Parameter Estimation from Quantized Observations in Multiplicative Noise Environments

Jiang Zhu[*], Xiaokang Lin[*], Rick S. Blum[†], and Yuantao Gu[*]

Received Sep. 6, 2014; revised Mar. 7, 2015; accepted May 6, 2015;
to appear in *IEEE Transactions on Signal Processing*

Abstract

The problem of distributed parameter estimation from binary quantized observations is studied when the unquantized observations are corrupted by combined multiplicative and additive Gaussian noise. These results are applicable to sensor networks where the sensors observe a parameter in combined additive and nonadditive noise and to a case where dispersed receivers are employed with analog communication over fading channels where the receivers employ binary quantization before noise-free digital communications to a fusion center. We first discuss the case in which all the quantizers use an identical threshold. The parameter identifiability condition is given, and, surprisingly, it is shown that unless the common threshold is chosen properly and the parameter lies in an open interval, the parameter will not generally be identifiable, in contrast to the additive noise case. The best achievable mean square error (MSE) performance is characterized by deriving the corresponding Cramér-Rao Lower Bound (CRLB). A closed-form expression describing the corresponding maximum likelihood (ML) estimator is presented. The stability of the performance of the ML estimators is improved when a nonidentical threshold strategy is utilized to estimate the unknown parameter. The thresholds are designed by maximizing the minimum asymptotic relative efficiency (ARE) between quantized and unquantized ML estimators. Although the ML estimation problem is nonconvex, it is shown that one can relax the optimization to make it convex. The solution to the relaxed problem is used as an initial solution in a gradient algorithm to solve the original problem. Next, the case where both the variances of the additive noise and multiplicative noise are unknown is studied. The corresponding CRLB is obtained, and the ML estimation problem is transformed to a convex optimization, which can be solved efficiently. Finally, numerical simulations are performed to substantiate the theoretical analysis.

Keywords: Parameter estimation, multiplicative noise, quantization, CRLB.

[*]Jiang Zhu, Xiaokang Lin, and Yuantao Gu are with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, CHINA. The corresponding author of this paper is Yuantao Gu (gyt@tsinghua.edu.cn).

[†]Rick S. Blum is with the Electrical and Computer Engineering Department, Lehigh University, Bethlehem, PA 18015, USA.

1 Introduction

The problem of estimating a deterministic parameter from quantized observations is encountered in several applications including target tracking [1], target localization [2], environmental monitoring [3], and radar applications [4] using wireless sensor networks (WSNs). WSNs consist of a large number of spatially distributed nodes which are characterized by several constraints such as power, communication bandwidth and computational capability. In distributed parameter estimation, each node is able to obtain local measurements which are either transmitted to the fusion center (WSNs with a fusion center) or shared among nodes (ad hoc WSNs). In the case of classical multisensor fusion, the data collected by sensors is communicated to the fusion center without any pre-processing. As proposed in much recent work [7, 8], each node will quantize the measurement coarsely to reduce complexity with little performance loss. For bandwidth-constrained parameter estimation in additive white Gaussian noise, it is well-known that the variance of the estimator based on binary measurements can have a variance as small as $\pi/2$ times that of the sample mean estimator based on unquantized measurements [5]. This motivates researchers to propose various strategies to approach the lower bounds [5] [7, 8]. While the majority of investigations considered the well-studied parameter-in-additive-noise estimation problem, other estimation problems, including state estimation, have also received attention with quantized measurements [6].

1.1 Related Work

Signal parameter estimation from quantized data has attracted a lot of attention in past years. Most of these works consider the case in which multiplicative noise is absent. In [5], various strategies are proposed to estimate the unknown deterministic parameter, including using noise dithering, a periodic known threshold or a feedback signal added prior to quantization. Various numerical algorithms have been introduced to estimate the unknown parameter, including the Expectation Maximization (EM) algorithm. When a nonidentical threshold strategy is utilized to estimate the unknown parameter, it is demonstrated that the threshold spacing should be comparable to the standard deviation of the additive noise [7]. In addition, when the noise variance or even the noise probability density function (PDF) is unknown, some strategies are proposed, including nonparametric methods [8]. When the noise is bounded, an isotropic universal decentralized estimation scheme is proposed for a bandwidth constrained ad hoc sensor network, where each sensor node uses the same algorithm independent of noise distributions [9–11]. In [12], signal parameter estimation from 1-bit dithered samples is studied for a general case, and the theoretical properties of the estimators are established. The optimal distribution of additive noise for parameter estimation based on quantized samples is characterized by formulating the estimation problem as a Cramèr Rao lower bound (CRLB) minimization problem [13], which shows that the optimal additive noise is of constant level.

Another related problem is to estimate the unknown parameter in the presence of multiplicative noise. In a linear regression model, the sensing matrix may be subject to errors. One can model the sensing matrix as a random matrix [14,15], where the mean and variance of each element of the matrix are known. The random components of the sensing matrix are a source of multiplicative noise, and a ML estimator is utilized to estimate the unknown parameters. In [16], two approximate ML estimators were proposed to estimate the direction of arrival (DOA) in the presence of multiplicative noise. In [17], a vector parameter is estimated from 1-bit measurements with multiplicative noise. One should note that the threshold of the quantizer used in their model is always zero. However, as we will show later, the threshold provides an additional degree of freedom that may improve the performance of the ML estimator significantly. Different from [17], this paper focuses on improving the performance of scalar parameter estimation with optimization of the identical or nonidentical threshold strategies. Significant improvements are demonstrated in this paper and a robust optimization approach is given and extended for the case where the variances of the multiplicative and additive noise are unknown.

1.2 Contributions

This paper focuses on the problem of deterministic parameter estimation from binary quantized observations of the parameter in combined additive and multiplicative Gaussian noise. The model we employ can model multiplicative noise in a number of practical situations, see [18,19] and the references therein. Thus, one can view our results as extending previous results involving quantization of a parameter in additive noise to cases with multiplicative and additive noise. It is very interesting that this relatively minor change leads to cases with an unidentifiable parameter unless the quantizer is designed properly and the parameter is known to lie in an open interval, which does not occur in the additive noise case. Since similar things may also occur in other estimation problems, these results are especially informative and expand the accumulated knowledge concerning signal processing with quantized data in an important way. As one very concrete example of a practical application, the problem we study could also model analog transmission of a measurement θ over a fading channel where the transmission is received at a bunch of spatially separated receivers placed to get independent fading. As shown in Fig. 1, the receivers make hard binary decisions that are sent to a fusion center to develop an estimate $\hat{\theta}(\mathbf{y})$. We assume the binary transmissions are sent error free. This will simplify things and will be a justified approximation if the transmissions are sent slowly enough that we can use strong codes and obtain nearly error free communications. If the noise in the original analog transmissions is just thermal noise which is the typical assumption in communication systems, it will produce an i.i.d. sequence. This is the most commonly used simple model in communication systems and it makes sense to consider this model. Even if the noise is not just thermal, the model is reasonable if the receivers are reasonably separated in space.

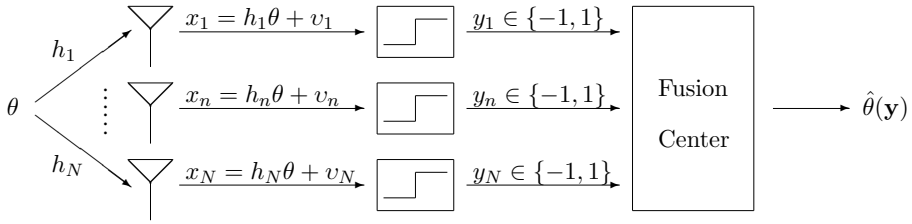

Figure 1: Block diagram for estimating a noisy measurement θ at the fusion center when the measurement θ is transmitted using analog communication over parallel fading channels with gains $h_1, \ldots, h_N$. These transmissions are received by spatially separated receivers which make hard decisions that are sent to the fusion center (error free) to develop an estimate $\hat{\theta}(\mathbf{y})$.

The best achievable MSE performance of an unbiased estimator is analyzed in the scenario of identical thresholds. By minimizing the corresponding CRLB, it is shown that the optimal threshold is not equal to the true value of θ which is the case with additive noise as shown in previous research [5] [7]. Then, the ML estimator is utilized to estimate the unknown parameter. It is shown that the ML estimator is asymptotically efficient when the threshold satisfies certain conditions. Secondly, a nonidentical threshold strategy is used to improve the performance of the ML estimation. The corresponding CRLB is derived, and it is shown that the multiplicative noise may improve the performance of the ML estimator under rather mild condition in this nonidentical scenario. The nonidentical thresholds are designed by maximizing the minimum asymptotic relative efficiency (ARE). To solve the corresponding ML estimation problem, we re-parameterize and relax the original problem to find a good initial point. Thirdly, the case in which some nuisance parameters are present is studied. The nuisance parameters describe the distributions of the additive noise and the multiplicative noise. The corresponding CRLB is obtained, and the ML estimation problem is transformed to a convex optimization, which can be solved efficiently.

Note that some approximations have been adopted in order to simplify the analysis. In Subsection 3.3, we have used the Chernoff bound to find the approximately optimal threshold given known θ. Besides, for the nonidentical thresholds strategy, we also use the Chernoff bound to simplify the analysis in Subsection 4.1. These approximations are useful to understand this problem, and the impact of the approximations are discussed in Section 6.

Notation

The $N \times 1$ vector of ones is $\mathbf{1}_N$, and the $N \times N$ identity matrix is $\mathbf{I}_N$. For an unknown deterministic parameter θ, θ_0 denotes its true value. For a random vector $\mathbf{y}$, $p(\mathbf{y}; \theta)$ denotes the probability density function (PDF) of $\mathbf{y}$ parameterized by θ, and $\mathrm{E}_{\mathbf{y}}[\cdot]$ denotes the expectation taken with respect to $\mathbf{y}$. For $\mathbf{x} = [x_1, \cdots, x_n]^{\mathrm{T}}$ and a continuous and differentiable

function $f : \mathbb{R}^n \to \mathbb{R}$, $\nabla_{\mathbf{x}} f$ and $\nabla^2_{\mathbf{x}} f$ denotes its gradient and Hessian. For a vector function $\mathbf{g} : S \to \mathbb{R}^r$ defined on a set S in $\mathbb{R}^s$, $\partial \mathbf{g}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ denotes its Jacobian matrix $[\partial g_i(\theta)/\partial \theta_j]_{r\times s}$. For any appropriate matrix $\mathbf{A}$, $\mathrm{tr}(\mathbf{A})$ denotes its trace, $\mathbf{A} \succeq \mathbf{0}$ ($\mathbf{A} \succ \mathbf{0}$) means that $\mathbf{A}$ is positive semidefinite (positive definite), and $\mathbf{A} \succeq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B} \succeq \mathbf{0}$.

The rest of this paper is organized as follows. In Section II, the parameter estimation problem with unquantized measurements is studied. The theoretical CRLB and the closed-form expression of the ML estimator based on binary measurements is derived, and the theoretical performance limits are analyzed in Section III. Section IV deals with estimation using the nonidentical threshold strategy. The thresholds are selected via a robust optimization procedure to ensure the performance of the ML estimator is stable. Numerical algorithms are proposed to find the ML estimator. In Section V, the underlying parameter is estimated when some nuisance parameters describing the distributions of the additive noise and the multiplicative are present. In Section VI, numerical results are presented. Finally we conclude the paper in Section VII.

2 Preliminaries

In this section, the signal acquisition model is introduced first. Then the ML estimation problem based on unquantized measurements is studied. The corresponding CRLB is derived, and the ML estimator is numerically found.

As for the model previously described in Figure 1, for the nth receiver, its received signal can be described as

$$x_n = h_n\theta + \upsilon_n, \quad n = 1, \cdots, N, \tag{1}$$

where θ is the underlying parameter to be estimated, υ_n is the additive noise at the nth receiver satisfying $\upsilon_n \sim \mathcal{N}(0, \sigma_\upsilon^2)$, and the noise is assumed independent and identically distributed across the receivers. h_n is a flat fading channel coefficient, x_n is the local measurement at receiver n, and N denotes the number of receivers. In our setting, the channel coefficient h_n is modeled as an *i.i.d.* Gaussian random variable which satisfies $h_n \sim \mathcal{N}(1, \sigma_e^2)$ and is independent of υ_n. Consequently, model (1) is written as

$$x_n = (1 + e_n)\theta + \upsilon_n, \tag{2}$$

where e_n is a random Gaussian perturbation satisfying $e_n \sim \mathcal{N}(0, \sigma_e^2)$. Note that the channel coefficient h_n can be described by two terms. The first term corresponds to the line-of-sight path, and the second term corresponds to the scattered path. As a result, $1/\sigma_e^2$ is the ratio of the power of the line of sight component to that of the remaining nonspecular multipath [20]. Another perspective is to view the random component e_n as a perturbation added to the deterministic component 1. Consequently, we may view σ_e^2 as the strength of a perturbation in the following text.

