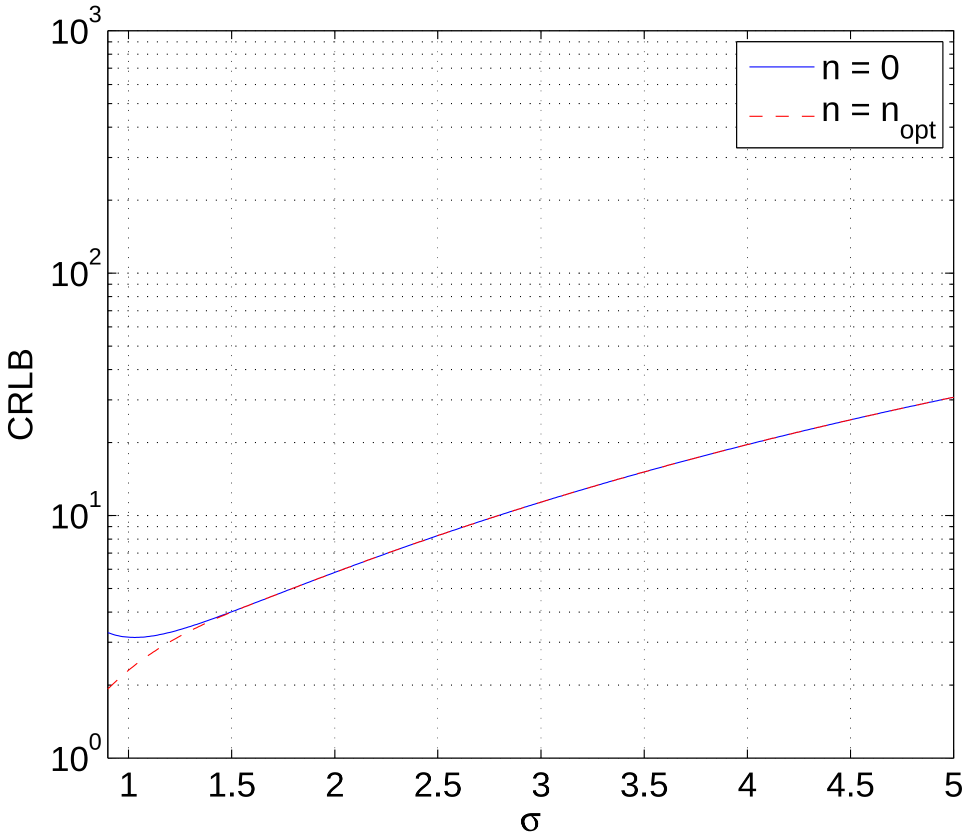

Figure 2.8: Example 2: CRLB versus σ for $n = 0$ and $n = n_{\text{opt}}$.

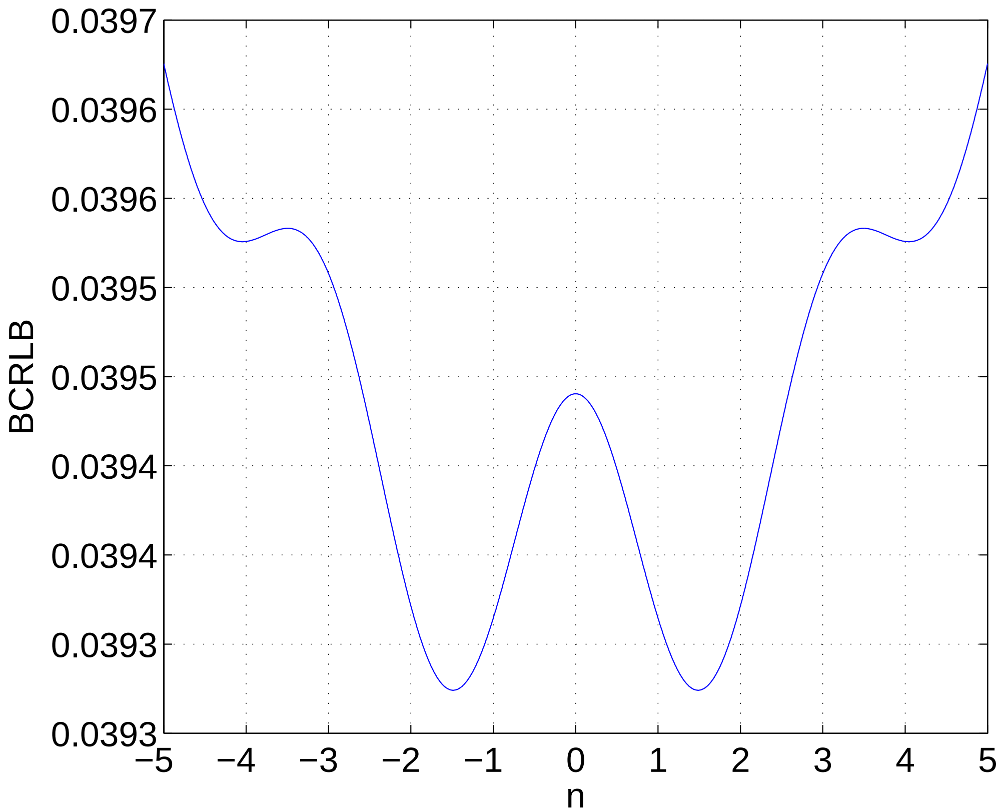

Figure 2.9: Example 2: BCRLB versus additive "noise" n for various values of the mean-shift parameter θ.

Example 3. Variance of Symmetric Gaussian Distributed Observation

In our third example, we consider a scalar observation x, whose p.d.f. and c.d.f. are given by

$$p_X(x;\theta) = 0.5\gamma(x;-\mu,\theta^2) + 0.5\gamma(x;\mu,\theta^2) \ , \tag{2.47}$$

and

$$F_X(x;\theta) = 0.5Q\left(\frac{-x+\mu}{\theta}\right) + 0.5Q\left(\frac{-x-\mu}{\theta}\right) \ , \tag{2.48}$$

respectively, where $\mu = 0.2$. This time, the threshold values of the 4-level quantizer are set to $\tau_1 = -1$, $\tau_2 = 0$ and $\tau_3 = 1$. In Figure 2.10, the CRLB is plotted versus additive "noise" for $\theta = 0.3$ and $\theta = 1$. For $\theta = 0.3$, it is observed that the CRLB is minimized by the additive noise $n = \pm 0.498$. However, for $\theta = 1$, the additive "noise" level required for CRLB minimization is zero. In addition, Figure 2.11 depicts the CRLB versus θ for different noise levels. Similar to Figure 2.3 in the Example 1, it is observed that the additive "noise" level required to minimize the CRLB changes for different values of θ. This result can be also seen in Figure 2.12, where the optimal additive "noise" level differs from zero for $0.51 \leq \sigma \leq 1.51$. Since the behavior of CRLB versus σ for $n = 0$ and $n = n_{\text{opt}}$ is similar to Figures 2.4 and 2.8, we can draw the same conclusions for Figure 2.12.

Assuming that θ is a random parameter having exponential distribution with parameter λ; that is,

$$w(\theta) = \lambda \exp\{-\lambda(\theta - \zeta)\} \ , \tag{2.49}$$

where $\theta \in [\zeta, \infty)$ and $\zeta \in \mathrm{R}^+$ is the shift variable, we consider the BCRLB for the estimate of θ. Choosing $\lambda = 1$ and $\zeta = 0.3$, the information obtained from the prior knowledge is computed as

$$\mathrm{J_P} = \lambda^2 \tag{2.50}$$

$$= 1 \ . \tag{2.51}$$

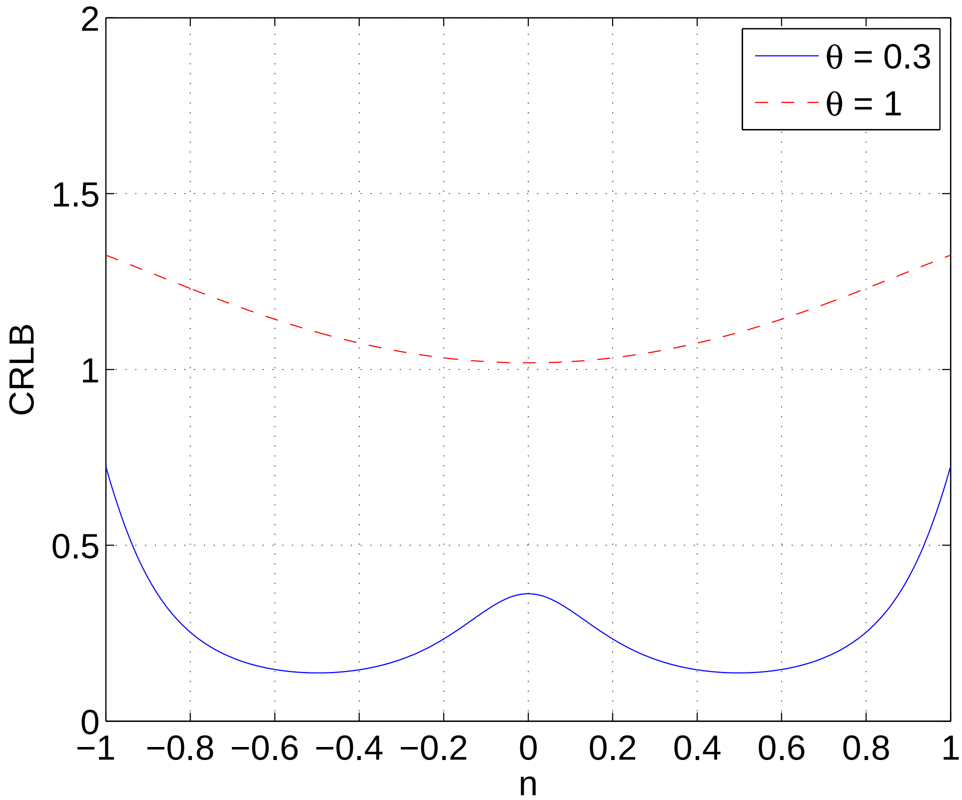

Figure 2.10: Example 3: CRLB versus additive “noise” n for various values of the standard deviation of the Gaussian mixture components θ.