To estimate the unknown parameter θ, the random component e_n is treated as a noise. Therefore, we may define the equivalent noise z_n as $z_n = e_n\theta + \upsilon_n$, which can be viewed as the sum of the additive noise and the multiplicative noise. Obviously, the multiplicative noise satisfies $e_n\theta \sim \mathcal{N}(0, \sigma_e^2\theta^2)$. Since the additive noise υ_n is independent of e_n, it can be shown that $z_n \sim \mathcal{N}(0, \sigma_z^2)$, where σ_z^2 is defined as

$$\sigma_z^2 = \theta^2\sigma_e^2 + \sigma_\upsilon^2. \tag{3}$$

To find the ML estimator of θ, ones needs to solve the following optimization problem

$$\tilde{\theta}_{\mathrm{ML}} = \underset{\theta}{\operatorname{argmin}} \ -\log p(\mathbf{x};\theta), \tag{4}$$

where $p(\mathbf{x};\theta)$ denotes the likelihood function based on unquantized measurements $\{x_n\}_{n=1}^N$. Differentiating the above log-likelihood function (4) and setting it equal to zero yields

$$N\sigma_e^4\theta^3 + \sigma_e^2 s\theta^2 + (N\sigma_\upsilon^2 - \sigma_e^2 t + N\sigma_e^2\sigma_\upsilon^2)\theta - \sigma_\upsilon^2 s = 0, \tag{5}$$

where t and s are constants defined as $t \triangleq \sum_{n=1}^N x_n^2$ and $s \triangleq \sum_{n=1}^N x_n$, respectively. The three roots of (5) can be calculated efficiently by several numerical methods. Then the root that minimizes the negative log-likelihood function is chosen as the ML estimator $\tilde{\theta}_{\mathrm{ML}}$.

The CRLB provides a lower bound on the mean square error (MSE) of any unbiased estimator. Therefore, it is meaningful to discuss the CRLB. The Fisher information (FI) is used to find bounds on unbiased estimators, and the CRLB follows by taking the inverse of the FI. To compute the corresponding FI, we utilize the result of the CRLB for the general Gaussian case [21], and finally obtain

$$\mathrm{CRLB}(\mathbf{x};\theta) = \frac{1}{N}\frac{\left(\theta^2\sigma_e^2 + \sigma_\upsilon^2\right)^2}{\sigma_\upsilon^2 + (1+2\sigma_e^2)\sigma_e^2\theta^2}. \tag{6}$$

Notice that $\mathrm{CRLB}(\mathbf{x};\theta) \leq \frac{\theta^2\sigma_e^2+\sigma_\upsilon^2}{N}$, where the righthand side term of the inequality denotes the variance of the unbiased sample mean estimator. The above result is expected, because the ML estimator is asymptotically optimal in the sense of unbiasedness.

3 ML Estimation Based on Identical Thresholds

In this section, the case where each sensor or receiver quantizes the signal with the same threshold τ is investigated. It is assumed that both the strength of perturbation σ_e^2 and the variance of the additive noise σ_υ^2 are known. First, the log-likelihood function is computed. Then the statistical identifiability of the estimation problem is analyzed, and a sufficient condition is provided to make the problem well-defined. Finally, the CRLB is derived, and the ML estimator is utilized to estimate the unknown parameter.

It is assumed that θ lies in $(-\Delta, \Delta)$, where Δ is known. This assumption is practical since quantizers have the inherent dynamic range limitations [5].

With a fixed threshold τ, the fusion center obtains the 1-bit measurements $\{y_n\}_{n=1}^N$ as

$$y_n = \text{sign}(x_n - \tau). \tag{7}$$

From (7), it can be seen that y_n follows a Bernoulli distribution. Given the observations $\{y_n\}_{n=1}^N$, let them be partitioned into the index sets $\mathcal{I}_+$ and $\mathcal{I}_-$, where $\mathcal{I}_+$ and $\mathcal{I}_-$ are defined as $\mathcal{I}_+ = \{n|y_n = 1, n = 1, \cdots, N\}$ and $\mathcal{I}_- = \{n|y_n = -1, n = 1, \cdots, N\}$, respectively. According to (2), the likelihood function can be calculated as

$$\begin{aligned}\Pr(\mathbf{y};\theta) &= \prod_{n\in\mathcal{I}_+} \Pr(\theta + z_n - \tau > 0) \prod_{n\in\mathcal{I}_-} \Pr(\theta + z_n - \tau \leq 0) \\ &= \prod_{n=1}^{N} \Phi\left(y_n \frac{\theta - \tau}{\sigma_z}\right),\end{aligned}$$

where $\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} \mathrm{e}^{-\frac{x^2}{2}} \mathrm{d}x$ is the cumulative distribution function of the standard Gaussian distribution and σ_z is defined in (3). Therefore, the corresponding log-likelihood function is

$$l(\mathbf{y};\theta) = \sum_{n=1}^{N} \log \Phi\left(y_n \frac{\theta - \tau}{\sigma_z}\right). \tag{8}$$

3.1 Statistical identifiability

In this subsection, the statistical identifiability of the model (7) is investigated, and the range of the threshold to make the model identifiable is determined. The concept of identifiability is related to the underlying parameter being uniquely determined from the distribution of the measurements. Mathematically, this means that the mapping $\theta \longrightarrow l(\mathbf{y};\theta)$ is injective for all $\mathbf{y}$, and the underlying parameter can be estimated accurately given that the number of measurements tends to be infinity [22, 23].

By introducing a new variable

$$\omega(\theta, \tau) \triangleq (\theta - \tau)/\sqrt{\theta^2\sigma_e^2 + \sigma_v^2} \tag{9}$$

the log-likelihood function (8) is more easily analyzed. The function $\log \Phi(\cdot)$ is monotone increasing. Therefore, to ensure that the estimation problem is identifiable, the following condition should be satisfied

$$\theta_1 \neq \theta_2 \Longrightarrow \omega(\theta_1, \tau) \neq \omega(\theta_2, \tau). \tag{10}$$

We first discuss three extreme cases. When the variance of the equivalent noise σ_z^2 is zero, we can only determine the interval to which θ belongs according to (2) and (7). When the strength of the multiplicative noise σ_e^2 is zero, condition (10) is always satisfied. Thus any threshold will make the estimation problem statistically identifiable. When the variance

of the additive noise σ_v^2 is zero, we find that there does not exist a threshold such that the estimation problem is identifiable.

Let us go back to the general case, but assume that both σ_v^2 and σ_e^2 are nonzero unless stated otherwise in the following text. A sufficient condition for (10) to be satisfied is that $\omega(\theta, \tau)$ is a strictly monotone function of θ in the interval $(-\Delta, \Delta)$. By straightforward calculation, it can be shown that the partial derivative of $\omega(\theta, \tau)$ with respect to θ is

$$\eta(\theta, \tau) \triangleq \partial\omega(\theta, \tau)/\partial\theta = (\sigma_v^2 + \theta\tau\sigma_e^2)/(\theta^2\sigma_e^2 + \sigma_v^2)^{\frac{3}{2}}. \tag{11}$$

Because $\theta \in (-\Delta, \Delta)$ and $\eta(0, \tau) > 0$, $\omega(\theta, \tau)$ is monotone increasing with θ if and only if

$$-\sigma_v^2/(\Delta\sigma_e^2) \leq \tau \leq \sigma_v^2/(\Delta\sigma_e^2). \tag{12}$$

As a consequence, (12) is a sufficient condition to make the estimation problem identifiable.

It can be checked that (12) is not applicable when the variance of the additive noise σ_v^2 is zero. When the perturbation does not exist, i.e., σ_e is zero, the estimation problem is identifiable for any choice of thresholds, which is consistent with (12). If σ_v^2 is zero, $\omega(\theta, \tau)$ is not differentiable at $\theta = 0$. For any nonzero σ_e and σ_v, selecting the threshold τ as zero can also make the estimation problem identifiable. The smaller Δ is, the larger the range of the threshold which will make estimation problem identifiable from (12). As smaller Δ reduces our uncertainty, this seems reasonable. At the end of Section III, we show a case where the parameter becomes unidentifiable when (12) is not satisfied.

3.2 Cramér-Rao Lower Bound

In this subsection, the performance with the approach in (7) is characterized using the CRLB on the MSE of unbiased estimators. A closed-form expression for the CRLB is presented in the following proposition. To simplify the expression, we let $p_{n_0}(\cdot)$ denote a standard normal PDF.

Proposition 1 *Consider the estimation of θ from the observations in (7) with identical thresholds at each sensor. The MSE $\mathrm{mse}(\hat{\theta}) = \mathrm{E}_{\mathbf{y}}[|\theta - \hat{\theta}|^2]$ for any unbiased estimator $\hat{\theta}$ must satisfy*

$$\mathrm{mse}(\hat{\theta}) \geq \frac{1}{N}\frac{1}{q(\theta, \tau)\eta^2(\theta, \tau)} \triangleq B(\theta, \tau), \tag{13}$$

where $q(\theta, \tau)$ is

$$q(\theta, \tau) = \frac{p_{n_0}^2\left(\frac{\theta-\tau}{\sigma_z}\right)}{\Phi\left(\frac{\theta-\tau}{\sigma_z}\right)\Phi\left(-\frac{\theta-\tau}{\sigma_z}\right)}, \tag{14}$$

$\eta(\theta, \tau)$ is defined as (11), and $B(\theta, \tau)$ is defined as the CRLB with the threshold being τ.

PROOF It can be checked that the regularity condition $\mathrm{E}_{\mathbf{y}}\left[\nabla_\theta l(\mathbf{y};\theta)\right]=0$ holds. Besides, the FI is computed through $J(\theta)=-\mathrm{E}_{\mathbf{y}}\left[\nabla_\theta^2 l(\mathbf{y};\theta)\right]$, where $\nabla_\theta^2 l(\mathbf{y};\theta)$ is given by

$$\begin{aligned}\nabla_\theta^2 l(\mathbf{y};\theta)=&-\sum_{i=1}^{N}\frac{1}{\Phi^2\left(y_i\frac{\theta-\tau_i}{\sigma_z}\right)}\eta^2(\theta,\tau_i)p_{n_0}^2\left(\omega(\theta,\tau_i)\right)\\&+\sum_{i=1}^{N_i}\frac{\sqrt{2\pi}y_i}{\Phi\left(y_i\frac{\theta-\tau_i}{\sigma_z}\right)}\nabla_\theta\left(\eta^2(\theta,\tau_i)p_{n_0}^2\left(\omega(\theta,\tau_i)\right)\right).\end{aligned}\tag{15}$$

It follows by straightforward calculation that the expectation of the second term of (15) is zero, and the FI is

$$J(\theta)=Nq(\theta,\tau)\eta^2(\theta,\tau).\tag{16}$$

Consequently, the CRLB (13) can be established. ∎

We now establish some links with results in prior related work. If the threshold τ is zero or the strength of perturbation σ_e^2 is zero, CRLB (13) is consistent with those results in [17] and [7], respectively.

Remark 1 *From the previous subsection, we know that if the parameter θ lies in $(-\Delta,\Delta)$ and the threshold is chosen according to (12), then the parameter θ is identifiable. Meanwhile, under the same conditions, it is interesting to find that the CRLB is always finite from (11) and (13).*

3.3 The approximately optimal threshold τ_{app}

In this subsection, the near optimal threshold τ_{opt} is determined in terms of minimizing the corresponding CRLB $B(\theta,\tau)$. However, it is difficult to directly analyze (13). Fortunately, by using the Chernoff bound for the cummulative distribution function (CDF) [24]

$$\Phi\left(-\frac{\theta-\tau}{\sigma_z}\right)\Phi\left(\frac{\theta-\tau}{\sigma_z}\right)\le\frac{1}{4}\mathrm{e}^{-\frac{(\theta-\tau)^2}{2\sigma_z^2}},\tag{17}$$

one can find a tight upper bound $U(\theta,\tau)$ for $B(\theta,\tau)$ by substituting (17) in (13) to obtain

$$B(\theta,\tau)\le\frac{\pi}{2N}\frac{\sigma_z^6}{\left(\sigma_v^2+\theta\tau\sigma_e^2\right)^2}\mathrm{e}^{\frac{(\theta-\tau)^2}{2\sigma_z^2}}\triangleq U(\theta,\tau).\tag{18}$$

We will find the approximately optimal threshold τ_{app} by minimizing the upper bound $U(\theta,\tau)$. Taking the natural logarithm of the above upper bound $U(\theta,\tau)$ and dropping term independent of τ, one has

$$\log U(\theta,\tau)\sim(\theta-\tau)^2/(2\sigma_z^2)-2\log|\sigma_v^2+\theta\tau\sigma_e^2|.$$

Setting the first derivative of the above right-hand side to zero, one can obtain the approximately optimal threshold as

$$\tau_{\text{opt}} \approx \tau_{\text{app}} = \left(1-\alpha+\sqrt{(1+\alpha)^2+8\sigma_e^2\,(1+\alpha)}\right)\frac{\theta}{2}, \tag{19}$$

where α is defined as $\alpha \triangleq \sigma_v^2/(\theta^2\sigma_e^2)$, which can be regarded as the ratio of the variance of the additive noise to that of the multiplicative noise. In fact, when $\theta \approx 0$ or $\alpha \gg 1$, by using a first order Talyor approximation, equation (19) reduces to

$$\begin{aligned}\tau_{\text{opt}} &\approx \left(1-\alpha+\alpha\sqrt{1+2(1+4\sigma_e^2)/\alpha+(1+8\sigma_e^2)/\alpha^2}\right)\theta \\ &\approx (1+2\sigma_e^2)\theta.\end{aligned}$$

This result demonstrates that the optimal threshold is not always equal to the true value of θ with the existence of multiplicative noise.

Although the optimal threshold depends on the true value of θ, it is still meaningful in helping us to estimate the parameter θ with adaptive approaches. Suppose that feedback can be utilized to improve the estimation. Then the fusion center can obtain an estimation $\hat{\theta}_t$ iteratively. According to (19), the fusion center can adjust the optimal threshold τ_{t+1} based on this estimation $\hat{\theta}_t$. As a consequence, the accuracy of new estimation can improve significantly, as revealed by [5, 25].

3.4 ML Estimation

Now a closed-form expression of the ML estimate of θ is presented. Meanwhile, the asymptotic efficiency of the ML estimator is established.

Finding the constrained ML estimate of θ amounts to the following optimization problem

$$\hat{\theta}_{\text{ML}} = \underset{\theta\in[-\Delta,\Delta]}{\operatorname{argmin}} \; -\sum_{n=1}^{N}\log\Phi\left(y_n\frac{\theta-\tau}{\sqrt{\theta^2\sigma_e^2+\sigma_v^2}}\right). \tag{20}$$

The ML estimation of θ can be obtained by first optimizing the variable $\omega(\theta,\tau)$ in (9). Suppose that $\{y_n\}_{n=1}^N$ has k ones, the ML estimation of $\omega(\theta,\tau)$ can be simplified to be

$$\hat{\omega}_\tau = \underset{\omega(\theta,\tau)}{\operatorname{argmin}} \; -k\log\Phi\left(\omega(\theta,\tau)\right)-(N-k)\log\Phi\left(-\omega(\theta,\tau)\right).$$

By setting the partial derivative of the above log-likelihood function with respect to $\omega(\theta,\tau)$ to zero, one can show that

$$\hat{\omega}_\tau = \Phi^{-1}\left(k/N\right), \tag{21}$$

where $\Phi^{-1}(\cdot)$ denotes the inverse function of $\Phi(\cdot)$. Considering the restrictions that $\theta \in (-\Delta,\Delta)$ and the threshold τ satisfies (12), $\omega(\theta,\tau)$ should satisfy

$$\Omega_1 \triangleq \omega(-\Delta,\tau) \le \omega(\theta,\tau) \le \omega(\Delta,\tau) \triangleq \Omega_2. \tag{22}$$

Combining (21) and (22), the constrained ML estimation of $\omega(\theta,\tau)$ is $\hat{\omega}_{\mathrm{ML}} = \mathcal{I}_{\Omega}(\hat{\omega}_{\tau})$, where $\mathcal{I}_{\Omega}(\cdot)$ is the following piecewise-linear limiter function

$$\mathcal{I}_{\Omega}(x) = \begin{cases} x, & \text{if} \quad \Omega_1 \leq x \leq \Omega_2 \\ \Omega_1, & \text{if} \quad x < \Omega_1 \\ \Omega_2, & \text{otherwise.} \end{cases}$$

According to (9) and by defining $\xi^2 = 1 + \frac{\sigma_v^2}{\sigma_e^2\tau^2}(1 - \hat{\omega}_{\mathrm{ML}}^2\sigma_e^2)$, where $\xi > 0$, $\theta_{1,2} = \frac{1 \pm \hat{\omega}_{\mathrm{ML}}\sigma_e\xi}{1-\hat{\omega}_{\mathrm{ML}}^2\sigma_e^2}\tau$, $\theta_3 = \frac{\sigma_v}{\sqrt{1-\hat{\omega}_{\mathrm{ML}}^2\sigma_e^2}}\hat{\omega}_{\mathrm{ML}}$, and $\theta_4 = \left(1 - \frac{\sigma_v^2}{\sigma_e^2\tau^2}\right)\frac{\tau}{2}$, one can obtain the ML estimation of θ shown in Table 1, where $\times$ denotes cases that never occur. Note that θ_3 is consistent with the result in the prior work [17]. The proof is presented in the Appendix in Section 8.1.

Table 1: Considering the restriction $\theta \in (-\Delta, \Delta)$, the ML estimator $\hat{\theta}_{\mathrm{ML}}$ with identical threshold τ.

$\tau \backslash \hat{\omega}_{\mathrm{ML}}$	$(-\infty, -1/\sigma_e)$	$\pm 1/\sigma_e$	$(-1/\sigma_e, 1/\sigma_e)$	$(1/\sigma_e, \infty)$
$\tau < 0$	$\times$	θ_4	θ_2	θ_2
$\tau = 0$	$\times$	$\times$	θ_3	$\times$
$\tau > 0$	θ_1	θ_4	θ_1	$\times$

In Figure 2, it is shown that the MSE performance of the ML estimator $\hat{\theta}_{\mathrm{ML}}$ (20) based on binary measurements does not deviate far away from $B(\theta,\tau)$ (13), the CRLB using unquantized measurements, provided an appropriate threshold is employed. The parameters are selected as: $\theta = 1$, $\sigma_e^2 = 0.5$, $\sigma_v^2 = 0.5$, $\tau = 1.5$, and $\Delta = 3$. Numerical computation shows that $\tau_{\mathrm{opt}} = 1.7$, thus the selected threshold is near the optimal threshold. It is demonstrated that the ML estimator $\tilde{\theta}_{\mathrm{ML}}$ (4) quickly approaches its corresponding CRLB, $\mathrm{CRLB}(\mathbf{x};\theta)$ (6), while there exists some gap between the CRLB $B(\theta,\tau)$ (13) and the MSE of the ML estimator $\hat{\theta}_{\mathrm{ML}}$ if $N < 50$. Therefore, optimization based on CRLB minimization is reliable if the number of observations is large. Besides, it is shown that with an appropriate threshold, the degradation of the performance of the ML estimator $\hat{\theta}_{\mathrm{ML}}$ based on binary quantized measurements is indeed small compared to the CRLB of unquantized measurements.

Now the asymptotic property of the ML estimator $\hat{\theta}_{\mathrm{ML}}$ is summarized in the following theorem.

Theorem 1 *Suppose that the unknown parameter θ lies in $(-\Delta, \Delta)$, and the threshold τ is chosen according to (12). Then the ML estimator $\hat{\theta}_{\mathrm{ML}}$ is asymptotically efficient, i.e., $\lim_{N\to\infty} N\mathrm{E}_{\mathbf{y}}[(\hat{\theta}_{\mathrm{ML}} - \theta)^2] = B_0(\theta,\tau)$, where $B_0(\theta,\tau)$ denotes the CRLB for estimating θ based on any one sample y_n and is equal to $NB(\theta,\tau)$.*

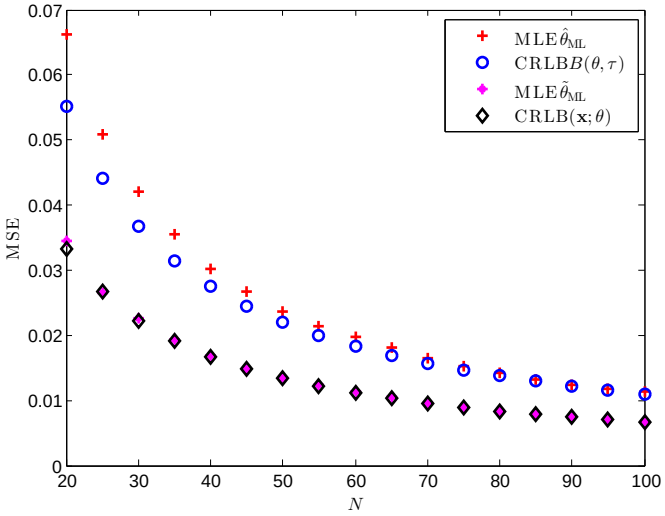

Figure 2: The MSE performance of the two ML estimators $\tilde{\theta}_{\mathrm{ML}}$ (4) and $\hat{\theta}_{\mathrm{ML}}$ (20) compared with their corresponding CRLBs CRLB$(\mathbf{x};\theta)$ (6) and $B(\theta,\tau)$ (13). The threshold is chosen as $\tau = 1.5$.

PROOF Since the derivative of the log-likelihood function exists, the parameter θ is identifiable, and the FI is nonzero for $\theta \in (-\Delta, \Delta)$ given τ satisfying (12), the ML estimator is asymptotically efficient based on the Theorem 7.3.1 in [26]. ∎

The Theorem 7.3.1 from [26] says that ML estimators are always asymptotically efficient under some necessary conditions. From [26], parameter identifiability is a necessary condition for the ML estimator to be asymptotically efficient, so it is important how the thresholds are selected. It is reasonable that parameter identifiability is needed. For example, consider the parameters set as follows: $\Delta = 2$, $\sigma_v^2 = 0.5$, $\sigma_e^2 = 1$ and the true value of θ is $\theta_0 = 1$. If the threshold is chosen as $\tau = -2$, then $l(\mathbf{y};\theta)|_{\theta=1} = l(\mathbf{y};\theta)|_{\theta=-0.2}$ for arbitrary observations. In such a case, a maximum likelihood estimator has no way to know if $\theta_0 = 1$ or $\theta_0 = -0.2$ and having larger N will not help.

4 ML Estimation Based on Nonidentical Thresholds

The variance of the ML estimator $\hat{\theta}_{\mathrm{ML}}$ given in Table 1 can be close to the variance of the ML estimator $\tilde{\theta}_{\mathrm{ML}}$ based on unquantized measurements when the threshold is close to the optimal threshold given the actual parameter θ. When the dynamic range of θ is large compared to the standard deviation of the noise, the identical threshold strategy can not ensure that the performance of the quantized ML estimator is close to that of the unquantized ML estimator for every θ [7]. Therefore, a nonidentical thresholds strategy is proposed to ensure that there will always exist a threshold close to the optimal threshold of the true parameter.

In this section, a nonidentical threshold strategy is utilized to estimate the unknown parameter θ. Firstly, the mathematical model is described and the corresponding CRLB is established. Secondly, the nonidentical thresholds are designed via robust optimization. Finally, numerical algorithms are proposed to find a global minimum of the negative log-likelihood function with high probability.

Suppose that the sensors are partitioned into N_g groups. For the ith group, there are N_i sensors which utilize the identical threshold τ_i. Let $\rho_i = N_i/N$ denote the corresponding fraction of sensors, where N is the total number of sensors. Clearly, one has $1 = \sum_{i=1}^{N_g} \rho_i$. Let $x_{i,j}$ be the original measurement of the jth sensor in the ith group, and $y_{i,j}$ be the corresponding 1-bit quantized measurement. It is also assumed that the fusion center knows the thresholds utilized by any sensor.

Similar to (8), the corresponding log-likelihood function is obtained as

$$l(\mathbf{y};\theta) = \sum_{i=1}^{N_g}\sum_{j=1}^{N_i} \log \Phi\left(y_{i,j}\frac{\theta-\tau_i}{\sqrt{\sigma_e^2\theta^2+\sigma_v^2}}\right), \tag{23}$$

where $\mathbf{y}$ denotes the binary measurements of all groups.

In the case of nonidentical thresholds, the CRLB is stated in the following proposition.