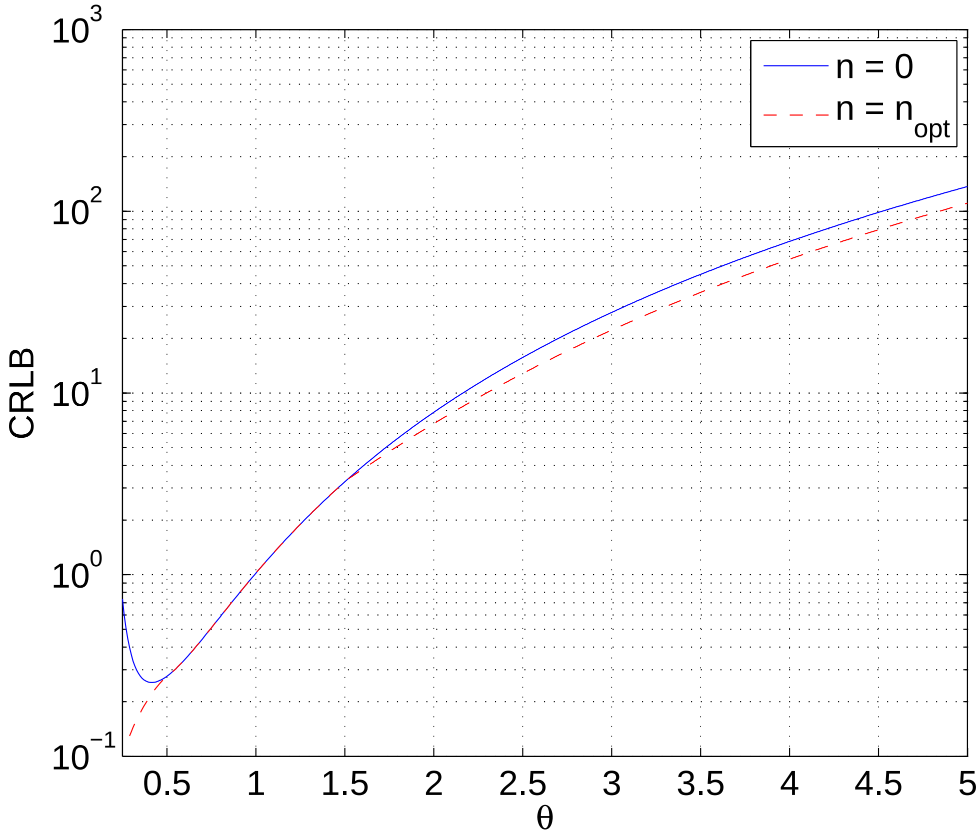

Figure 2.11: Example 3: CRLB versus θ for various values of additive "noise" n.

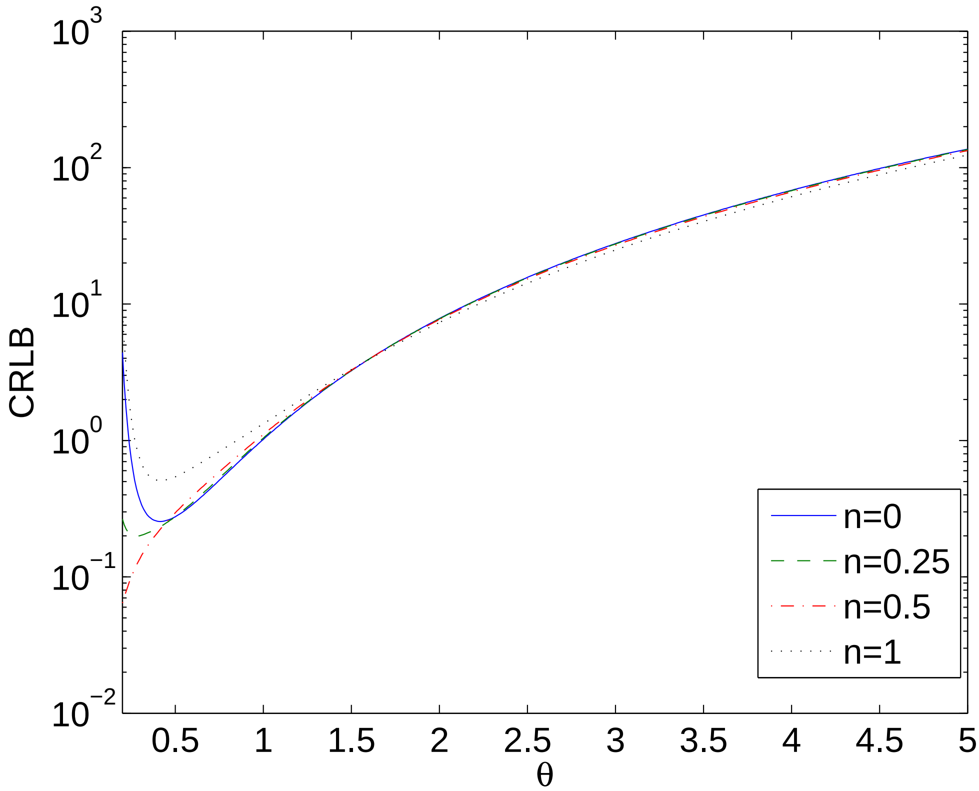

Figure 2.12: Example 3: CRLB versus θ for $n = 0$ and $n = n_{\mathrm{opt}}$.

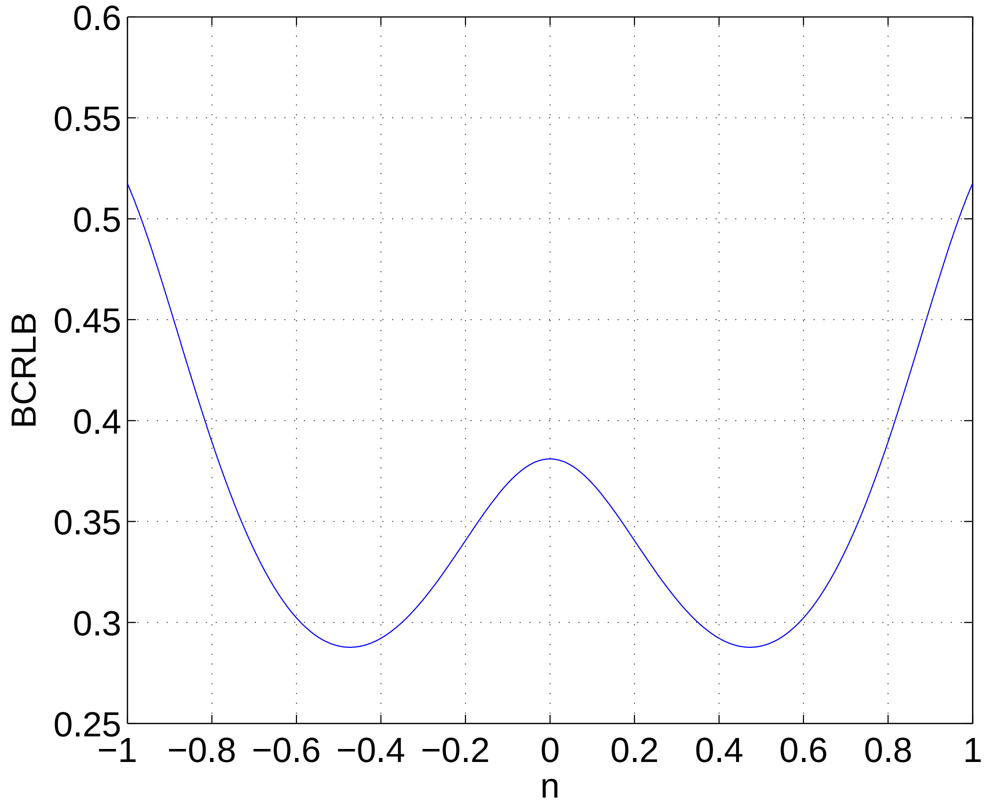

Figure 2.13: Example 3: BCRLB versus additive "noise" n for various values of the standard deviation parameter θ.

In Figure 2.13, the variation of the BCRLB with respect to n is shown, and it is observed that the minimum BCRLB is achieved at $n = \pm 0.4730$.

2.4.2 Comparison with Common Dithering Techniques

In some related studies in the literature, the benefits of additive "noise" in non-linear systems are observed by employing random noise, which can be Gaussian or uniformly distributed [12], [15], [20], [16], [22], [23]. In this section, we compare the optimal CRLB values obtained with optimal additive constant signal to the additive noise models, which are used in common dithering techniques, namely, Gaussian dithering and uniform dithering [20, 29]. As a Gaussian dither

Table 2.1: Optimal Gaussian dithering and uniform dithering versus optimal additive "noise" for Example 1.

	σ_N^{opt}	$\epsilon = 1$	$\epsilon = 0.5$	$\epsilon = 0.25$	$\epsilon = 0$	Optimal
CRLB ($\theta = 1$)	6.888 (σ_N^{opt}=0.645)	6.566	7.302	7.575	7.675	1.924
CRLB ($\theta = 3$)	1.571 (σ_N^{opt}=0)	2.146	1.705	1.604	1.571	1.571
BCRLB	0.8683 (σ_N^{opt}=0)	0.8762	0.8705	0.8689	0.8683	0.7573

signal, zero mean additive Gaussian noise with a standard deviation σ_N is employed. Since the random observations in our examples in the previous section have a Gaussian mixture distribution, the standard deviation of the sum of the observation and the additive noise can be described as

$$\sigma_{X+N} = \sqrt{\sigma^2 + \sigma_N^2} \tag{2.52}$$

where σ is the variance of the Gaussian mixture components of X. The standard deviation of the optimal additive Gaussian noise can be found as

$$\sigma_N^{\text{opt}} = \sqrt{(\sigma_{X+N}^{\text{opt}})^2 - \sigma^2} \tag{2.53}$$

where $(\sigma_{X+N}^{\text{opt}})^2$ represents the variance of the observation combined with the optimal noise. Since adding zero mean additive Gaussian noise has the same effect as increasing the variance, we can consider Figures 2.4, 2.8 and 2.12 as a comparison of the effects of the additive Gaussian noise and additive constant signal on the CRLB. In these figures, we can also consider the σ value yielding the minimum CRLB as $\sigma_{X+N}^{\text{opt}}$. Using σ values in these examples, we can find σ_N^{opt} for the optimal additive Gaussian noise. In addition to the Gaussian noise, additive uniform noise between $-\epsilon$ and ϵ is compared to additive constant noise. The results in Table 2.1, 2.2 and 2.3 reveal that the performance improvement in single parameter estimation by additive constant noise is significantly superior to Gaussian and uniform dithering.

Table 2.2: Optimal Gaussian dithering and uniform dithering versus optimal additive "noise" for Example 2.

	σ_N^{opt}	$\epsilon = 1$	$\epsilon = 0.5$	$\epsilon = 0.25$	$\epsilon = 0$	Optimal
CRLB ($\theta = 0$)	3.142 (σ_N^{opt}=0.247)	3.162	3.136	3.144	3.148	2.300
CRLB ($\theta = 0.5$)	2.880 (σ_N^{opt}=0)	3.087	2.929	2.892	2.880	2.300
BCRLB	0.0395 (σ_N^{opt}=0.142)	0.0395	0.0395	0.0395	0.0395	0.0393

Table 2.3: Optimal Gaussian dithering and uniform dithering versus optimal additive "noise" for Example 3.