Proposition 2 *Consider the estimation of θ with both σ_e^2 and σ_v^2 known. The nonidentical thresholds strategy is utilized to estimate θ. Consequently, the FI is*

$$J_{\mathrm{k}}(\theta) = N\sum_{i=1}^{N_g}\left(q(\theta,\tau_i)\eta^2(\theta,\tau_i)\rho_i\right), \tag{24}$$

and the MSE of any unbiased estimator based on binary measurements $\mathbf{y}$ must satisfy

$$\mathrm{mse}(\hat{\theta}) \geq \frac{1}{\left(\sum\limits_{i=1}^{N_g} q(\theta,\tau_i)\eta^2(\theta,\tau_i)\rho_i\right)N} \triangleq \mathrm{CRLB}(\mathbf{y};\theta) \tag{25}$$

where $q(\theta,\tau)$ and $\eta(\theta,\tau)$ are defined as (14) and (11), respectively.

PROOF The FI of the identical threshold strategy is (16) given in Proposition 1. Similarly, we can obtain the FI (24) in this nonidentical scenario, and the CRLB is (25). ∎

Given $\sigma_e^2 = 0$, we can find that (25) is consistent with the result in [7]. From Proposition 2, it can be seen that the nonidentical thresholds strategy can overcome some problems of the identical threshold approach. In particular, in Subsection 3.1, it is shown that the parameter is unidentifiable unless the threshold is chosen properly. In the nonidentical thresholds scenario, the CRLB (25) is always finite, which overcomes the limitations of the identical threshold approach. This occurs because the estimation problem is always identifiable with multiply thresholds. In fact, in Subsection 5.1, a closed form expression for the ML estimator is shown to exist with only two different thresholds even when the two

nuisance parameters σ_e^2 and σ_v^2 are unknown. This result demonstrates that the multiple threshold strategy provides some significant advantages for estimation. In the following subsection, we will show that it is possible to select the system parameters $(\boldsymbol{\tau}, \boldsymbol{\rho})$ such that the CRLB (25) is robust in the dynamic range of the parameter being estimated, overcoming this limitation of the ML estimator in the identical threshold scenario.

It is well-known that suitable additive noise can improve the MSE performance of the ML estimator in nonlinear systems [27–32]. In fact, it can be shown that increasing the variances of the multiplicative noise may improve the estimation accuracy in this nonidentical threshold scenario. An example is used to validate this statement. Suppose that the variance of the additive noise σ_v^2 is zero, and θ is nonzero. Assume each threshold τ_i is nonzero and satisfies $\tau_i - \theta \neq 0$. Since the nonidentical threshold strategy is utilized, one can obtain a consistent ML estimator even when the two nuisance parameters σ_v^2 and σ_e^2 are unknown, as shown in the Subsection 5.1. Therefore the sign of θ can be determined provided that the number of measurements N is large enough. Without loss of generality, we may restrict that $\theta \in (0, \Delta)$. The derivative of the log-likelihood function (23) exists for $\theta \in (0, \Delta)$, and the FI is nonzero according to (24). Thus the CRLB (25) can be used to characterize the performance of the ML estimator in the case of $\sigma_v^2 = 0$ [33].

From Proposition 2, one can conclude that the FI corresponding to the ith group is $J_{\mathrm{k}}(\theta, \tau_i) = Nq(\theta, \tau_i)\eta^2(\theta, \tau_i)\rho_i$. It can be shown that if the strength of the multiplicative noise tends to either zero or positive infinity, $J_{\mathrm{k}}(\theta, \tau_i)$ tends to zero, which implies that there exists no finite unbiased estimator for θ [34]. Except for the two cases, $J_{\mathrm{k}}(\theta, \tau_i)$ is nonzero. Thus an increase in the variance of the multiplicative noise is helpful when starting from zero variance multiplicative noise. In general, this phenomenon reveals that multiplicative noise has two opposing effects on estimation. On the one hand, multiplicative noise provides dynamic thresholds for the quantization, which is beneficial to the estimation. On the other hand, multiplicative noise increases the variance of the estimation. Therefore, there exists an optimal strength of multiplicative noise that balances the two opposing effects.

4.1 Threshold selection via robust optimization

In this subsection, the problem of designing the system parameters $(\boldsymbol{\tau}, \boldsymbol{\rho})$ is analyzed. The concept of asymptotic relative efficiency (ARE) is first introduced. Then the ARE is adopted as a criterion to design the system parameters.

The ARE is utilized to measure the relative performance between two estimators [37,38]. In our scenario, the ARE is defined as the ratio of the CRLBs obtained by $\mathbf{x}$ to that of $\mathbf{y}$, which is

$$\gamma(\theta) \triangleq \frac{\mathrm{CRLB}(\mathbf{x}; \theta)}{\mathrm{CRLB}(\mathbf{y}; \theta)}, \tag{26}$$

where $\mathrm{CRLB}(\mathbf{x}; \theta)$ is the CRLB based on unquantized measurements given by (6). Since the ML estimator asymptotically achieves the CRLB, the ARE represents the relative asymp-

totic performance of the two ML estimators based on quantized and unquantized measurements. It is obvious that the performance of the unquantized ML estimator is better than that of the quantized ML estimator, the ARE $\gamma(\theta)$ is less than 1 for every value of θ. It can also be seen that in the asymptotic sense, if the number of quantized measurements is $1/\gamma$ times that of unquantized measurements then the MSE performances of both estimators are the same [38].

Since the ARE depends on θ, which is unknown in practice. We consider the design of the parameters $(\boldsymbol{\tau}, \boldsymbol{\rho})$ that maximizes the minimum ARE over $\theta \in (-\Delta, \Delta)$. In other words, we wish to design the parameters such that

$$\begin{aligned} (\boldsymbol{\tau}^*, \boldsymbol{\rho}^*) = & \underset{\boldsymbol{\tau}, \boldsymbol{\rho}}{\operatorname{argmax}} \inf_{\theta \in (-\Delta, \Delta)} && \gamma(\theta, \boldsymbol{\tau}, \boldsymbol{\rho}), && \text{(27a)} \\ & \text{subject to} && \boldsymbol{\rho}^{\mathrm{T}} \mathbf{1} = 1, && \text{(27b)} \\ & && \boldsymbol{\rho} \succeq \mathbf{0}, && \text{(27c)} \end{aligned}$$

where $\gamma(\theta, \boldsymbol{\tau}, \boldsymbol{\rho})$ is produced by inserting (6) and (25) in (26)

$$\begin{aligned} \gamma(\theta, \boldsymbol{\tau}, \boldsymbol{\rho}) &= \frac{1}{\kappa(\theta)} \sum_{i=1}^{N_g} q(\theta, \tau_i)(\sigma_v^2 + \theta \tau_i \sigma_e^2)^2 \rho_i \triangleq \sum_{i=1}^{N_g} r(\theta, \tau_i) \rho_i \\ &\geq \frac{2}{\pi \kappa(\theta)} \sum_{i=1}^{N_g} \mathrm{e}^{-\frac{(\theta - \tau_i)^2}{2\sigma_z^2}} (\sigma_v^2 + \theta \tau_i \sigma_e^2)^2 \rho_i \triangleq \sum_{i=1}^{N_g} s(\theta, \tau_i) \rho_i, \end{aligned} \tag{28}$$

where $\kappa(\theta) = \left[\sigma_v^2 + (1 + 2\sigma_e^2)\sigma_e^2 \theta^2\right] \sigma_z^2$ and the inequality follows by the Chernoff bound (17). Note that (27) implies that the number of thresholds N_g is given. Meanwhile, it can be seen that problem (27) is a linear programming problem of $\boldsymbol{\rho}$ if $\boldsymbol{\tau}$ is fixed. If $\boldsymbol{\rho}$ is fixed, (27) is either convex or concave of $\boldsymbol{\tau}$. Therefore, it is difficult to find the optimal solutions of (27).

Note that thresholds can be optimized based on either CRLB or ARE criterion. One may use CRLB$(\mathbf{y}, \theta)$ as an optimization criterion, as [13] [35] [36] does. ARE can also be adopted as a criterion to design the quantizers as in [5] [29] concerning estimation in quantizer systems. The reasons that we use the ARE criterion instead of CRLB are as follows: (1) ARE reflects the quantization effects compared to the unquantized systems. Given the ARE, we can determine the least number of observations using binary quantizers whose MSE is comparable to that of the unquantized systems in the asymptotic sense. (2) Optimization via ARE also provides stability of estimation, as revealed by [5] [29].

The optimization problem (27) can be relaxed to obtain

$$(\rho^*(\tau)) = \underset{\rho(\tau) \in \mathcal{X}(\infty)}{\operatorname{argmax}} \inf_{\theta \in (-\Delta, \Delta)} \mathrm{E}_\tau \left[r(\theta, \tau) \right], \tag{29}$$

where τ is regarded as a random variable whose PDF is $\rho(\tau)$, $\mathcal{X}(o)$ denotes the following set

$$\mathcal{X}(o) = \left\{ \rho(\tau) \middle| \int_{-o}^{o} \rho(\tau) \mathrm{d}\tau = 1, \rho(\tau) \geq 0, \forall -o \leq \tau \leq o \right\}. \tag{30}$$

The problem (29) can be interpreted as follows: We should find the optimal PDF $\rho^*(\tau)$ which corresponds to the worst-case ARE for $\theta \in (-\Delta, \Delta)$. In the following analysis, the unknown PDF $\rho(\tau)$ is allowed to describe a random variable with a discrete part so that $\rho(\tau)$ may include some delta functions. It can be seen that the value of the optimized objective function of (29) should be larger than (27), because any feasible point in (27) is feasible for (29), but a feasible solution for (29) may not be feasible for (27).

It is difficult to explore the solution $\rho^*(\tau)$ of (29). Instead, we will analyze $\rho_l^*(\tau)$ in (29) with $r(\theta, \tau)$ replaced by $s(\theta, \tau)$, where $s(\theta, \tau)$ is a lower bound of $r(\theta, \tau)$ given by (28). For the optimization problem (29) and its lower bound version, we have the following result.

Proposition 3 *There exists an optimal $\rho^*(\tau)$ and $\rho_l^*(\tau)$ both of which are symmetric around the origin, i.e., $\rho^*(\tau)$ and $\rho_l^*(\tau)$ satisfy $\rho^*(\tau) = \rho^*(-\tau)$ and $\rho_l^*(\tau) = \rho_l^*(-\tau)$, respectively. Further, the support of $\rho_l^*(\tau)$ is bounded, i.e., $\rho_l^*(\tau) = 0$, $\forall |\tau| \geq \tau_{\max}$, where $\tau_{\max}$ is a finite constant given shortly in (45).*

Proof The proof is presented in the Appendix in Subsection 8.3. ∎

In the following analysis, the support of $\rho(\tau)$ is required to be $[-\tau_{\max}, \tau_{\max}]$. This requirement can be interpreted as follows: First, the threshold should be chosen such that it is near the optimal threshold given θ. Since the dynamic range of θ is bounded, the whole set of optimal thresholds is bounded. The thresholds we select should not be far away from the optimal thresholds set. This intuition may demonstrate that $\rho(\tau)$ is bounded. Secondly, the Chernoff bound is a good approximation. The bounded nature of $\rho_l(\tau)$ may be applied to $\rho(\tau)$, and the support of $\rho(\tau)$ may be approximated by $[-\tau_{\max}, \tau_{\max}]$. Thirdly, numerical simulation demonstrates that the ARE is robust under this requirement.

Assume that the support of $\rho(\tau)$ is $[-\tau_{\max}, \tau_{\max}]$, the epigraph form of problem (29) is

$$\underset{\rho(\tau)\in\mathcal{X}(\tau_{\max}),t}{\text{minimize}} \quad t, \tag{31a}$$

$$\text{subject to} \quad -t - \int_{-\tau_{\max}}^{\tau_{\max}} r(\theta,\tau)\rho(\tau)\mathrm{d}\tau \leq 0, \quad \forall \theta \in (-\Delta, \Delta), \tag{31b}$$

which is linear in t and the functional $\rho(\tau)$. Thus the above optimization problem is an infinite dimensional linear programming problem, and any local optimum corresponds to the global optimum.

For the original optimization problem (31), strong duality holds [39]. We derive the dual in the Appendix in Subsection 8.4. Now we give an interpretation of the dual for the original optimization problem (27). For the dual problem (46), $\lambda(\theta)$ can be viewed as a PDF of θ. It can be shown that the dual problem (46) is equivalent to

$$\min_{\lambda(\theta)\in\mathcal{X}(\Delta)} \max_{\tau\in[-\tau_{\max},\tau_{\max}]} \mathrm{E}_\theta \left[r(\theta, \tau) \right]. \tag{32}$$

The inner maximization means that for a given PDF $\lambda(\theta)$, we select the optimal threshold in the sense of Bayesian ARE. The outer minimization means that the worst-case PDF is selected. Thus the dual problem means that we should find a worst-case PDF $\lambda^*(\theta)$ for θ and the corresponding optimal threshold. This gives another interpretation in the original optimization problem.

Next we will discuss the numerical algorithms to solve the primal (31) and its dual (32). First, we will determine $\tau_{\max}$ (45). Then we solve a finite dimensional optimization problem by discretization. We consider the case corresponding to $\theta > 0$ and $\tau > 0$. Note that $s(\theta, -\tau)$ seems negligible compared to $s(\theta, -\tau)$ with θ and τ being large in (44). In this setting, $\tau_{\max}(\theta)$ can be approximated by $\tau_{\text{app}}(\theta)$ (19), where $\tau_{\text{app}}(\theta)$ is a critical point of $s(\theta, \tau)$. Besides, the approximately optimal threshold $\tau_{\text{app}}(\theta)$ is monotonously increasing with respect to $|\theta|$, which can be checked by showing $\partial\tau_{\text{app}}/(\theta)\partial\theta > 0, \forall\theta > 0$. Thus the largest approximately optimal threshold $\tau_{\text{app,max}}$ is obtained at $\theta = \Delta$. In the numerical implementation, the following approximation $\tau_{\max} \approx \tau_{\text{app,max}}$ is utilized, and numerical simulation shows that the ARE is robust for $\theta \in (-\Delta, \Delta)$.

We solve the primal (31) and its dual (32) numerically. The method we adopt here is to represent θ as belonging to the interval $(-\Delta, \Delta)$ and taking on the value of a finite number of discrete points. Let θ take on values in $[\theta_1, \cdots, \theta_M]$, where $\theta_1 = -\Delta$, $\theta_M = \Delta$, and M controls the discretization step. Meanwhile, the thresholds belonging to the interval $[-\tau_{\max}, \tau_{\max}]$ are also discretized as $[\tau_1, \cdots, \tau_T]$, where $\tau_1 = -\tau_{\max}$ and $\tau_T = \tau_{\max}$. Let the discrete points be equally spaced. Then we may use CVX to solve the discretized version of problems (31) and (32) [41]. Once we obtain the optimal threshold probability mass function (PMF) of the discretized version of problem (31), we can construct a symmetric threshold PMF $\rho^*_{\text{ds}}(\tau)$. To reduce the number of nonidentical thresholds, we discard the thresholds whose fraction is small, and combine the nearby thresholds to form one threshold. In addition, the sum of the fraction of the nonidentical thresholds are normalized to be unity. Finally, the modified $\rho^*_{\text{ds}}(\tau)$ and the thresholds which corresponds to $\rho^*_{\text{ds}}(\tau) \neq 0$ are adopted as an approximate solution of (27). Similarly, $\lambda^*_{\text{ds}}(\theta)$ is obtained as an approximate solution of worst-case parameter PMF $\lambda^*(\theta)$.

4.2 ML estimation

Finding the ML estimate of θ is equivalent to the following optimization problem

$$\hat{\theta}^{\text{k}}_{\text{ML}} = \underset{\theta}{\text{argmin}} \; -\sum_{i=1}^{N_g}\sum_{j=1}^{N_i} \log\Phi\left(y_{i,j}\frac{\theta - \tau_i}{\sqrt{\sigma_e^2\theta^2 + \sigma_v^2}}\right). \tag{33}$$

Since the ML estimator $\hat{\theta}^{\text{k}}_{\text{ML}}$ can not be found in closed-form, numerical algorithms are proposed to find the optimal solution of problem (33). The modified Newton-Raphson method is utilized by replacing the Hessian with the negative of the FI to yield

$$\theta_{t+1} = \theta_t + J_{\text{k}}^{-1}(\theta_t)\nabla_\theta l(\mathbf{y}; \theta_t),$$

where $J_{\mathrm{k}}(\cdot)$ is given in (24) and t denotes the iteration index. However, the numerical algorithms are not guaranteed to converge to the global optimum, because the ML estimation problem (33) may have multiple local optimum points. As a consequence, a good initial point is key to ensuring the numerical algorithms converge to the global optimum with high probability. A strategy is proposed to find a good initial point. The idea is to relax the original non-convex optimization problem to a convex optimization problem, then the corresponding optimal point is chosen as an initial point for a gradient search on the original non-convex objective function. Numerical simulations show that the initial point is typically near the global optimum. Besides, the initial point is also a consistent ML estimator as will be shown later. Thus it is expected that this method can help us find the global optimum.

Although the original optimization problem (33) is non-convex, it can be relaxed to become a convex optimization problem. By introducing two new variables

$$u = \theta / \sqrt{\sigma_e^2 \theta^2 + \sigma_v^2} \tag{34a}$$

and

$$v = 1 / \sqrt{\sigma_e^2 \theta^2 + \sigma_v^2}, \tag{34b}$$

the above problem (33) can be equivalently transformed to become

$$\underset{u \in \mathbb{R}, v > 0}{\text{minimize}} \quad - \sum_{i=1}^{N_g} \sum_{j=1}^{N_i} \log \Phi \left(y_{i,j} \left(u - v \tau_i \right) \right), \tag{35a}$$

$$\text{subject to} \quad u^2 \sigma_e^2 + v^2 \sigma_v^2 = 1. \tag{35b}$$

This objective function is convex with respect to u and v. However, (35b) describes an ellipse which implies a non-convex constraint. Thus problem (35) is still non-convex [40]. The method we propose is to discard the non-convex constraint (35b), then the original ML estimation problem is relaxed to become a convex optimization problem. The objective function of (35a) is strictly convex [17], thus the local optimum is the unique global optimum. The gradient and the Hessian of (35a) are derived in the Appendix in Section 8.2. Then the gradient and Hessian are provided to the MATLAB *fminunc* function to find the global optimum (u^*, v^*) [44] which will provide an initial point for the following ML optimization. Consequently, the initial point is chosen as

$$\theta_{\mathrm{init}} = u^* / v^*. \tag{36}$$

Next, the ML estimator $\hat{\theta}_{\mathrm{ML}}^{\mathrm{k}}$ is numerically found by solving the original problem (33) with the above initial point θ_{init}.

5 ML Estimation Based on Nonidentical Thresholds with Parameters Unknown

In this section, it is assumed that both the variance of the additive noise σ_v^2 and the strength of the perturbation σ_e^2 are unknown. In this scenario, it can be seen that estimation of θ is impossible with the identical threshold strategy, because for an estimated $\hat{\omega}_\tau$ (21), there exists an infinite number of solutions $(\hat{\theta}, \hat{\sigma}_e, \hat{\sigma}_v)$ according to (9). Thus multiple thresholds are utilized to estimate the unknown parameter. We assume that the scenario is the same as in Section 4. The corresponding CRLB is derived and analyzed, and a numerical algorithm is proposed to solve the ML estimation problem.

The closed-form expression of the corresponding CRLB is presented in the following proposition.

Proposition 4 *Consider the estimation of θ with both σ_e^2 and σ_v^2 unknown. The nonidentical threshold strategy is utilized to estimate θ. Consequently, the MSE of any unbiased estimator must satisfy*

$$\begin{aligned} \mathrm{mse}(\hat{\theta}) &\geq \mathrm{CRLB}_{\mathrm{u}}(\mathbf{y};\theta) \\ &= \frac{1}{N}\frac{\sum_{i=1}^{N_g}\left(q(\theta,\tau_i)\rho_i\theta^2 - 2q(\theta,\tau_i)\tau_i\rho_i\theta + q(\theta,\tau_i)\tau_i^2\rho_i\right)}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} q(\theta,\tau_i)q(\theta,\tau_j)(\tau_j-\tau_i)\tau_j\rho_i\rho_j}\sigma_z^2, \end{aligned} \tag{37}$$

where $q(\theta,\tau)$ is defined as (14).

Proof The proof is presented in the Appendix in Section 8.5. ∎

Note that if $\sigma_e^2 = 0$, $N_g = 2$ and $\rho_1 = \rho_2 = 1/2$, the CRLB (37) is consistent with the result in [8]. From Proposition 4, it can be seen that the CRLB (37) is related to the variance of the equivalent noise σ_z. This is because both the multiplicative noise and the additive noise are Gaussian distributed. The strength of the perturbation and the variance of the additive noise are unknown. As a consequence, the estimation accuracy is related to the variance of the equivalent noise, which is also consistent with common sense.

5.1 ML Estimation

Finding the ML estimator $\hat{\theta}_{\mathrm{ML}}^{\mathrm{u}}$ is equivalent to the following optimization problem

$$\hat{\theta}_{\mathrm{ML}}^{\mathrm{u}} = \underset{\theta,\sigma_v,\sigma_e}{\operatorname{argmin}} \; -\sum_{i=1}^{N_g}\sum_{j=1}^{N_i}\log\Phi\left(y_{i,j}\frac{\theta-\tau_i}{\sqrt{\sigma_e^2\theta^2+\sigma_v^2}}\right). \tag{38}$$

According to (34), we reformulate another optimization problem

$$\underset{u\in\mathbb{R},v>0,\sigma_v,\sigma_e}{\text{minimize}} \quad -\sum_{i=1}^{N_g}\sum_{j=1}^{N_i}\log\Phi\left(y_{i,j}\left(u-v\tau_i\right)\right), \tag{39a}$$

$$\text{subject to} \quad u^2\sigma_e^2 + v^2\sigma_v^2 = 1. \tag{39b}$$

For the problem (38) and (39), we have the following result.

Proposition 5 *The optimization problem (38) is equivalent to (39).*

PROOF Suppose that the optimal value of (38) is f^*, and the optimal solution is $(\theta_f^*, \sigma_{e,f}^*, \sigma_{v,f}^*)$. Meanwhile, the optimal value of (39) is g^*, and the optimal solution is $(u^*, v^*, \sigma_{e,g}^*, \sigma_{v,g}^*)$. Once we find the optimal solution $(\theta_f^*, \sigma_{e,f}^*, \sigma_{v,f}^*)$, we can construct $(u_1, v_1, \sigma_{e,f}^*, \sigma_{v,f}^*)$ given by (34), and the corresponding optimal value is given by g_1. Thus we have $f^* = g_1 \geq g^*$. Similarly, we have $f^* \leq g^*$ by construction. In summary, $f^* = g^*$ and the two optimization problems are equivalent. ∎

Note that problem (39) is similar to (35). However, in problem (39), the nuisance parameters σ_e and σ_v are optimization variables. Since the optimization variables σ_e and σ_v appear only in the constraint (39b), we may solve the optimization problem (39) to obtain $(u_{\text{opt}}, v_{\text{opt}})$ after discarding the non-convex constraint (39b). Consequently, the optimal nuisance parameters $\sigma_{e,\text{opt}}$ and $\sigma_{v,\text{opt}}$ need to satisfy $u_{\text{opt}}^2\sigma_{e,\text{opt}}^2 + v_{\text{opt}}^2\sigma_{v,\text{opt}}^2 = 1$. The ML estimator $\hat{\theta}_{\text{ML}}^{\text{u}}$ is

$$\hat{\theta}_{\text{ML}}^{\text{u}} = u_{\text{opt}}/v_{\text{opt}}, \tag{40}$$

which is the initial point (36) for problem (33).

We now discuss further the reason that the non-convex constraint (39b) can be discarded. Since the strength of the perturbation and the variance of the additive noise are unknown, the multiplicative noise becomes an additive noise. As a consequence, we can not estimate the nuisance parameters σ_e^2 and σ_v^2. However, the variance of the equivalent noise σ_z^2 can be estimated by $1/v_{\text{opt}}^2$. In this setting, the constraint (39b) is trivial and can be discarded.

It can be seen that solving the ML estimation problem (33) is much easier than (38). This result is consistent with known results concerning the corresponding linear regression model. In the linear regression model the least squares solution is the ML estimator when the two nuisance parameters σ_v^2 and σ_e^2 are unknown. When the two parameters are known, one has to resort to numerical algorithms to find the ML estimator. In addition, when σ_e^2 is nonzero the least squares estimator and the ML estimator are consistent. In our scenario, we also find that both estimators $\hat{\theta}_{\text{ML}}^{\text{k}}$ and $\hat{\theta}_{\text{ML}}^{\text{u}}$ are consistent.