	σ_N^{opt}	$\epsilon = 1$	$\epsilon = 0.5$	$\epsilon = 0.25$	$\epsilon = 0$	Optimal
CRLB ($\theta = 0.3$)	0.2551 (σ_N^{opt}=0.279)	1.218	0.352	0.3411	0.3621	0.1369
CRLB ($\theta = 1$)	1.0186 (σ_N^{opt}=0)	2.149	1.234	1.069	1.0186	1.0186
BCRLB	0.3810 (σ_N^{opt}=0)	0.6425	0.4266	0.3919	0.3810	0.2877

2.4.3 ML and MAP Estimation Performance

For the estimation performance evaluation in practical cases, we compare our results with the performance of the maximum likelihood (ML) estimator for the fixed parameter case and maximum a-posteriori probability (MAP) estimator for the random parameter case. ML and MAP estimates are known to be asymptotically efficient [25]. This means that

$$\lim_{l \to +\infty} \mathrm{E}\big\{(\hat{\theta}_{ML}(\mathbf{y}) - \theta)^2\big\} = \mathrm{J_D}^{-1} \tag{2.54}$$

and

$$\lim_{l \to +\infty} \mathrm{E}\big\{(\hat{\theta}_{MAP}(\mathbf{y}) - \theta)^2\big\} = (\mathrm{J_D} + \mathrm{J_P})^{-1} \tag{2.55}$$

for $l = 1, \ldots, L$, where L is the number of observations, $\mathbf{y} = [y_1\ y_2 \cdots y_L]$ and $\hat{\theta}_{ML}(\mathbf{y})$ and $\hat{\theta}_{MAP}(\mathbf{y})$ are the ML and MAP estimates of parameter θ, respectively. Therefore, it is expected that the asymptotical performance of both estimators will improve with the reduced CRLB and BCRLB. The ML and MAP estimates for a parameter θ are defined as

$$\hat{\theta}_{ML}(\mathbf{y}) = \arg\max_{\theta}\, p_{\mathbf{Y}}(\mathbf{y};\theta) \tag{2.56}$$

and

$$\hat{\theta}_{MAP}(\mathbf{y}) = \arg\max_{\theta} p_{\mathbf{Y}}(\mathbf{y};\theta)w(\theta) \ , \tag{2.57}$$

respectively. For the i.i.d. case of the observations, the p.d.f. of $\mathbf{Y}$ is calculated as $p_{\mathbf{Y}}(\mathbf{y};\theta) = \prod_{l=1}^{L} p_Y(y;\theta)$ and the Fisher information obtained from the data becomes $\mathrm{J_D} = L\mathrm{J}_\theta$, where J is the Fisher information obtained from one observation Y. The probability distribution of Y can be expressed as

$$p_Y(i;\theta) = F_X(\tau_{i+1} - n;\theta) - F_X(\tau_i - n;\theta) \ . \tag{2.58}$$

For the fixed and random parameter cases, we have performed a series of Monte Carlo trials in order to evaluate the MSEs of the ML and MAP estimates of parameter θ, where the settings of the first example in Section 2.4.1 are employed. For the evaluation of the ML and MAP estimator performance, L realizations of the observation Y are generated for $\theta = 1$ in the fixed parameter case and for an exponential distributed random θ characterized by the p.d.f.

$$w(\theta) = \lambda \exp\{-\lambda\theta\} \ , \tag{2.59}$$

where $\lambda = 1$, in the random parameter case. The RMSEs of both estimates with and without optimal noise enhancement are compared to their lower bounds in Figure 2.14 and Figure 2.15[1]. The asymptotic efficiency of the ML and MAP estimates are evident in the figures, since they approach to their lower bounds for an increasing number of observations. Furthermore, since noise enhancement reduces the CRLB (BCRLB), it is observed that the MSE performances of the estimators significantly improve. Hence, the optimization of the CRLB using additive noise can be an effective alternative to the optimization of the MSE of the estimate itself.

[1]In Figure 2.14, the RMSE of the ML estimate in the absence of additive noise can get lower than the CRLB for small numbers of observations, since it turns out to be a biased estimator in those cases.

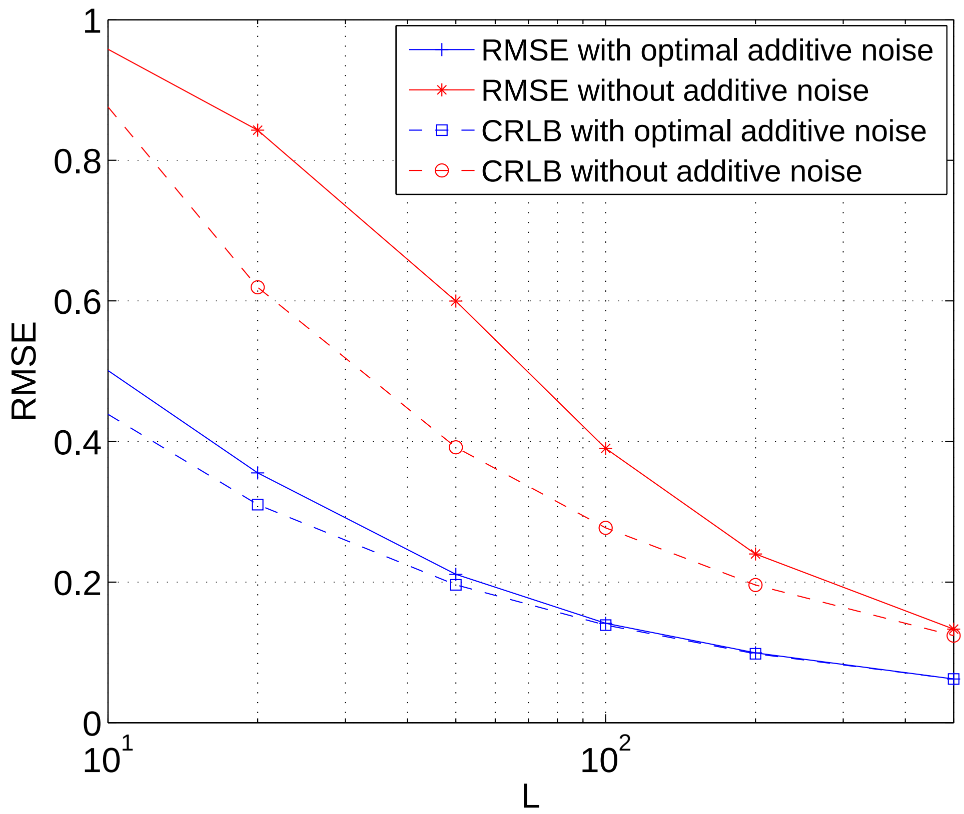

Figure 2.14: RMSE versus CRLB for ML estimates with and without additive "noise". The observations are generated for $\theta = 1$.

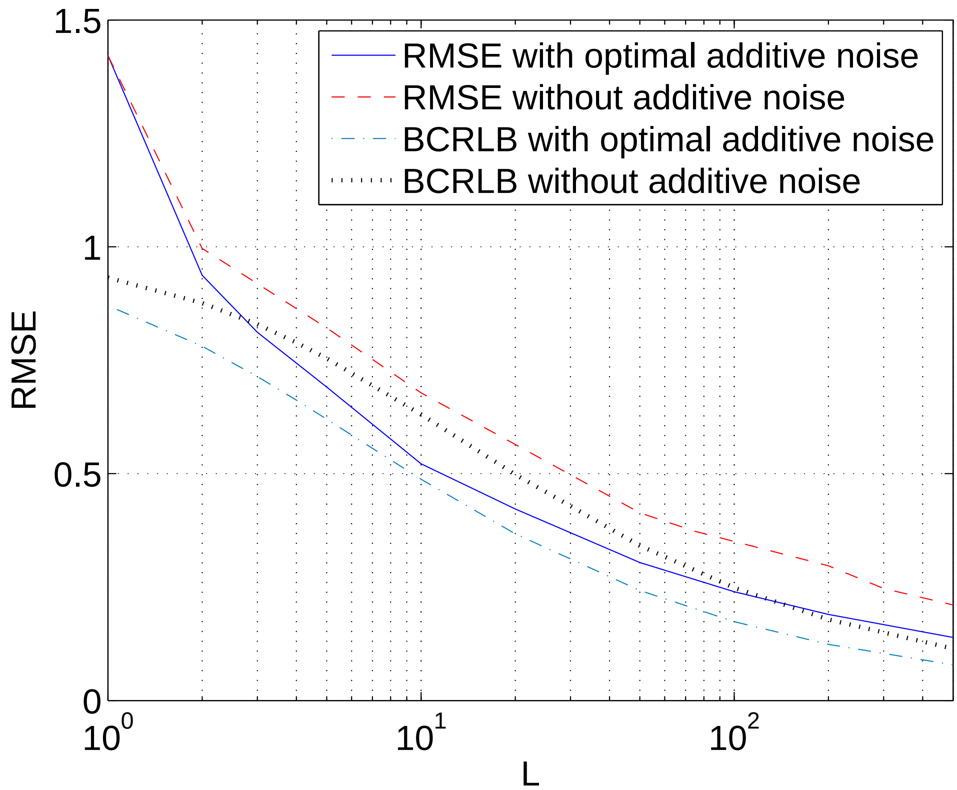

Figure 2.15: RMSE versus BCRLB for MAP estimates with and without additive "noise". The observations are generated for $w(\theta) = \lambda \exp\{-\lambda\theta\}$ with $\theta \in [0, \infty)$.

2.5 Conclusions

In this chapter, it has been proven that in the noise enhanced estimation problem based on quantized observations, the best improvement can be obtained by adding the optimal constant "noise" among all possible dither signals, when the aim is to improve the estimation performance in terms of CRLB. Since Propositions 1 and 2 state that optimal additive "noise" can be represented by a constant signal level, it has been concluded that the CRLB (BCRLB) is minimized by shifting the original observation, which can also be interpreted as shifting the thresholds of the quantizer by a constant value (cf. (2.2)). In other words, among all possible p.d.f.s for the additive noise in Figure 2.1, the ones with a single mass point, i.e., constant "noise" levels, can be used to achieve the minimum CRLB (BCRLB). Therefore, randomization among different noise components are not necessary to obtain the lowest bounds, which is a useful result for practical implementations.

In Section 2.4, where three examples of different parameter types have been investigated, it has been seen that the improvability of the estimation accuracy in terms of CRLB (BCRLB) and the optimal additive "noise" level depends on the probability distribution of the observation. For some observation p.d.f.s, additive noise may degrade the estimation performance. However, this can be interpreted as $p_{\mathbf{N}}^{\text{opt}}(\mathbf{n}) = \delta(\mathbf{n})$, which is still consistent with our theoretical results.

Moreover, the comparison of Gaussian and uniform dithering with optimal additive constant "noise" in the aforementioned examples reveals that the optimal additive constant "noise" outperforms these dithering types in every case, which confirms our theoretical results.

Finally, it has been observed that reducing the CRLB and the BCRLB can yield significant improvements of the MSE performance of asymptotically efficient estimators such as ML and MAP estimators.