If $N_g = 2$, the closed-form expression of the ML estimator $\hat{\theta}_{\text{ML}}^{\text{u}}$ exists. Suppose that $\{y_{i,j}\}_{j=1}^{N_i}$ has k_i ones for a given i, then one has $u_{\text{opt}} - v_{\text{opt}}\tau_i = \Phi^{-1}(k_i/N_i)$, $i = 1, 2$. Therefore, the ML estimator is

$$\hat{\theta}_{\text{ML}}^{\text{u}} = \frac{\Phi^{-1}(k_2/N_2)\,\tau_1 - \Phi^{-1}(k_1/N_1)\,\tau_2}{\Phi^{-1}(k_2/N_2) - \Phi^{-1}(k_1/N_1)}. \tag{41}$$

Similar results, consistent with (41), are obtained in [8], which focuses on the additive noise only case.

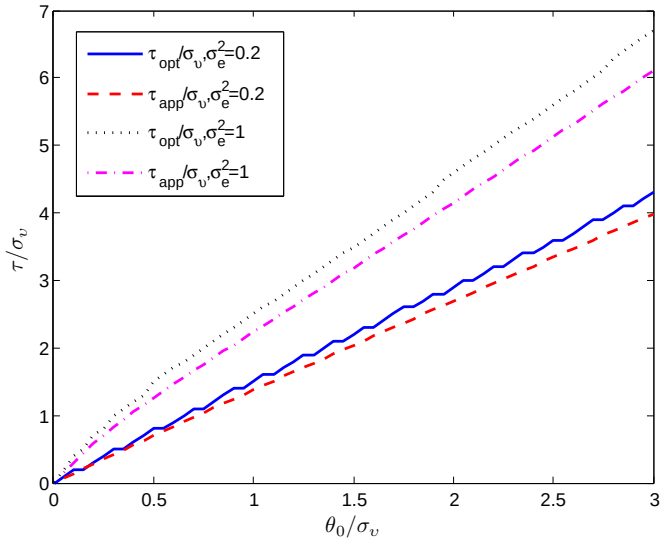

Figure 3: The optimal and approximate threshold (19) as a function of the true value θ_0 normalized by the standard deviation of the additive noise σ_υ.

6 Numerical Simulations

In this section, various numerical calculations are performed to substantiate the corresponding theoretical analysis.

The first calculation shows that (18) is a good approximation of (13). Due to symmetry, $\theta \geq 0$ is considered only. We consider two cases corresponding to $\sigma_e^2 = 0.2$ and $\sigma_e^2 = 1$, and the optimal threshold τ_{opt} is numerically found. The results are plotted in Figure 3. It is shown that the approximately optimal threshold given by (19) approximates the optimal threshold well, especially when either θ/σ_υ or σ_e is small. In addition, the approximately optimal threshold τ_{app} always underestimates the optimal τ_{opt} for $\theta > 0$, and the optimal threshold τ_{opt} is not equal to θ_0, which is different from the situation of additive noise only.

The effectiveness of the relaxed method described in Subsection 4.2 for finding the ML estimator $\hat{\theta}_{\text{ML}}^{\text{k}}$ is demonstrated in the second calculation. The parameters are set as follows: $\sigma_e^2 = 1$, $\sigma_\upsilon^2 = 1$, $\theta_0 = 2$, $\Delta = 3$, $N_g = 2$, and $N_i = 100, i = 1, 2$. The thresholds are chosen as the two boundary points $\pm\Delta$. From Figure 4, it can be seen that all the local minimum points are lying in the range $(-\Delta, \Delta)$, while there exists a unique global minimum of the negative log-likelihood function. The optimum of the relaxed optimization problem is near the global minimum. Thus numerical algorithms will be well positioned to find the ML estimator $\hat{\theta}_{\text{ML}}^{\text{k}}$ by choosing the initial point θ_{init} (36).

The significance of selecting appropriate nonidentical thresholds is validated next. The dynamic range of the parameter is $\Delta = 2$, the number of discrete values for $\theta \in (-\Delta, \Delta)$ and $\tau \in [-\tau_{\max}, \tau_{\max}]$ are $M = 601$ and $T = 201$, respectively. We consider two scenarios.

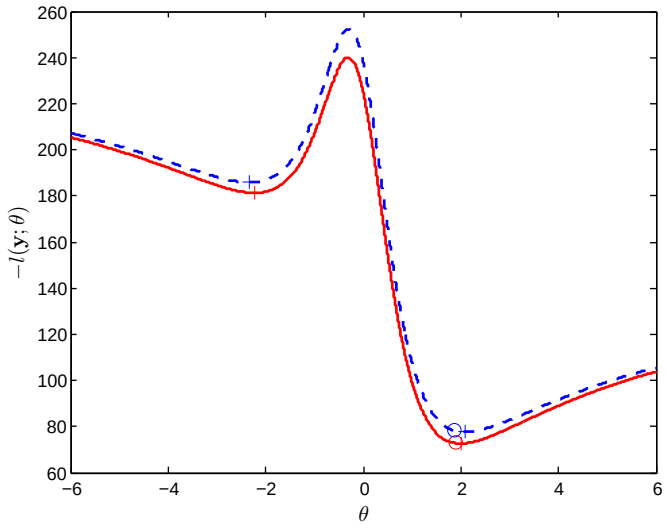

Figure 4: The effectiveness of the relaxed method. The red solid line and the blue dashed line represents two realizations of the negative log-likelihood functions. The marks + and ∘ denote the local minimum of the negative log-likelihood function and the optimal point of the relaxed problem, respectively. Note that the above negative log-likelihood function has two local optimum, and one of the local optimum is the global optimum.

One corresponds to $\sigma_v^2 = 0.1$ and $\sigma_e^2 = 0.1$, Thus the optimal threshold at $\theta = \Delta$ is approximated by $\tau_{\max} \approx 2.35$ by (19). The other is $\sigma_v^2 = 1$ and $\sigma_e^2 = 0.5$. It can be calculated that $\tau_{\max} \approx 3.37$. In Figure 5, we plot the ARE γ with respect to θ. It can be seen that setting the single threshold to zero works well when θ is near zero. However, this works poorly when θ is near the boundary. For the optimized nonidentical thresholds case, the ARE is almost unchanged over the dynamic range of θ, which shows that the thresholds we optimized are robust for estimation.

Meanwhile, we plot the approximate PMF $\rho_{\rm ds}^*(\tau)$ and the approximate parameter PMF $\lambda_{\rm ds}^*(\theta)$ in Figure 6. Note that the nearby thresholds or the parameters are combined together and represented by their mean, and some thresholds whose probability is small are abandoned. It can be seen that although the range of the threshold is larger in the more noisy scenario, the number of thresholds is smaller. In fact, the optimal thresholds should be matched to the additive and multiplicative noise such that the worst-case ARE is maximized. Intuitively, when the variance of noise increases, the optimal threshold spacing could become larger. The threshold spacing increases faster than the range of the threshold in this case, thus the number of thresholds is smaller in the more noisy scenario. It is seen that $\rho_{\rm ds}^*(\tau)$ is large at the boundary. In fact, this phenomenon can be explained by the dual variable $\lambda_{\rm ds}^*(\theta)$. Since the approximate PMF $\lambda_{\rm ds}^*(\theta)$ is large at the boundary, a large fraction of receivers is allocated with a large threshold to ensure that the ARE is robust in the parameter range.

Finally, the performances of the two ML estimators $\hat{\theta}_{\rm ML}^{\rm k}$ (33) and $\hat{\theta}_{\rm ML}^{\rm u}$ (38) are compared

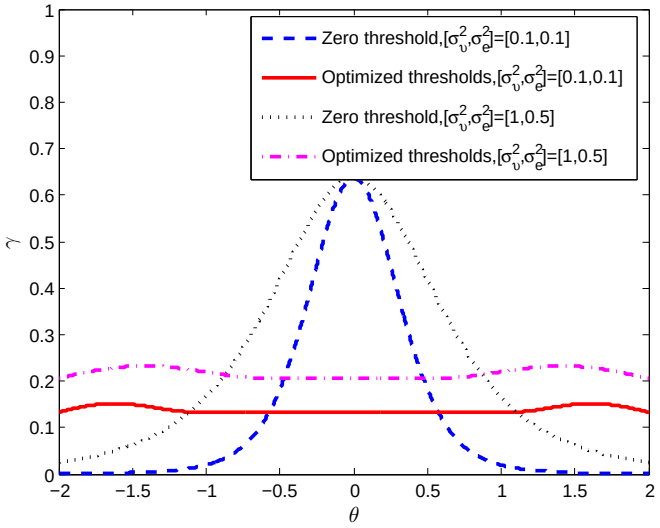

Figure 5: The ARE with different thresholds versus θ. The blue dashed line and the red solid line denote the ARE with the identical threshold being zero and the optimized thresholds, respectively, in the case of $[\sigma_v^2, \sigma_e^2] = [0.1, 0.1]$. The black dotted line and the magenta dash-dot line denote the ARE with the identical threshold being zero and the optimized thresholds, respectively, in the case of $[\sigma_v^2, \sigma_e^2] = [1, 0.5]$.

with their corresponding CRLBs. The parameters are the same as the previous investigation, and we consider the more noisy scenario corresponding to $[\sigma_v^2, \sigma_e^2] = [1, 0.5]$. The true value θ_0 is drawn randomly from the interval $(-\Delta, \Delta)$. In this simulation, $\theta_0 = 1.26$. From Figure 6, the thresholds are approximated by $\boldsymbol{\tau} = [-3.3, -1.1, 1.1, 3.3]^{\mathrm{T}}$, and the threshold PMF can be approximated by $\boldsymbol{\rho}(\tau) = [0.25, 0.25, 0.25, 0.25]^{\mathrm{T}}$. The MSEs and biases are averaged over 5000 Monte Carlo simulations, and the results are plotted in Figure 7 and Figure 8. It can be seen that the MSE performance of the ML estimator $\hat{\theta}_{\mathrm{ML}}^{\mathrm{u}}$ is smaller than the corresponding $\mathrm{CRLB}_{\mathrm{u}}(\mathbf{y};\theta)$ (37) when the total number of observations is less than 100. The reason is that the CRLB is a lower bound for unbiased estimators. When the number of observations is small, the ML estimator $\hat{\theta}_{\mathrm{ML}}^{\mathrm{u}}$ is biased, which can be seen in Figure 8. As a consequence, the MSE performance of the ML estimator $\hat{\theta}_{\mathrm{ML}}^{\mathrm{u}}$ could be smaller than $\mathrm{CRLB}_{\mathrm{u}}(\mathbf{y};\theta)$ (25) [42]. For the ML estimator $\hat{\theta}_{\mathrm{ML}}^{\mathrm{k}}$, it is biased and its MSE is larger than $\mathrm{CRLB}(\mathbf{y};\theta)$ when N is small. This phenomenon is documented [43]. As the number of observations increases, the two ML estimators become asymptotically unbiased and achieve their corresponding CRLB asymptotically.

7 Conclusion

In this paper, we have studied the problem of estimating a deterministic parameter from binary quantized observations from a combined additive and multiplicative noise environment.

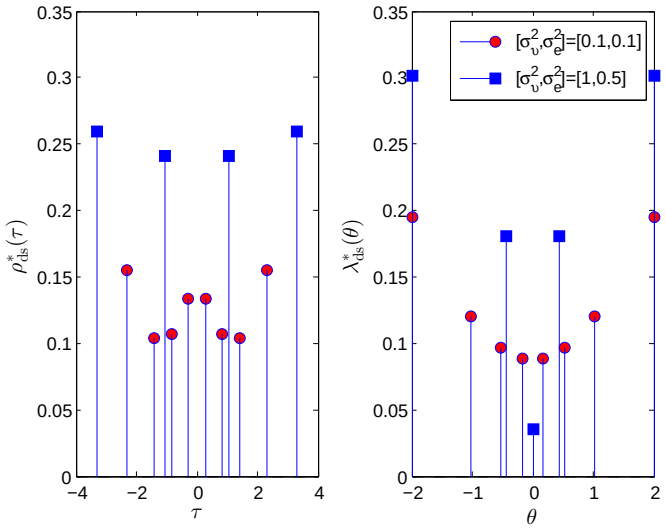

Figure 6: The approximate threshold PMF $\rho^*_{\rm ds}(\tau)$ and the approximate parameter PMF $\lambda^*_{\rm ds}(\theta)$ in two scenarios. The red circle marker, the blue square marker denote the $[\sigma_v^2, \sigma_e^2] = [0.1, 0.1]$, $[\sigma_v^2, \sigma_e^2] = [1, 0.5]$ cases, respectively.

Firstly, it was shown that the ML estimator may have two optimum points when each sensor or receiver utilized an identical threshold, in contrast to the extensively studied additive noise only case. It is very interesting that this relatively minor change of adding multiplicative noise leads to cases with an unidentifiable parameter unless the common threshold is designed properly and the parameter is known to lie in an open interval, which does not occur in the additive noise case. Since similar things may also occur in other estimation problems, these results are especially informative and expand the accumulated knowledge concerning signal processing with quantized data in an important way. The closed-form expression of the ML estimator was derived to estimate the unknown parameter and it was proved to be asymptotically efficient provided the common threshold is designed properly and the parameter is known to lie in an open interval. The CRLB was utilized to analyze the performance of the ML estimator. It was demonstrated that the optimal threshold was not equal to the true value of θ in contrast with the additive noise only case. Next, a strategy where groups of senors or receivers can employ different thresholds was proposed to estimate the underlying parameter. The CRLB was established in the situation where the variances of the additive noise and the multiplicative noise was known. Then the ARE criterion was used to optimize the nonidentical thresholds. Since the original ML estimation problem was non-convex, a relaxed method was proposed to find a good initial point for a gradient search. Thirdly, it was demonstrated that the parameter could still be estimated when the variance of the additive noise and the strength of the multiplicative noise were unknown. The corresponding CRLB was derived. Finally, the ML estimation problem was transformed to be a convex optimization problem, which can be solved efficiently. It was

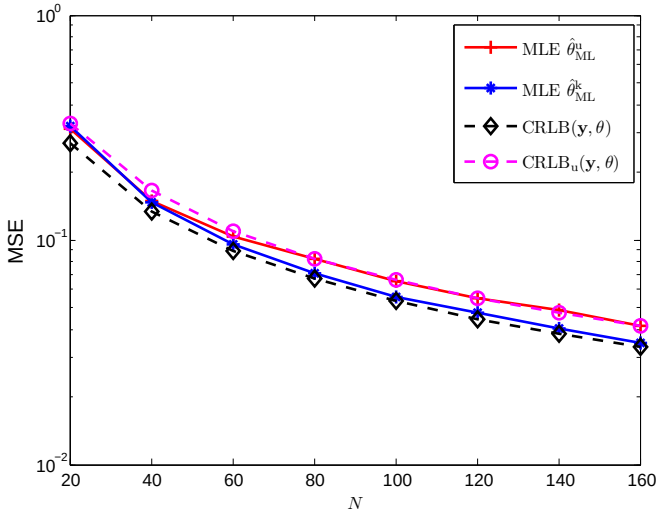

Figure 7: The MSE performance of the two ML estimators compared with their corresponding CRLB. The red solid line denotes the ML estimator $\hat{\theta}^{\mathrm{k}}_{\mathrm{ML}}$ when the nuisance parameters σ_e^2 and σ_v^2 are known, and the dashed magenta line denotes its corresponding $\mathrm{CRLB}_{\mathrm{u}}(\mathbf{y};\theta)$. While the blue solid line denotes the ML estimator $\hat{\theta}^{\mathrm{u}}_{\mathrm{ML}}$ when the two nuisance parameters are unknown, and the black dashed line denotes its corresponding $\mathrm{CRLB}(\mathbf{y};\theta)$.

demonstrated that both the known variance and the unknown variance ML estimators are consistent, and asymptotically achieve the corresponding CRLB. Although we focused on the binary quantization and scalar parameter estimation, it is possible to obtain the CRLB of multi-bit quantization. As for vector parameter estimation, it is more challenging but a very interesting topic where it seems similar approaches can be applied, see [46].

8 Appendix

8.1 Derivation of the ML Estimator $\hat{\theta}_{\mathrm{ML}}$ (20)

Proof We consider three cases corresponding to $\tau = 0$, $\tau > 0$, and $\tau < 0$.

For the case $\tau = 0$, one can see that $|\hat{\omega}_{\mathrm{ML}}| < 1/\sigma_e$, and the ML estimator is θ_3.

For the case $\tau < 0$, $\omega(\theta,\tau)$ is monotone increasing in the interval $\theta \in (-\infty, -\sigma_v^2/(\tau\sigma_e^2)]$, and $\omega(\theta,\tau)$ is monotone decreasing in the interval $\theta \in [-\sigma_v^2/(\tau\sigma_e^2), \infty)$. Meanwhile, we have $-\sigma_v^2/(\tau\sigma_e^2) \geq \Delta$ and $\lim_{\theta\to\infty} \omega(\theta,\tau) = 1/\sigma_e$. When $1/\sigma_e < \hat{\omega}_{\mathrm{ML}} \leq \Omega_2$, there exists two roots θ_1 and θ_2, and the smaller root lies in the interval $(-\Delta, \Delta)$. It can be checked that $\theta_1 > \theta_2$. Therefore, θ_2 is the ML estimator. When $\hat{\omega}_{\mathrm{ML}} = 1/\sigma_e$, θ_4 is the ML estimator. When $|\hat{\omega}_{\mathrm{ML}}| < 1/\sigma_e$, it can be shown that $(\theta_1 - \tau)\hat{\omega}_{\mathrm{ML}} < 0$ and $(\theta_2 - \tau)\hat{\omega}_{\mathrm{ML}} > 0$. Thus θ_2 is the ML estimator. Since $\hat{\omega}_{\mathrm{ML}} \geq \Omega_1 > -1/\sigma_e$, there does not exist other cases.

The case that $\tau > 0$ can be analyzed similarly as $\tau < 0$.

To sum up, the ML estimation of θ is shown in Table 1. ∎

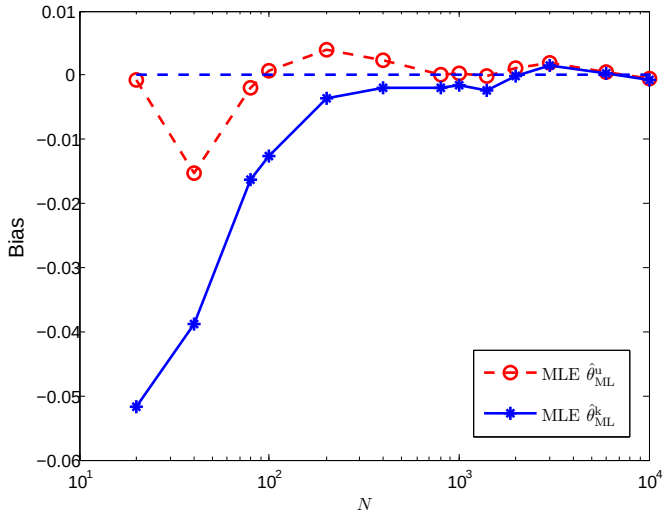

Figure 8: The bias of the two ML estimators $\hat{\theta}^{\mathrm{k}}_{\mathrm{ML}}$ and $\hat{\theta}^{\mathrm{u}}_{\mathrm{ML}}$, where the bias is defined as $\mathrm{Bias}(\theta) = \mathrm{E}_{\mathbf{y}}(\hat{\theta}) - \theta$.

8.2 Computation of the Gradient and Hessian of (35a)

In this appendix, both the Gradient and Hessian of (35a) are computed. For brevity, let $f(\tilde{\mathbf{x}})$ denote the function (35a), which is $f(\tilde{\mathbf{x}}) = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_i} \log \Phi\left(y_{i,j}\mathbf{a}_i^{\mathrm{T}}\tilde{\mathbf{x}}\right)$, where $\tilde{\mathbf{x}} = [u, v]^{\mathrm{T}}$ and $\mathbf{a}_i = [1, -\tau_i]^{\mathrm{T}}$. Its gradient and Hessian are

$$\nabla_{\tilde{\mathbf{x}}} f(\tilde{\mathbf{x}}) = -\frac{1}{\sqrt{2\pi}} \sum_{i=1}^{N_g}\sum_{j=1}^{N_i} \frac{y_{i,j}}{\Phi\left(y_{i,j}\mathbf{a}_i^{\mathrm{T}}\tilde{\mathbf{x}}\right)} \mathrm{e}^{-\frac{\left(\mathbf{a}_i^{\mathrm{T}}\tilde{\mathbf{x}}\right)^2}{2}} \mathbf{a}_i,$$

and

$$\begin{aligned}\nabla^2_{\tilde{\mathbf{x}}} f(\tilde{\mathbf{x}}) = &\frac{1}{\sqrt{2\pi}} \sum_{i=1}^{N_g}\sum_{j=1}^{N_i} \frac{y_{i,j}}{\Phi\left(y_{i,j}\mathbf{a}_i^{\mathrm{T}}\tilde{\mathbf{x}}\right)} \mathrm{e}^{-\frac{\left(\mathbf{a}_i^{\mathrm{T}}\tilde{\mathbf{x}}\right)^2}{2}} (\mathbf{a}_i^{\mathrm{T}}\tilde{\mathbf{x}})\mathbf{a}_i\mathbf{a}_i^{\mathrm{T}} \\ &+ \frac{1}{2\pi} \sum_{i=1}^{N_g}\sum_{j=1}^{N_i} \frac{1}{\Phi^2\left(y_{i,j}\mathbf{a}_i^{\mathrm{T}}\tilde{\mathbf{x}}\right)} \mathrm{e}^{-\left(\mathbf{a}_i^{\mathrm{T}}\tilde{\mathbf{x}}\right)^2} \mathbf{a}_i\mathbf{a}_i^{\mathrm{T}}, \end{aligned} \tag{42}$$

respectively.

8.3 Proof of Proposition 3

Proof We first show that there exists an optimal $\rho^*(\tau)$ symmetric around the origin. Suppose that $\rho_o(\tau)$ is optimum. It can be shown that $\rho_o(-\tau) \in \mathcal{X}(\infty)$. Besides, for the objective function of (29), we have

$$\begin{aligned}&\inf_{\theta\in(-\Delta,\Delta)} \mathrm{E}_{\tau}\left[r(\theta,\tau)\right] = \inf_{\theta\in(-\Delta,\Delta)} \mathrm{E}_{-\tau}\left[r(\theta,-\tau)\right] \\ =&\inf_{\theta\in(-\Delta,\Delta)} \mathrm{E}_{-\tau}\left[r(-\theta,\tau)\right] = \inf_{\theta\in(-\Delta,\Delta)} \mathrm{E}_{-\tau}\left[r(\theta,\tau)\right], \end{aligned} \tag{43}$$

where the first equality follows by using a change of variable $\tau = -\tau$, the second equality follows by $r(\theta, -\tau) = r(-\theta, \tau)$, and the last equality follows from $\theta \in (-\Delta, \Delta)$. Equation (43) demonstrates that the optimized value of the objective function (29) with $\rho_o(-\tau)$ is the same as that of $\rho_o(\tau)$. As a consequence, $\rho_o(-\tau)$ is also optimum. We can construct a new optimal PDF $\rho^*(\tau) = (\rho_o(\tau) + \rho_o(-\tau))/2$ which is symmetric around the origin. Similarly, we can find an optimal PDF $\rho_l^*(\tau)$ symmetric around the origin.

Next we will show that the support of $\rho_l^*(\tau)$ is bounded. We may require that $\rho_l(\tau) = \rho_l(-\tau)$, thus $\mathrm{E}_\tau[s(\theta, \tau)]$ can be expressed as

$$\mathrm{E}_\tau[s(\theta, \tau)] = \int_0^\infty (s(\theta, \tau) + s(\theta, -\tau))\rho_l(\tau)\mathrm{d}\tau. \tag{44}$$

By defining $f(\theta, \tau) \triangleq s(\theta, \tau) + s(\theta, -\tau)$, it can be shown that for any $\theta > 0$, there exist a $\tau_{\max}(\theta)$ such that $f(\theta, \tau)$ is decreasing in $[\tau_{\max}(\theta), \infty)$. Straightforward calculation shows that $\tau_{\max}(0) = 0$. By defining

$$\tau_{\max} = \sup_{\theta \in [0, \Delta)} \tau_{\max}(\theta), \tag{45}$$

$f(\theta, \tau)$ is decreasing in the interval $[\tau_{\max}, \infty)$ for any $\theta \in [0, \Delta)$. Since $\partial f(\theta, \tau)/\partial\tau \sim \partial f(-\theta, \tau)/\partial\tau$, similar analysis will show that $f(\theta, \tau)$ is decreasing in the interval $[\tau_{\max}, \infty)$ for any $\theta \in (-\Delta, 0]$. Thus $f(\theta, \tau)$ is decreasing in the interval $[\tau_{\max}, \infty)$ for any $\theta \in (-\Delta, \Delta)$. To ensure that the value of the optimized objective function of the lower bound version of (29) attains its maximum, the support of $\rho_l^*(\tau)$ is bounded and satisfies $\rho_l^*(\tau) = 0,\ \forall |\tau| \geq \tau_{\max}$. ∎

8.4 Derivation of the dual of (31)

Proof We will give the dual of problem (31). By introducing the dual variables $\lambda(\theta)$, ν and $\chi(\tau)$, we construct the Lagrangian

$$\begin{aligned} L(t, \rho(\tau), \lambda(\theta), \nu) = t - \int_{-\Delta}^{\Delta} \left[t + \int_{-\tau_{\max}}^{\tau_{\max}} r(\theta, \tau)\rho(\tau)\mathrm{d}\tau \right] \lambda(\theta)\mathrm{d}\theta \\ + \nu \left(\int_{-\tau_{\max}}^{\tau_{\max}} \rho(\tau)\mathrm{d}\tau - 1 \right) - \int_{-\tau_{\max}}^{\tau_{\max}} \rho(\tau)\chi(\tau)\mathrm{d}\tau. \end{aligned}$$

Through the Karush-Kuhn-Tucker (KKT) conditions, we obtain the dual problem as

$$\begin{aligned} \underset{\lambda(\theta) \in \mathcal{X}(\Delta), \chi(\tau), \nu}{\text{maximize}} \quad & -\nu, && \text{(46a)} \\ \text{subject to} \quad & \nu - \int_{-\Delta}^{\Delta} r(\theta, \tau)\lambda(\theta)\mathrm{d}\theta - \chi(\tau) = 0, \\ & \forall \tau \in [-\tau_{\max}, \tau_{\max}], && \text{(46b)} \\ & \chi(\tau) \geq 0, \forall \tau \in [-\tau_{\max}, \tau_{\max}]. && \text{(46c)} \end{aligned}$$

∎

8.5 Proof of Proposition 4

PROOF The corresponding log-likelihood function is (23). To obtain the closed-form expression of the CRLB, the following conclusion is utilized in the case of vector parameter CRLB for transformation.