Chapter 3

OPTIMAL ADDITIVE NOISE IN MULTIPLE PARAMETER ESTIMATION PROBLEMS

3.1 Problem Formulation

Consider the multi-parameter version of the system in Figure 2.1, where the vector parameter $\boldsymbol{\theta} = [\theta_1 \cdots \theta_K]$ is to be estimated instead of a single parameter. As in the previous chapter, the noise modified version of the observation is to be used as in Figure 3.1 in order to enhance the estimation performance of the system, where the additive noise $\mathbf{n}$ and the observation $\mathbf{x}$ are independent of each other. The aim is the same as in the previous chapter, which is to find the optimal probability distribution of the noise that minimizes the estimation accuracy of the system in Figure 3.1.

In this chapter, the following representations are used: $\mathbf{x}$, $\mathbf{n}$, $\mathbf{y}$ and $\varphi(\cdot)$ are defined as in Section 2.1, but $\mathbf{x}$ and $\mathbf{y}$ are characterized by p.d.f.s $p_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta})$ and $p_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta})$. The relation between the input and the output of the quantizer is

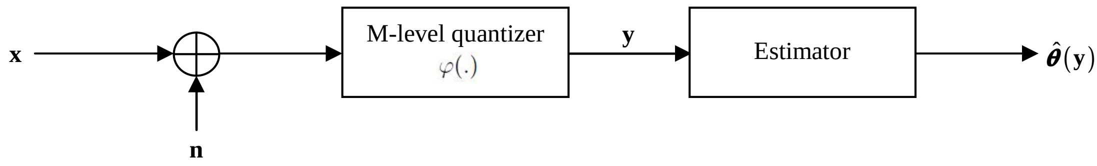

Figure 3.1: The block diagram of the quantization process of the noise enhanced signal and estimation of a set of parameters of the input signal.

described as in the (2.2). Hence, the p.d.f. of $\mathbf{Y}$ can be written as

$$p_{\mathbf{Y}}(\boldsymbol{i}\,;\boldsymbol{\theta}) = \int_{\mathbb{R}^L} \mathrm{P}(\tau_{i_1} - n_1 < X_1 \leq \tau_{i_1+1} - n_1, \ldots, \tau_{i_L} - n_L < X_L \leq \tau_{i_L+1} - n_L)\, p_{\mathbf{N}}(\mathbf{n})\, d\mathbf{n}\,. \tag{3.1}$$

Note that the difference between (3.1) and (2.3) lies in the fact that $\boldsymbol{\theta}$ in (3.1) is a vector parameter.

The aim is to obtain the optimal additive noise p.d.f. that minimizes the CRLB. A generic expression for the CRLB on the covariance matrix of unbiased estimators of $\boldsymbol{\theta}$ is stated as [30]

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}) \geq \mathbf{J}_{\boldsymbol{\theta}}^{-1}\,, \tag{3.2}$$

where $\mathrm{Cov}(\hat{\boldsymbol{\theta}}) \geq \mathbf{J}_{\boldsymbol{\theta}}^{-1}$ means that $\mathrm{Cov}(\boldsymbol{\theta}) - \mathbf{J}_{\boldsymbol{\theta}}^{-1}$ is positive semidefinite, $\mathbf{J}_{\boldsymbol{\theta}}$ is defined as the Fisher information matrix (FIM) given by

$$\mathbf{J}_{\boldsymbol{\theta}} = \mathrm{E}\left\{ (\nabla_{\boldsymbol{\theta}} \log p_{\mathbf{Y}}(\boldsymbol{i}\,;\boldsymbol{\theta}))\,(\nabla_{\boldsymbol{\theta}} \log p_{\mathbf{Y}}(\boldsymbol{i}\,;\boldsymbol{\theta}))^T \right\} \tag{3.3}$$

with

$$\nabla_{\boldsymbol{\theta}} \log p_{\mathbf{Y}}(\boldsymbol{i}\,;\boldsymbol{\theta}) \triangleq \left[\frac{\partial \log p_{\mathbf{Y}}(\boldsymbol{i}\,;\boldsymbol{\theta})}{\partial \theta_1} \quad \cdots \quad \frac{\partial \log p_{\mathbf{Y}}(\boldsymbol{i}\,;\boldsymbol{\theta})}{\partial \theta_K} \right]^T . \tag{3.4}$$

As a special case, if the components of $\mathbf{X}$ and $\mathbf{N}$ are independent, the quantizer output $\mathbf{y}$ has independent components, as well. Therefore, the FIM in (3.3) can be expressed as [30]

$$\mathbf{J}_{\boldsymbol{\theta}} = \sum_{l=1}^{L} \mathbf{J}_{\boldsymbol{\theta}}^{Y_l}\,, \tag{3.5}$$

where $\mathbf{J}_{\boldsymbol{\theta}}^{Y_l}$ represents the FIM due to the lth observation; that is,

$$\mathbf{J}_{\boldsymbol{\theta}}^{Y_l} = \mathrm{E}\left\{(\nabla_{\boldsymbol{\theta}} \log p_{Y_l}(i\,;\boldsymbol{\theta}))\,(\nabla_{\boldsymbol{\theta}} \log p_{Y_l}(i\,;\boldsymbol{\theta}))^T\right\} \ . \tag{3.6}$$

Note that (3.5) reduces to

$$\mathbf{J}_{\boldsymbol{\theta}} = L\,\mathbf{J}_{\boldsymbol{\theta}}^{Y_1}\,, \tag{3.7}$$

when $Y_1, \ldots, Y_L$ are independent and identically distributed (i.i.d.).

The CRLB in (3.2) imposes a lower bound on the mean-squared error (MSE) of an unbiased estimator. Specifically, the MSE of an unbiased estimator is limited by the trace of the CRLB matrix, as shown in the following equations [30]:

$$\begin{aligned} \mathrm{MSE} = \mathrm{E}\left\{\|\hat{\boldsymbol{\theta}}(\mathbf{y}) - \boldsymbol{\theta}\|^2\right\} &= \sum_{i=1}^{K} \mathrm{E}\left\{(\hat{\theta}_i(\mathbf{y}) - \theta_i)^2\right\} && (3.8)\\ &= \sum_{i=1}^{K} \mathrm{Var}(\hat{\theta}_i) && (3.9)\\ &\geq \sum_{i=1}^{K} \left[\mathbf{J}_{\boldsymbol{\theta}}^{-1}\right]_{ii} = \mathrm{trace}\left\{\mathbf{J}_{\boldsymbol{\theta}}^{-1}\right\} \ . && (3.10) \end{aligned}$$

Note that the unbiasedness property of the estimator is employed to obtain (3.9) from (3.8), and (3.2) and (3.5) are used to obtain the lower bound in (3.10). For independent $\mathbf{X}$ and $\mathbf{N}$ components, (3.10) reduces to

$$\mathrm{trace}\left\{\left(\sum_{l=1}^{L} \mathbf{J}_{\boldsymbol{\theta}}^{Y_l}\right)^{-1}\right\} \ . \tag{3.11}$$

From (3.6) and (3.10), the p.d.f. of the optimal additive noise can be calculated from

$$p_{\mathbf{N}}^{\mathrm{opt}}(\mathbf{n}) = \arg\min_{p_{\mathbf{N}}(\cdot)} \ \mathrm{trace}\left\{\left(\mathrm{E}\left\{(\nabla_{\boldsymbol{\theta}} \log p_{\mathbf{Y}}(\boldsymbol{i}\,;\boldsymbol{\theta}))\,(\nabla_{\boldsymbol{\theta}} \log p_{\mathbf{Y}}(\boldsymbol{i}\,;\boldsymbol{\theta}))^T\right\}\right)^{-1}\right\} \ , \tag{3.12}$$

where $p_{\mathbf{Y}}(\cdot\,;\boldsymbol{\theta})$ is as in (3.1). After some manipulation, (3.12) can also be expressed as

$$p_{\mathbf{N}}^{\mathrm{opt}}(\mathbf{n}) = \arg\min_{p_{\mathbf{N}}(\cdot)} \ \mathrm{trace}\left\{\left(\sum_{\boldsymbol{i}\in\mathcal{I}} \frac{1}{p_{\mathbf{Y}}(\boldsymbol{i}\,;\boldsymbol{\theta})}\,\mathbf{D}_{\boldsymbol{\theta}}^{\mathbf{Y},\boldsymbol{i}}\right)^{-1}\right\} \ , \tag{3.13}$$

where $\mathcal{I} \triangleq \{0, 1, \ldots, M-1\}^L$ and $\mathbf{D}_{\boldsymbol{\theta}}^{\mathbf{Y},\boldsymbol{i}}$ is a $K \times K$ matrix with its element in row k_1 and column k_2 being given by

$$\left[\mathbf{D}_{\boldsymbol{\theta}}^{\mathbf{Y},\boldsymbol{i}}\right]_{k_1 k_2} = \frac{\partial p_{\mathbf{Y}}(\boldsymbol{i}\,;\boldsymbol{\theta})}{\partial \theta_{k_1}} \frac{\partial p_{\mathbf{Y}}(\boldsymbol{i}\,;\boldsymbol{\theta})}{\partial \theta_{k_2}} \;. \tag{3.14}$$

For independent $\mathbf{Y}$ components, the optimal additive noise can be characterized with the p.d.f.

$$p_{\mathbf{N}}^{\text{opt}}(\mathbf{n}) = \arg\min_{p_{\mathbf{N}}(\cdot)} \operatorname{trace}\left\{\left(\sum_{l=1}^{L}\sum_{i=0}^{M-1} \frac{1}{p_{Y_l}(i\,;\boldsymbol{\theta})}\,\mathbf{D}_{\boldsymbol{\theta}}^{Y_l,i}\right)^{-1}\right\} \;, \tag{3.15}$$

where $\mathbf{D}_{\boldsymbol{\theta}}^{Y_l,i}$ is a $K \times K$ matrix with its element in row k_1 and column k_2 being given by

$$\left[\mathbf{D}_{\boldsymbol{\theta}}^{Y_l,i}\right]_{k_1 k_2} = \frac{\partial p_{Y_l}(i\,;\boldsymbol{\theta})}{\partial \theta_{k_1}} \frac{\partial p_{Y_l}(i\,;\boldsymbol{\theta})}{\partial \theta_{k_2}} \;. \tag{3.16}$$

When $Y_1, \ldots, Y_L$ are i.i.d., $p_{Y_l}(i;\boldsymbol{\theta}) = p_Y(i;\boldsymbol{\theta})$ for $l = 1, \ldots, L$ can be used to reduce (3.13) to

$$p_{N}^{\text{opt}}(n) = \arg\min_{p_{N}(\cdot)} \operatorname{trace}\left\{\left(\sum_{i=0}^{M-1} \frac{1}{p_{Y}(i\,;\boldsymbol{\theta})}\,\mathbf{D}_{\boldsymbol{\theta}}^{Y,i}\right)^{-1}\right\} \;. \tag{3.17}$$

Note that in the i.i.d. case, the same noise n is added to all components of $\mathbf{x}$. In other words, a scalar variable can be considered as in (3.17), which results in a significantly simpler optimization problem than that in (3.13).

3.2 Optimal Noise in the Absence of Prior Information

First, the following functions are introduced:

$$H_{\boldsymbol{i}}^{\boldsymbol{\theta}}(\mathbf{n}) \triangleq \mathrm{P}(\tau_{i_1} - n_1 < X_1 \le \tau_{i_1+1} - n_1, \ldots, \tau_{i_L} - n_L < X_L \le \tau_{i_L+1} - n_L) \;, \tag{3.18}$$

$$G_{\boldsymbol{i}}^{\theta_k}(\mathbf{n}) \triangleq \frac{\partial H_{\boldsymbol{i}}^{\boldsymbol{\theta}}(\mathbf{n})}{\partial \theta_k} \;, \quad \text{for } k = 1, \ldots, K. \tag{3.19}$$

Note that (3.18) and (3.19) are the multiple parameter versions of (2.9) and (2.10). Based on the definitions in (3.18) and (3.19), the marginal p.m.f. in (3.1) and its partial derivatives can be expressed as

$$p_{\mathbf{Y}}(\boldsymbol{i}\,;\boldsymbol{\theta}) = \mathrm{E}\{H_{\boldsymbol{i}}^{\boldsymbol{\theta}}(\mathbf{N})\} \ , \qquad \frac{\partial p_{\mathbf{Y}}(\boldsymbol{i}\,;\boldsymbol{\theta})}{\partial \theta_k} = \mathrm{E}\{G_{\boldsymbol{i}}^{\theta_k}(\mathbf{N})\} \ . \tag{3.20}$$

with $0 \leq H_{\boldsymbol{i}}^{\boldsymbol{\theta}}(\mathbf{n}) \leq 1$ and $\sum_{\boldsymbol{i}\in\mathcal{I}} H_{\boldsymbol{i}}^{\boldsymbol{\theta}}(\mathbf{n}) = 1$.

Based on (3.18) and (3.19), the optimization problem in (3.13) can be expressed as

$$p_{\mathbf{N}}^{\mathrm{opt}}(\mathbf{n}) = \arg\min_{p_{\mathbf{N}}(\cdot)} \operatorname{trace}\left\{\left(\sum_{\boldsymbol{i}\in\mathcal{I}} \frac{1}{\mathrm{E}\{H_{\boldsymbol{i}}^{\boldsymbol{\theta}}(\mathbf{N})\}}\, \mathbf{D}_{\boldsymbol{\theta}}^{\mathbf{Y},\boldsymbol{i}}\right)^{-1}\right\} \ , \tag{3.21}$$

where $\mathbf{D}_{\boldsymbol{\theta}}^{Y_l,i}$ in (3.14) is given by

$$\left[\mathbf{D}_{\boldsymbol{\theta}}^{\mathbf{Y},\boldsymbol{i}}\right]_{k_1k_2} = \mathrm{E}\left\{G_{\boldsymbol{i}}^{\theta_{k_1}}(\mathbf{N})\right\}\mathrm{E}\left\{G_{\boldsymbol{i}}^{\theta_{k_2}}(\mathbf{N})\right\} \ . \tag{3.22}$$

Then, the following proposition describes the form of the optimal noise p.d.f.

Proposition 3: *Assume that $H_{\boldsymbol{i}}^{\boldsymbol{\theta}}(\cdot)$ in (3.18) and $G_{\boldsymbol{i}}^{\theta_k}(\cdot)$ in (3.19) are continuous functions and that the additional noise components take finite values specified by $n_l \in [a_l, b_l]$, $l = 1, \ldots, L$, for some finite a_l and b_l. Then, the optimal additive noise p.d.f. in (3.21) can be expressed as*

$$p_{\mathbf{N}}^{\mathrm{opt}}(\mathbf{n}) = \sum_{j=1}^{(M^L-1)(K+1)+1} \lambda_j\, \delta(\mathbf{n} - \mathbf{n}_j) \ , \tag{3.23}$$

where $\lambda_j \geq 0$ and $\sum_{j=1}^{(M^L-1)(K+1)+1} \lambda_j = 1$.

In addition, if the observation vector and the additive noise vector both consist of i.i.d. components, then each component of the optimal additive noise has the same p.d.f. that is in the form of

$$p_N^{\mathrm{opt}}(n) = \sum_{j=1}^{(M-1)(K+1)+1} \nu_j\, \delta(n - n_j) \ , \tag{3.24}$$

where $\nu_j \geq 0$ and $\sum_{j=1}^{(M-1)(K+1)+1} \nu_j = 1$.

Proof: Optimization problems that involve functions of expectations of a number of functions have been investigated in various studies in the literature [5], [6], [31], [32]. Under the conditions in the proposition, it can be shown that the optimal solution of (3.21) can be represented by a randomization of at most $(M^L - 1)(K + 1) + 1$ different noise values as a result of Carathéodory's theorem [33], [34]. Hence, the optimal additive noise PDF can be expressed as in (3.23). The number $(M^L - 1)(K + 1) + 1$ of mass points comes from the facts that there are a total of $K + 1$ different functions for a given value of $\boldsymbol{i} \in \mathcal{I}$; namely, $H_{\boldsymbol{i}}^{\boldsymbol{\theta}}(\cdot), G_{\boldsymbol{i}}^{\theta_1}(\cdot), \ldots, G_{\boldsymbol{i}}^{\theta_K}(\cdot)$, and that there are $M^L - 1$ different functions corresponding to different values of $\boldsymbol{i}$. It should be noted that -1 is used since $\sum_{\boldsymbol{i}\in\mathcal{I}} H_{\boldsymbol{i}}^{\boldsymbol{\theta}}(\mathbf{n}) = 1$ and $G_{\boldsymbol{i}}^{\theta_k}(\mathbf{n}) = \partial H_{\boldsymbol{i}}^{\boldsymbol{\theta}}(\mathbf{n})/\partial\theta_k$.

In the case of i.i.d. observations and i.i.d. components of the additive noise, the problem is separable as shown in (3.17). In that case, there are $(K+1)(M-1)$ different functions, resulting in $(K + 1)(M - 1) + 1$ mass points as a result of Carathéodory's theorem; hence, the expression in (3.24) follows. $\square$

Proposition 3 states that discrete probability distributions with a finite number of mass points solve the optimal additive noise problem under certain conditions. Therefore, it implies that it is not necessary to search over all possible probability distributions in order to obtain the optimal noise, which simplifies the optimization problem significantly. In the next section, this result is used in numerical evaluations to calculate the probability distribution of the optimal additive noise.

3.3 Numerical Results

Consider a scalar observation x in Figure 3.1 with a Gaussian mixture distribution that consists of p components expressed as

$$p_X(x;\theta_1,\theta_2) = \sum_{k=1}^{p} a_k \gamma(x;\mu_k-\theta_1,\theta_2^2)\ , \tag{3.25}$$

where

$$\gamma(x;\theta_1,\theta_2^2) \triangleq \frac{1}{\sqrt{2\pi}\,\theta_2}\exp\left\{-\frac{(x-\theta_1)^2}{2\theta_2^2}\right\}\ . \tag{3.26}$$

In this case, $H_{\boldsymbol{i}}^{\boldsymbol{\theta}}(\mathbf{n})$ in (3.18) is expressed as $H_i^{\boldsymbol{\theta}}(n) = F_X(\tau_{i+1}-n;\theta_1,\theta_2) - F_X(\tau_i-n;\theta_1,\theta_2)$ for $i=0,1,\ldots,M-1$, where the c.d.f. of X for a given value of $\boldsymbol{\theta}=[\theta_1\ \theta_2]^T$ is calculated as

$$F_X(x;\theta_1,\theta_2) = \sum_{k=1}^{p} a_k\, Q\left(\frac{-x+\mu_k-\theta_1}{\theta_2}\right)\ . \tag{3.27}$$

Also, $G_i^{\theta_1}$ and $G_i^{\theta_2}$ can be obtained in a straightforward manner as the derivatives of $H_i^{\boldsymbol{\theta}}$ with respect to θ_1 and θ_2, respectively. In addition, the quantizer has three levels (i.e., $M=3$), which are specified by the thresholds $\tau_1=-8$ and $\tau_2=8$.

First, the optimal additive noise is investigated for $p=3$, $\boldsymbol{a}=[0.4\ 0.4\ 0.2]^T$, $\boldsymbol{\mu}=[-4\ \ -1\ 4]^T$ and $\boldsymbol{\theta}=[0\ 2]^T$. Using these values, the p.d.f. of X given in (3.25) becomes

$$p_X(x;\theta_1=0,\theta_2=2) = 0.4\gamma(x;-4,4)+0.4\gamma(x;-1,4)+0.2\gamma(x;4,4)\,, \tag{3.28}$$

which is depicted in Figure 3.2. According to Proposition 3, the optimal solution is in the form of

$$p_N^{\text{opt}}(n) = \sum_{j=1}^{7}\nu_j\,\delta(n-n_j)\ . \tag{3.29}$$

The optimization problem in (3.21) simplifies based on (3.29), and it can be solved by using global optimization techniques such as particle-swarm optimization (PSO) [35]-[38], genetic algorithms and differential evolution [39]. In this

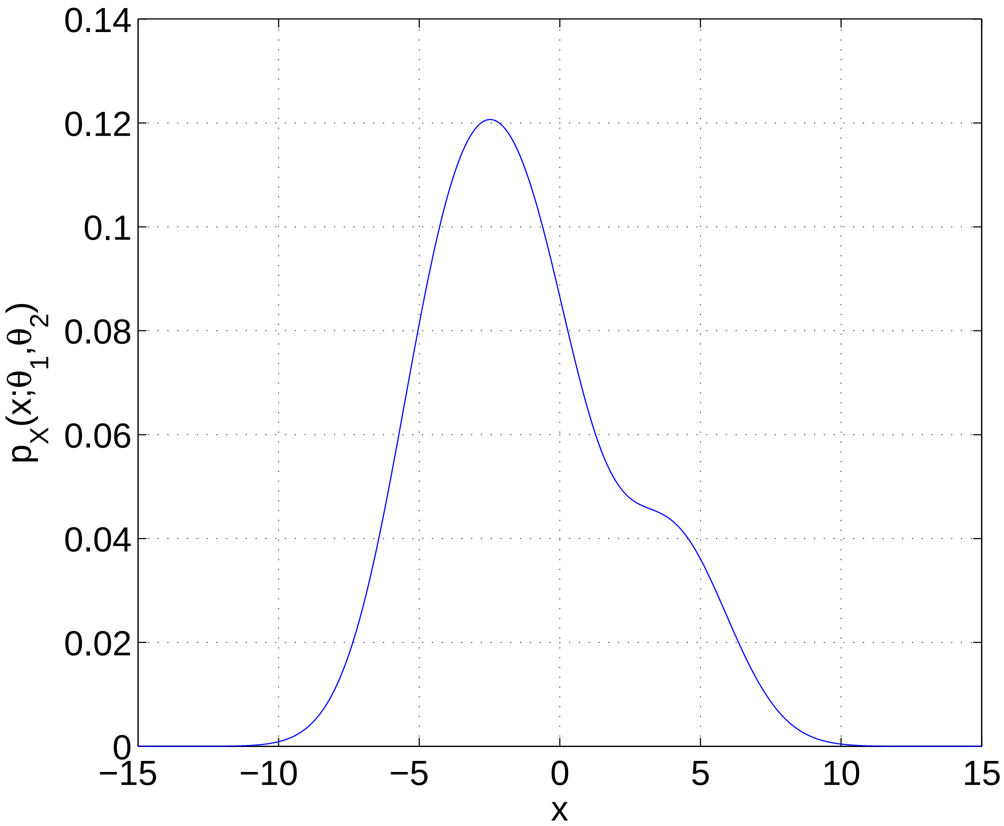

Figure 3.2: The p.d.f. of the Gaussian mixture distributed observation X.