Assume that we wish to estimate $\boldsymbol{\theta} = \mathbf{g}(\tilde{\mathbf{x}})$, where $\mathbf{g}$ is a r-dimensional function and $\tilde{\mathbf{x}}$ is a s-dimensional parameter vector. Then the CRLB of $\boldsymbol{\theta}$ from $\tilde{\mathbf{x}}$ is given by [45]

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}) \succeq \partial \mathbf{g}(\tilde{\mathbf{x}})/\partial \tilde{\mathbf{x}} \left(\mathbf{J}(\tilde{\mathbf{x}})\right)^{-1} \left(\partial \mathbf{g}(\tilde{\mathbf{x}})/\partial \tilde{\mathbf{x}}\right)^{\mathrm{T}}, \tag{47}$$

where $\mathbf{J}(\tilde{\mathbf{x}})$ is the FI matrix of $\tilde{\mathbf{x}}$.

In our case, $\tilde{\mathbf{x}} = [u, v]^{\mathrm{T}}$, and $\theta = g(\tilde{\mathbf{x}}) = u/v$. The FI matrix $\mathbf{J}(\tilde{\mathbf{x}})$ is obtained as $\mathbf{J}(\tilde{\mathbf{x}}) = \mathrm{E}_{\mathbf{y}}[\nabla^2_{\tilde{\mathbf{x}}} f(\tilde{\mathbf{x}})]$, where $\nabla^2_{\tilde{\mathbf{x}}} f(\tilde{\mathbf{x}})$ is (42). Therefore, the FI matrix is calculated to be

$$\mathbf{J}(\tilde{\mathbf{x}}) = \begin{bmatrix} \sum_{i=1}^{N_g} q(\theta, \tau_i)\rho_i & -\sum_{i=1}^{N_g} q(\theta, \tau_i)\tau_i\rho_i \\ -\sum_{i=1}^{N_g} q(\theta, \tau_i)\tau_i\rho_i & \sum_{i=1}^{N_g} q(\theta, \tau_i)\tau_i^2\rho_i \end{bmatrix} N, \tag{48}$$

where $q(\theta, \tau)$ is defined as (14). The Jacobian is

$$\partial \mathbf{g}(\tilde{\mathbf{x}})/\partial \tilde{\mathbf{x}} = [1/v, -u/v^2]. \tag{49}$$

The CRLB is computed by substituting (48) and (49) in (47). Therefore, the desired result (37) is established. ∎

References

[1] B. S. Rao, H. F. Durrant-Whyte and J. A. Sheen,"A fully decentralized multi-sensor system for tracking and surveillance," *Int. J. Robot. Res.*, vol. 12, no. 1, pp. 20-44, Feb. 1993.

[2] R. Niu and P. K. Varshney, "Target location estimation in sensor networks with quantized data," *IEEE Trans. Signal Process.*, vol. 54, no. 12, Dec. 2006.

[3] S. Kim, S. Pakzad, D. Culler, J. Demmel, G. Fenves, S. Glaser and M. Turon, "Health monitoring of civil infrastructures using wireless sensor networks", in *6th International Symposium on Information Processing in Sensor Networks,* April 2007.

[4] M. Yeddanapudi, Y. Bar-Shalom and K. R. Pattipati, "IMM estimation for multi-target-multisensor air traffic surveillance," *Proc. IEEE*, vol. 85, no. 1, pp. 120-138, Jan. 1997.

[5] H. C. Papadopoulos, G. W. Wornell and A. V. Oppenheim, "Sequential signal encoding from noisy measurements using quantizers with dynamic bias control," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 978-1002, Mar. 2001.

[6] A. Ribeiro, I. D. Schizas, S. I. Roumeliotis, and G. B. Giannakis, "Kalman filtering in wireless sensor networks: Incorporating communication cost in state estimation problems," *IEEE Control Syst. Mag.*, vol. 30, pp. 66-86, 2010.

[7] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor Networks-part I: Gaussian case," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1131-1143, Mar. 2006.

[8] A. Ribeiro, and G.B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks-part II: Unknown probability density function," *IEEE Trans. Signal Process.*, vol. 54, no. 7, pp. 2784-2796, July, 2006.

[9] Z. Luo, "Universal decentralized estimation in a bandwidth constrained sensor network," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2210-2219, June 2005.

[10] Z. Luo, "An isotropic universal decentralized estimation scheme for a bandwidth constrained ad hoc sensor network," *IEEE J. Sel. Area. Comm.*, vol. 23, no. 4, pp. 735-744, Apr. 2005.

[11] Z. Luo and J. Xiao, "Decentralized estimation in an inhomogeneous sensing environment," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3564-3575, Oct. 2005.

[12] O. Dabeer and A. Karnik, "Signal parameter estimation using 1-bit dithered quantization," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5389-5405, Dec. 2006.

[13] G. O. Balkan and S. Gezici, "CRLB based optimal noise enhanced parameter estimation using quantized observations," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 477-480, May 2010.

[14] A. Wiesel, Y. C. Eldar and A. Beck, "Maximum likelihood estimation in linear models with a Gaussian model matrix," *IEEE Signal Process. Lett.*, vol. 13, no. 5, pp. 292-295, May 2006.

[15] A. Wiesel, Y. C. Eldar and A. Yeredor, "Linear regression with Gaussian model uncertainty: Algorithms and bounds," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2194-2205, Jun. 2008.

[16] O. Besson, F. Vincent, P. Stocia and A. B. Gershman, "Approximate maximum likelihood estimators for array processing in multiplicative noise environments," *IEEE Trans. Signal Process.*, vol. 48, no. 9, pp. 2506-2517, Sep. 2000.

[17] J. Zhu, X. Wang, X. Lin and Y. Gu, "Maximum likelihood estimation from sign measurements with sensing matrix perturbation," *IEEE Trans. Signal Process.,* vol. 62, no. 15, pp. 3741-3753, Aug. 2014.

[18] I. Song and S. A. Kassam, "Locally optimum detection of signals in a generalized observation model: The known signal case," *IEEE Trans. Inf. Theory,* vol. 36, no. 3, pp. 502-515, May 1990.

[19] I. Song and S. A. Kassam, "Locally optimum detection of signals in a generalized observation model: The random signal case," *IEEE Trans. Inf. Theory,* vol. 36, no. 3, pp. 516-530, May 1990.

[20] G. L. Stuber, *Principles of Mobile Communication,* Boston, MA: Kluwer, 1996.

[21] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory,* Englewood Cliffs, NJ: Prentice Hall, pp. 47-48, 1993.

[22] E. Tsakonas, J. Jaldén, N. Sidiropoulos and B. Ottersten, "Sparse conjoint analysis through maximum likelihood estimation," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5704-5715, Nov. 2013.

[23] C. D. M. Paulino, and C. A. B. Pereira, "On identifiability of parametric statistical models," *Journal of the Italian Statistical Society*, pp. 125-151, 1994.

[24] J .G. Proakis, *Digital Communications*, McGraw-Hill, New York, pp. 42, 2001.

[25] J. Fang and H. Li, "Distributed adaptive quantization for wireless sensor networks: From Delta modulation to maximum likelihood," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5246-5257, Oct. 2008.

[26] E. L. Lehmann, *Elements of Large-Sample Theory,* Springer-Verlag New York, inc., pp. 456-497, 1999.

[27] J. K. Douglass, L.Wilkens, E. Pantazelou and F. Moss, "Noise enhancement of information transfer in crayfish mechanoreceptors by stochastic resonance," *Nature*, vol. 365, pp. 337-340, Sep. 1993.

[28] M. DeWeese and W. Bialek, "Information flow in sensory neurons," *Nuovo Cimento Soc. Ital. Fys.*, vol. 17D, no. 7-8, pp. 733-741, July-Aug. 1995.

[29] J. Levin and J. Miller, "Broadband neural encoding in the cricket sensory system enhanced by stochastic resonance," *Nature*, vol. 380, no. 6570, pp. 165-168, Mar. 1996.

[30] H. Chen, P. K. Varshney, S. M. Kay and J. H. Michels, "Theory of the stochastic resonance effect in signal detection: Part Ifixed detectors," *IEEE Trans. Signal Process.*, vol. 55, no. 7, July 2007.

[31] S. Bayram and S. Gezici, "On the performance of single-threshold detectors for binary communications in the presence of Gaussian mixture noise," *IEEE Trans. Commun.*, vol. 58, no. 11, Nov. 2010.

[32] D. Rousseau, F. Duan and F. Chapeau-Blondeau, "Suprathreshold stochastic resonance and noise-enhanced Fisher information in arrays of threshold devices", *Phys. Rev. E*, vol 68, no.3, Sep. 2003.

[33] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory,* Englewood Cliffs, NJ: Prentice Hall, pp. 167, 1993.

[34] P. Stoica and T. L.Marzetta, "Parameter estimation problems with singular information matrices," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 87-90, Jan. 2001.

[35] M. L. Fowler and M. Chen, "Fisher-information-based data compression for estimation using two sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 3, pp. 1131-1137, Jul. 2005.

[36] P. Venkitasubramaniam, G. Mergen, L. Tong, and A. Swami, "Quantization for distributed estimation in large scale sensor networks" in *Proc. ICISIP*, Bangalore, India, Dec. 2005, pp. 121-126.

[37] G. Casella and R. Berger, *Statistical Inference*, Pacific Grove, CA: Brooks/Cole, pp. 476-477, 1990.

[38] P. Venkitasubramaniam, L. Tong, and A. Swami, "Quantization for maximin ARE in distributed estimation," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3596-3605, July, 2007.

[39] D. G. Luenberger, *Optimization by Vector Space Methods*, John Wiley & Sons, 1969.

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

[41] M. Grant and S. Boyd, *CVX: Matlab software for disciplined convex programming [Online]*, Available: http://stanford.edu/ boyd/cvx Dec. 2008.

[42] H. L. Van Trees, K. L. Bell and Z. Tian, *Detection, Estimation, and Modulation Theory, Part I*, John Wiley & Sons, pp. 307-310, 2013.

[43] H. L. Van Trees, K. L. Bell and Z. Tian, *Detection, Estimation, and Modulation Theory, Part I*, John Wiley & Sons, pp. 259-260, 2013.

[44] A. Geletu, "Solving Optimization Problems using the Matlab Optimization Toolbox - a Tutorial," available at http://www.tu-ilmenau.de/fileadmin/media/simulation/Lehre/Vorlesungsskripte/ Lecture_materials_Abebe/OptimizatioWithMatlab.pdf, Dec, 2007.

[45] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory,* Englewood Cliffs, NJ: Prentice Hall, pp. 45-46, 1993.

[46] J. Fang and H. Li, "Hyperplane-based vector quantization for distributed parameter estimation in wireless sensor networks," *IEEE Trans. Inf. Theory,* vol. 55, no. 12, pp. 5682-5699, Dec. 2009.