work, the optimal solution is searched by using the PSO algorithm. The PSO algorithm can be described as follows. Consider the minimization of an objective function $f(.)$ over parameter $\mathbf{q}$. A set of parameters are called *particles* and their values $\{\mathbf{q}_i\}_{i=1}^{P}$ express the positions of the particles, where P is called the population size (i.e., the number of particles). First, the particles are generated. Then, iterations are performed, where at each iteration the position of each particle is updated with the addition of the velocity vectors $\boldsymbol{\upsilon}_i$ to the last position of the particle according to the following equations [35]:

$$\boldsymbol{\upsilon}_i^{k+1} = \chi\left(\omega\boldsymbol{\upsilon}_i^k + c_1\rho_{i1}^k\left(\mathbf{p}_i^k - \mathbf{q}_i^k\right) + c_2\rho_{i2}^k\left(\mathbf{p}_g^k - \mathbf{q}_i^k\right)\right) , \tag{3.30}$$

$$\mathbf{q}_i^{k+1} = \mathbf{q}_i^k + \boldsymbol{\upsilon}_i^{k+1} , \tag{3.31}$$

for $i = 1, \ldots, P$, 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 in the range of $[0, 1]$ [36]. In (3.30), $\mathbf{p}_i^k$ denotes the particle position corresponding to the smallest $f(\mathbf{q})$ value until the kth iteration of the ith particle, and $\mathbf{p}_g^k$ represents the position achieved at the global minimum among all the particles until the kth iteration. After a number of iterations, $\mathbf{p}_g^k$ is selected as the optimizer of the optimization problem.

By employing various approaches, such as penalty functions, PSO can be extended to constrained optimization problems [37], [38]. In the penalty function approach, the particle position is set to a large value, if it becomes infeasible.

In order to find the optimal solution for (3.29), the objective function in (3.17) can be rewritten as

$$\begin{bmatrix} \mathbf{n} \\ \boldsymbol{\nu} \end{bmatrix} = \arg \min_{\substack{0 \le \nu_j \le 1 \\ \sum_{j=1}^{7} \nu_j = 1 \\ n_j \in [a,b]}} \operatorname{trace}\left\{\left(\sum_{i=0}^{2} \frac{1}{\sum_{j=1}^{7} \nu_j H_i^{\boldsymbol{\theta}}(n_j)} \mathbf{D}_{\boldsymbol{\theta}}^{Y,i}\right)^{-1}\right\} , \tag{3.32}$$

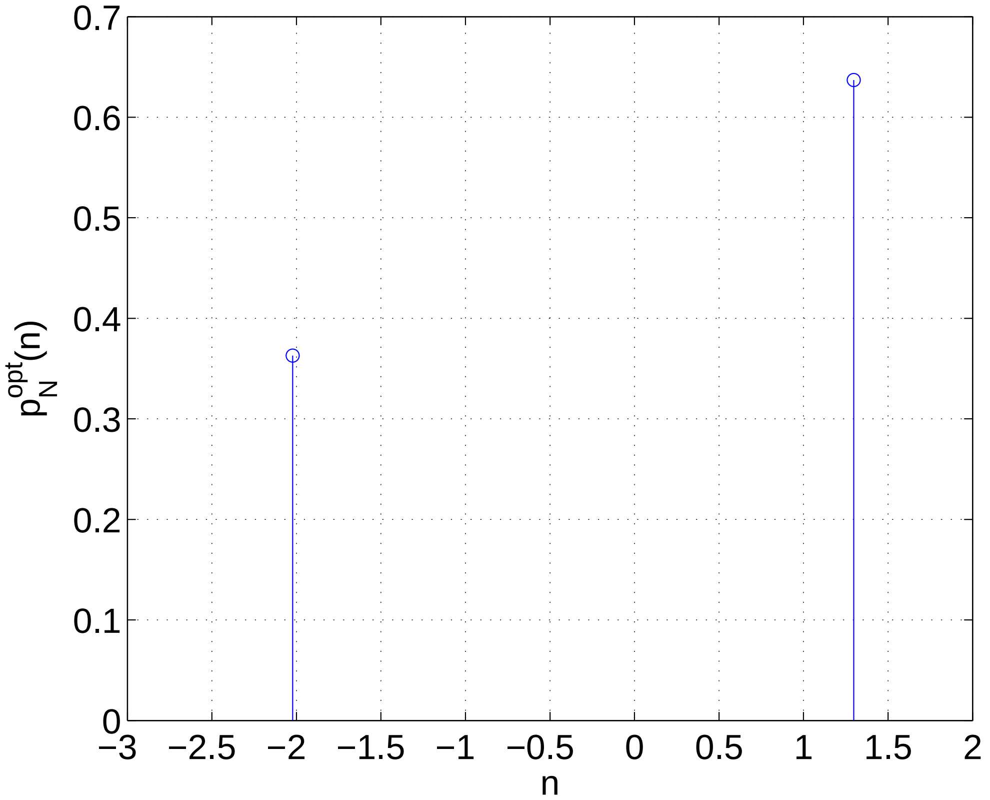

Figure 3.3: The p.d.f. of the optimal additive noise.

where $\left[\mathbf{D}_{\boldsymbol{\theta}}^{Y,i}\right]_{k_1k_2} = (\sum_{j=1}^{7}\nu_j G_i^{\theta_{k_1}}(n_j))(\sum_{j=1}^{7}\nu_j G_i^{\theta_{k_2}}(n_j))$. The mass points of the optimal noise p.d.f., n_j, and their weights ν_j are taken as the position of the particles. Performing 500 iterations for 50 particles, $\chi = 0.72984$, ω linearly decreasing from 1.2 to 0.1 with respect to the number of iteration, $c_1 = 2.05$, $c_2 = 2.05$, $-a = b = 12$, the optimal probability distribution of the additive noise is found as $p_N^{\text{opt}}(n) = 0.6371\,\delta(n-1.2970) + 0.3629\,\delta(n-2.0224)$, where $\nu_3 = \nu_4 = \nu_5 = \nu_6 = \nu_7 = 0$ in (3.29), as depicted in Figure 3.3.

In order to compare the performance of the 2-mass point optimal noise to the one mass point noise, which can also be achieved by shifting the quantizer thresholds, the constant noise performance is also computed as shown in Figure 3.4. The optimal CRLB is achieved with the constant noise $n^{\text{opt}} = 0.3750$. The CRLB values for the optimal noise distribution, the optimal constant noise and no

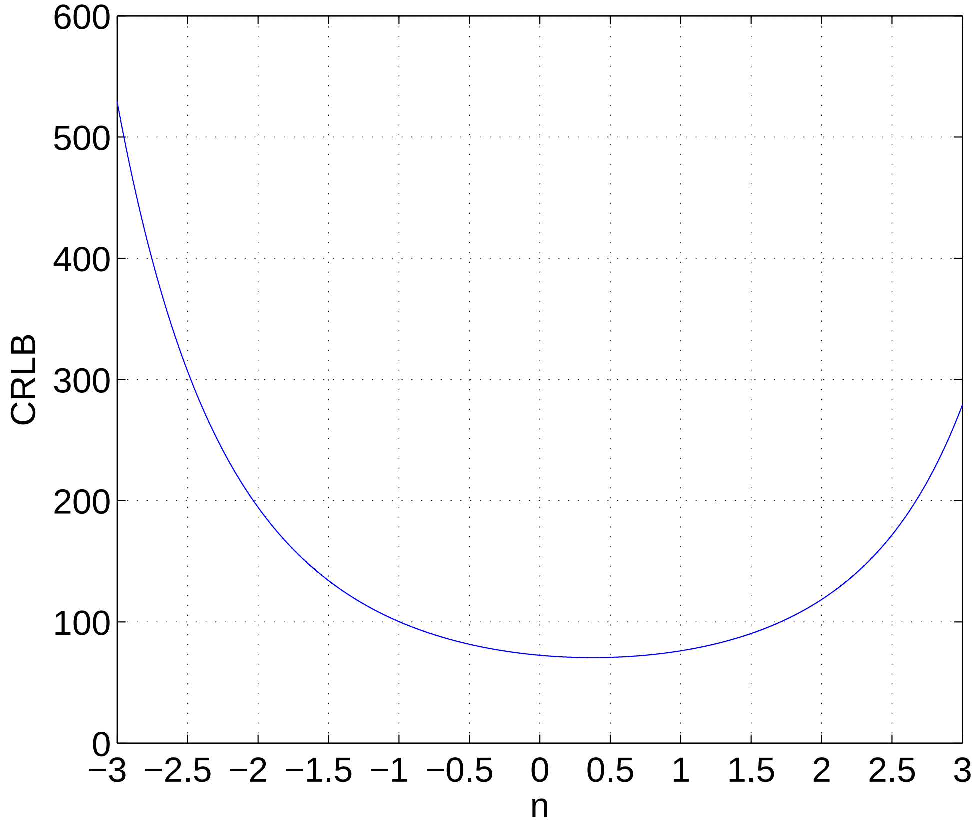

Figure 3.4: CRLB versus additive constant noise n.

Table 3.1: CRLB values for optimal additive random noise, optimal additive constant noise and without additive noise.

	$p_N^{\text{opt}}(n)$	$n = n^{\text{opt}}$	$n = 0$
CRLB	63.6959	70.5280	72.4618

noise cases are given in Table 3.1. In conclusion, the performance of the system in Figure 3.1 is still improvable with a additive constant noise; however, the optimal random noise outperforms the additive constant noise in this example. Moreover, it can be inferred that employing a random noise may result in a better performance enhancement than adding a constant signal or shifting the quantizer thresholds, if our aim is to achieve the best performance enhancement in terms of CRLB at the estimation of multiple parameters.

For the estimation performance evaluation in practical cases, we consider the MSE of the ML estimates of parameters θ_1 and θ_2, which is calculated as

$$\text{MSE} = \text{E}\left\{(\hat{\theta}_1(\mathbf{y}) - \theta_1)^2\right\} + \text{E}\left\{(\hat{\theta}_2(\mathbf{y}) - \theta_2)^2\right\}. \tag{3.33}$$

Since ML estimates are efficient, it is expected that the MSE of the ML estimates asymptotically achieves the CRLB. The ML estimates of the two parameters are calculated as

$$\hat{\boldsymbol{\theta}}_{ML}(\mathbf{y}) = \arg\max_{\theta_1,\theta_2} p_{\mathbf{Y}}(\mathbf{y};\theta_1,\theta_2)\ . \tag{3.34}$$

For i.i.d. case of the observations, the pd.f. of $\mathbf{Y}$ is calculated as $p_{\mathbf{Y}}(\mathbf{y};\theta) = \prod_{l=1}^{L} p_{Y_l}(y_l;\theta)$. For fixed parameters $\boldsymbol{\theta} = [0\ 2]^T$, using the same settings of the example considered for the comparison of the CRLB values, we have performed a series of Monte Carlo trials in order to evaluate the MSE for the ML estimates of parameter $\boldsymbol{\theta}$. As in Section 2.4.3, L realizations of observation Y are generated for the evaluation of the ML estimator performance. The RMSE of the estimates computed for the optimal random noise, the optimal constant noise and no noise cases are compared to their CRLBs in Figure 3.5. As expected, the estimation performance obtained by adding optimal random noise outperforms the optimal

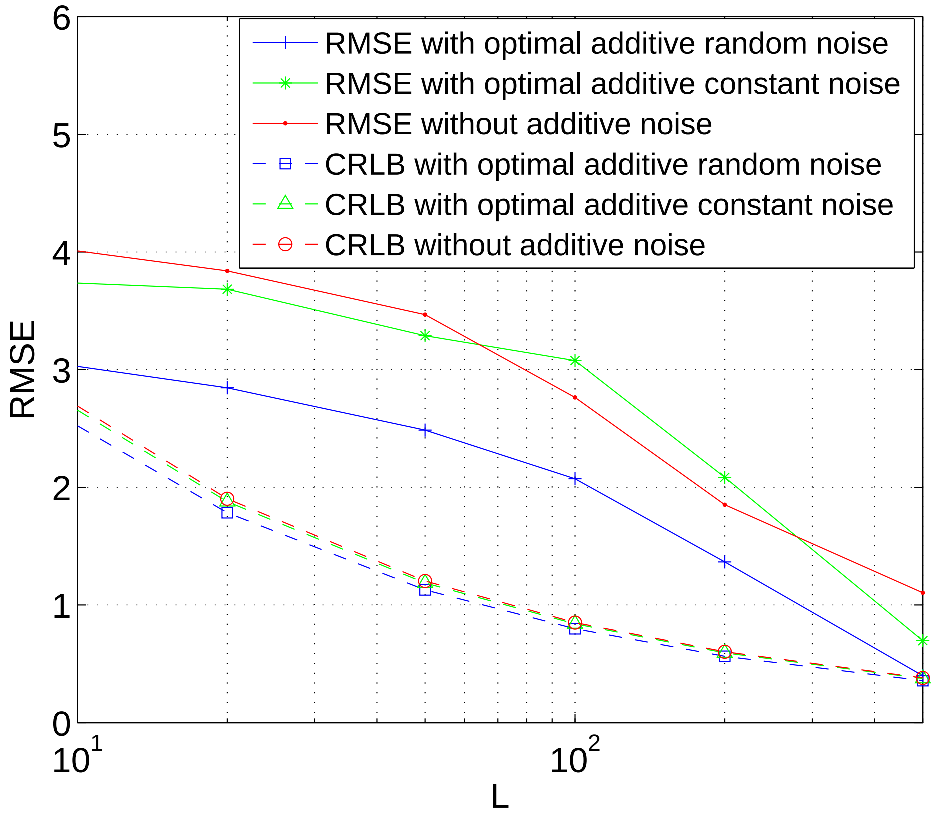

Figure 3.5: RMSE versus CRLB for ML estimates with optimal additive random noise, constant noise and without additive noise. Observations are generated for $\theta_1 = 0$ and $\theta_2 = 2$.

constant noise, which results a slight improvement in comparison to the no noise case.

3.4 Optimal Additive Constant Noise in the Presence of Prior Information

In some practical cases, prior information for the parameters to be estimated can be available. Then, the generic expression for the CRLB on the covariance matrix of the estimators of $\boldsymbol{\theta}$ in (3.2) becomes

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}) \geq \mathbf{J}^{-1} , \tag{3.35}$$

with $\mathbf{J} = \mathbf{J}_{\boldsymbol{\theta}} + \mathbf{J}_{\mathrm{P}}$, where $\mathbf{J}_{\boldsymbol{\theta}}$ denotes the Fisher information obtained from the data expressed as in (3.3) and $\mathbf{J}_{\mathrm{P}}$ from the prior knowledge described as

$$\mathbf{J}_{\mathrm{P}} = \mathrm{E}\left\{ \left(\nabla_{\boldsymbol{\theta}} \log w(\boldsymbol{\theta})\right) \left(\nabla_{\boldsymbol{\theta}} \log w(\boldsymbol{\theta})\right)^{T} \right\} . \tag{3.36}$$

Note that in the Bayesian case, the expectation operator $\mathrm{E}\{\cdot\}$ in (3.3) is over both $\mathbf{y}$ and $\boldsymbol{\theta}$. As a special case, if the components of $\mathbf{X}$ and $\mathbf{N}$ are independent, the FIM can be still rewritten as in (3.5), where (3.6) is valid. For the case in which prior information for the parameters exist, BCRLB imposes a lower limit on the MSE as shown in the following equation:

$$\mathrm{MSE} \geq \sum_{i=1}^{K} \left[\mathbf{J}^{-1}\right]_{ii} = \mathrm{trace}\left\{ (\mathbf{J}_{\boldsymbol{\theta}} + \mathbf{J}_{\mathrm{P}})^{-1} \right\} . \tag{3.37}$$

Since we are looking for the optimal additive constant noise, the p.d.f. of the optimal additive constant noise and optimal noise level can be calculated as

$$p_{\mathbf{N}}^{\mathrm{opt}}(\mathbf{n}) = \delta(\mathbf{n} - \mathbf{n}_{\mathrm{o}}) , \tag{3.38}$$

$$\mathbf{n}_{\mathrm{o}} = \mathrm{trace}\left\{ (\mathbf{J}_{\boldsymbol{\theta}} + \mathbf{J}_{\mathrm{P}})^{-1} \right\} . \tag{3.39}$$

After some manipulation, (3.39) can also be expressed as

$$\mathbf{n}_{\mathrm{o}} = \arg\min_{\mathbf{n}} \mathrm{trace}\left\{ \left(\sum_{\boldsymbol{i} \in \mathcal{I}} \frac{1}{p_{\mathbf{Y}}(\boldsymbol{i}\,;\boldsymbol{\theta})} \mathbf{D}_{\boldsymbol{\theta}}^{\mathbf{Y},\boldsymbol{i}} + \mathbf{J}_{\mathrm{P}} \right)^{-1} \right\} , \tag{3.40}$$

where $\mathcal{I} \triangleq \{0, 1, \ldots, M-1\}^{L}$ and $\mathbf{D}_{\boldsymbol{\theta}}^{\mathbf{Y},\boldsymbol{i}}$ is a $K \times K$ matrix defined as in (3.14). For independent $\mathbf{Y}$ components, the optimal additive noise can be characterized with the p.d.f.

$$\mathbf{n}_{\mathrm{o}} = \arg\min_{\mathbf{n}} \mathrm{trace}\left\{ \left(\sum_{l=1}^{L} \sum_{i=0}^{M-1} \frac{1}{p_{Y_l}(i\,;\boldsymbol{\theta})} \mathbf{D}_{\boldsymbol{\theta}}^{Y_l,i} + \mathbf{J}_{\mathrm{P}} \right)^{-1} \right\} , \tag{3.41}$$

where $\mathbf{D}_{\boldsymbol{\theta}}^{Y_l,i}$ is a $K \times K$ matrix with its element in row k_1 and column k_2 being given by (3.16).

When $Y_1, \ldots, Y_L$ are i.i.d., $p_{Y_l}(i;\boldsymbol{\theta}) = p_Y(i;\boldsymbol{\theta})$ for $l = 1, \ldots, L$ can be used to reduce (3.13) to

$$n_{\mathrm{o}} = \arg\min_{n} \mathrm{trace}\left\{ \left(\sum_{i=0}^{M-1} \frac{1}{p_{Y_l}(i\,;\boldsymbol{\theta})} \mathbf{D}_{\boldsymbol{\theta}}^{Y,i} + \mathbf{J}_{\mathrm{P}} \right)^{-1} \right\} . \tag{3.42}$$

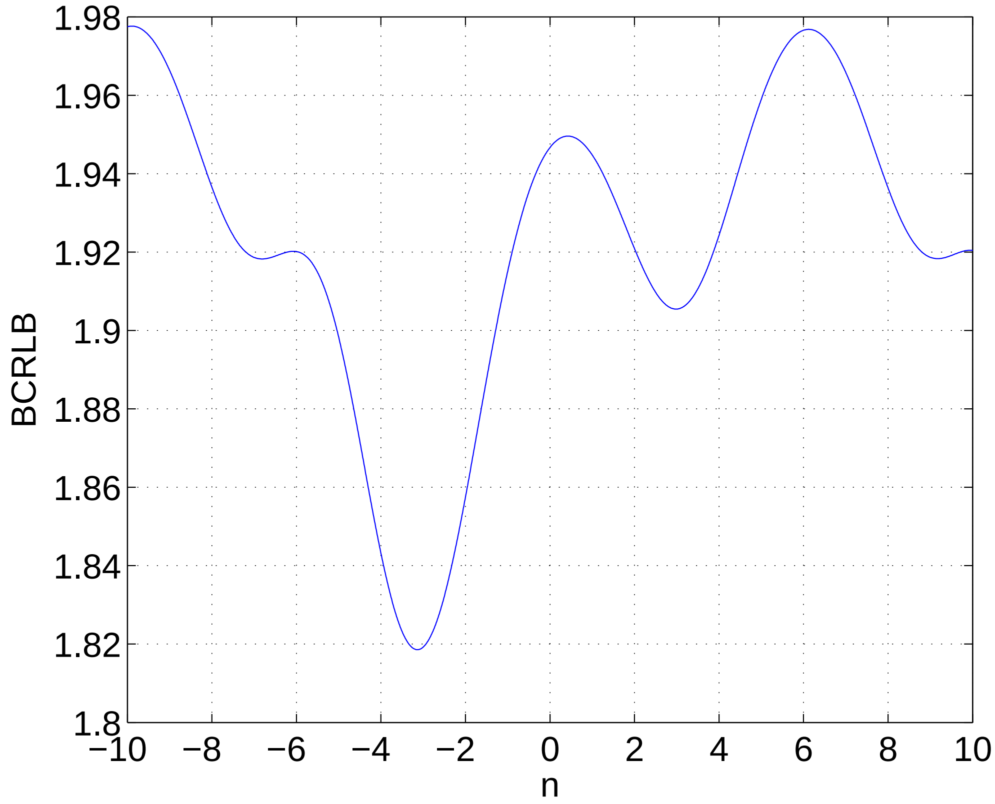

Figure 3.6: BCRLB versus additive constant noise n.

Consider the Bayesian version of the example given in Section 3.3, where the parameters are characterized with the p.d.f.s

$$w_1(\theta_1) = \frac{\exp\left\{-(\theta_1 - \mu_{\theta_1})^2/(2\sigma_{\theta_1}^2)\right\}}{(\sqrt{2\pi}\,\sigma_{\theta_1})} \tag{3.43}$$

and

$$w_2(\theta_2) = \lambda \exp\left\{-\lambda(\theta_2 - \zeta)\right\} \ , \tag{3.44}$$

with $\theta_1 \in \mathbb{R}$, $\theta_2 \in [\zeta, \infty)$, $\mu_{\theta_1} = 0$, $\sigma_{\theta_1} = 1$, $\zeta = 0.5$ and $\lambda = 1$. The lower bound on the MSE of the estimates is depicted in Figure 3.6. The best noise enhanced estimation performance is achieved at $n_{\mathrm{o}} = -3.140$, which shows that the estimation accuracy in the Bayesian case of the multiple parameter estimation problem in Section 3.2 can be improved by adding a constant noise or shifting quantizer thresholds.

3.5 Conclusions

In this chapter, it has been shown that the optimal additive noise, which maximizes the estimation performance (in terms of the CRLB) of multiple parameters based on quantized observations in the absence of the prior information, is a random noise with a discrete probability distribution. In addition, it has been observed that the number of mass points in this discrete probability distribution depends on the numbers of quantization levels, parameters and observations. The estimation performance improvement achieved via optimal additive random noise can outperform the optimal additive constant noise or shifting quantizer thresholds. This result can be also verified in terms of the MSE, since the MSE of the estimators achieves the CRLB asymptotically under certain conditions [30]. In addition, it has been shown that the multiple parameter estimation performance in the presence of prior information can be improved by adding a constant noise or shifting quantizer thresholds.

Chapter 4

CONCLUSIONS AND FUTURE WORK

4.1 Concluding Remarks

In this thesis, noise benefits in parameter estimation problems based on quantized observations have been studied in order to achieve the optimal performance in terms of CRLB and BCRLB. In the single parameter estimation problem, it has been proven that the optimal noise, which should be added to the observation before quantization in order to maximize the estimation accuracy, is a constant signal both in the absence and presence of prior information. In addition, numerical results have shows that constant "noise" always outperforms Gaussian and uniform dither signals. The benefit of this result can also be seen in the MSE performance of the asymptotically efficient estimators such as ML and MAP estimators. It can be concluded that for the optimal estimation performance, no randomization among different noise components is required, which can also be interpreted as shifting quantizer thresholds and is a very useful result for practical implementations.

On the other hand, if multiple parameters are to be estimated, then the optimal additive noise may not be expressed as a constant signal. Instead, the optimal additive noise can be characterized with a p.m.f., whose number of mass points depends on the numbers of quantizer levels, observations, and parameters. In this case, the optimal additive random noise results in better estimation accuracy than optimal additive constant noise. This result can also be seen numerically in the MSE performance of the ML estimator. Furthermore, in the presence of prior information, it has been shown that a additive constant noise can be beneficial to the multiple parameter estimation accuracy.

4.2 Future Work

In Section 3.4, the optimal additive constant noise, which enhances the estimation performance for multiple parameters nested in quantized observations, is investigated in terms of a BCRLB minimization problem. As future work, the theoretical derivation of the optimal additive random noise can be rigorously studied. Also, the effects of the noise enhanced estimation accuracy to the MSE performance of an asymptotically efficient estimator are worth being investigated. Furthermore, the optimal noise in the single and multiple parameter estimation problems in Chapters 2 and 3 may also be formulated in terms of other bounds such as Weiss-Weinstein, Barankin, Abel or Bhattacharya bounds, and estimation performance achieved via optimal noise based on these bounds can be compared.

Bibliography

[1] R. Benzi, A. Sutera, and A. Vulpiani, "The mechanism of stochastic resonance," *J. Phys. A: Math. Gen.*, vol. 14, pp. 453–457, 1981.

[2] L. Gammaitoni, P. Hanggi, P. Jung, and F. Marchesoni, "Stochastic resonance," *Rev. Mod. Phys.*, vol. 70, pp. 223–287, Jan. 1998.

[3] S. Zozor, P.-O. Amblard, and C. Duchene, "On pooling networks and fluctuation in suboptimal detection framework," *Fluctuation and Noise Letters*, vol. 7, no. 1, pp. L39–L60, 2007.

[4] G. O. Balkan and S. Gezici, "CRLB based optimal noise enhanced parameter estimation using quantized observations," *IEEE Sig. Processing Lett.*, vol. 17, pp. 477–480, May 2010.

[5] H. Chen, P. K. Varshney, S. M. Kay, and J. H. Michels, "Theory of the stochastic resonance effect in signal detection: Part I–Fixed detectors," *IEEE Trans. Sig. Processing*, vol. 55, pp. 3172–3184, July 2007.

[6] A. Patel and B. Kosko, "Optimal noise benefits in Neyman-Pearson and inequality-constrained signal detection," *IEEE Trans. Sig. Processing*, vol. 57, pp. 1655–1669, May 2009.

[7] S. M. Kay, J. H. Michels, H. Chen, and P. K. Varshney, "Reducing probability of decision error using stochastic resonance," *IEEE Sig. Processing Lett.*, vol. 13, pp. 695–698, Nov. 2006.

[8] S. Bayram and S. Gezici, "On the improvability and nonimprovability of detection via additional independent noise," *IEEE Sig. Processing Lett.*, vol. 16, pp. 1001–1004, Nov. 2009.

[9] 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.

[10] S. Bayram and S. Gezici, "Effects of additional independent noise in binary composite hypothesis-testing problems," in *Proc. International Conference on Signal Processing and Commun. Systems*, (Omaha, Nebraska), Sep. 2009.

[11] S. Zozor and P.-O. Amblard, "Stochastic resonance in locally optimal detectors," *IEEE Trans. Signal Process.*, vol. 51, pp. 3177–3181, Dec. 2003.

[12] O. Dabeer and A. Karnik, "Signal parameter estimation using 1-bit dithered quantization," *IEEE Trans. Information Theory*, vol. 52, pp. 5389–5405, Dec. 2006.

[13] S. Zozor and P.-O. Amblard, "Noise-aided processing: Revisiting dithering in a sigma-delta quantizer," *IEEE Trans. Signal Process.*, vol. 53, pp. 3202–3210, Aug. 2005.

[14] M. D. McDonnell, D. Abbott, and C. E. M. Pearce, "An analysis of noise enhanced information transmission in an array of comparators," *Microelectronics Journal*, vol. 33, pp. 1079–1089, 2002.

[15] F. Chapeau-Blondeau and D. Rousseau, "Noise-enhanced performance for an optimal Bayesian estimator," *IEEE Trans. Sig. Processing*, vol. 52, pp. 1327–1334, May 2004.

[16] F. Chapeau-Blondeau, "Noise-aided nonlinear Bayesian estimation," *Physical Review E*, vol. 66, pp. 1–3, Sep. 2002.

[17] H. Chen, P. K. Varshney, and J. H. Michels, "Noise enhanced parameter estimation," *IEEE Trans. Sig. Proc.*, vol. 56, pp. 5074–5081, Oct. 2008.

[18] S. Marano, V. Matta, and P. Willett, "Quantizer precision for distributed estimation in a large sensor network," *IEEE Trans. Sig. Processing*, vol. 54, pp. 4073–4078, Oct. 2006.

[19] M. D. McDonnell, "Signal estimation via averaging of coarsely quantised signals," in *Proc. Information, Decision and Control (IDC)*, (Adelaide, Australia), pp. 100–105, Feb. 2007.

[20] H. C. Papadopoulos, G. W. Wornell, and A. V. Oppenheim, "Sequential signal encoding from noisy measurements using quantizers with dynamic bias control," *IEEE Trans. Info. Theory*, pp. 978–1002, 2001.

[21] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks – Part II: Unknown probability density function," *IEEE Trans. Sig. Processing*, vol. 54, pp. 2784–2796, July 2006.

[22] D. Rousseau, F. Duan, and F. Chapeau-Blondeau, "Suprathreshold stochastic resonance and noise-enhanced Fisher information in arrays of threshold devices," *Physical Review E*, vol. 68, pp. 1–10, Sep. 2003.

[23] D. Rousseau, G. V. Anand, and F. Chapeau-Blondeau, "Nonlinear estimation from quantized signals: Quantizer optimization and stochastic resonance," in *International Symposium on Physics in Signal and Image Processing*, (Grenoble, France), pp. 89–92, Jan. 2003.

[24] G. O. Balkan and S. Gezici, "Average fisher information optimization for quantized measurements using additive independent noise," in *IEEE Signal Processing and Communications Applications Conference*, (Diyarbakir, Turkey), Apr. 2010.

[25] H. L. V. Trees, *Detection, Estimation, and Modulation Theory, Part I.* Wiley-Interscience, 2001.

[26] P. Tichavsky, C. H. Muravchik, and A. Nehorai, "Posterior Cramer-Rao bounds for discrete-time nonlinear filtering," *IEEE Trans. Sig. Processing*, vol. 46, pp. 1386–1396, May 1998.

[27] E. Weinstein and A. J. Weiss, "A general class of lower bounds in parameter estimation," *IEEE Trans. Inform. Theory*, vol. 34, pp. 338–342, Mar. 1988.

[28] A. Renaux, P. Forster, P. Larzabal, C. D. Richmond, and A. Nehorai, "A fresh look at the Bayesian bounds of the Weiss-Weinstein family," *IEEE Trans. Sig. Processing*, vol. 56, pp. 5334–5352, Nov. 2008.

[29] R. A. Wannamaker, S. P. Lipshitz, and J. Vanderkooy, "A theory of non-subtractive dither," *IEEE Trans. Sig. Processing*, vol. 48, pp. 449–516, Feb. 2000.

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

[31] S. Bayram, S. Gezici, and H. V. Poor, "Noise enhanced hypothesis-testing in the restricted Bayesian framework," *IEEE Trans. Sig. Processing*, vol. 58, Aug. 2010.

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

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

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

[35] K. E. Parsopoulos and M. N. Vrahatis, *Particle swarm optimization method for constrained optimization problems*, pp. 214–220. IOS Press, 2002. in

Intelligent Technologies–Theory and Applications: New Trends in Intelligent Technologies.

[36] 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.

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

[38] 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.

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