DISTRIBUTED ESTIMATION IN GAUSSIAN NOISE FOR BANDWIDTH-CONSTRAINED WIRELESS SENSOR NETWORKS*

Alejandro Ribeiro and Georgios B. Giannakis

Dept. of ECE, University of Minnesota
200 Union Street, Minneapolis, MN 55455, USA

ABSTRACT

We study deterministic mean-location parameter estimation when only quantized versions of the original observations are available, due to bandwidth constraints. When the dynamic range of the parameter is small or comparable with the noise variance, we introduce a class of maximum likelihood estimators that require transmitting just one bit per sensor to achieve an estimation variance close to that of the (clairvoyant) sample mean estimator. When the dynamic range is comparable or large relative to the noise standard deviation, we show that an optimum quantization step exists to achieve the best possible variance for a given bandwidth constraint. We will also establish that in certain cases the sample mean estimator formed by quantized observations is preferable for complexity reasons. We finally address implementation issues and guarantee that all the numerical maximizations required by the proposed estimators are concave.

1. INTRODUCTION

Wireless Sensor Networks (WSN) comprise a large number of geographically distributed nodes characterized by low power constraints and limited computation capability. However, with sensor collaboration, potentially powerful networks can be constructed to monitor and control environments [2]. While a number of works address sensor collaboration for distributed detection (see e.g., [8] and references therein), the equally challenging problem of distributed estimation has not received much attention. In distributed estimation based on data collected by a WSN, each sensor has available a subset of the observations that must be either transmitted to a central node (WSN with a fusion center) or shared among nodes (ad-hoc WSN).

Under either WSN configuration, attention has focused on decentralized algorithms exploiting spatial correlation to reduce transmission requirements, e.g., [1, 6]. A not so well studied issue is that bandwidth limits necessitate the estimator to be formed using quantized versions of the original observations. In this setup, quantization becomes an integral part of the estimation process, since one may think of quantization as a means of constructing *binary observations*. We then deal with *parameter estimation given a set of binary observations*. When the noise pdf is known, transmitting a single bit per sensor can lead to minimal loss in the estimator variance compared with the clairvoyant estimator [5]. When the noise pdf is unknown, pdf-unaware estimators based on quantized data have been introduced in [4].

Our focus is on bandwidth-constrained distributed mean-location parameter estimation in Additive White Gaussian Noise (AWGN). We seek Maximum Likelihood Estimators (MLE) and benchmark their variances with the Cramer-Rao Lower Bound (CRLB); that is asymptotically achieved by the MLE. We will show that the deciding factor in the choice of the estimator is a proper form of Signal Noise Ratio (SNR), defined here as the dynamic range of the parameter square over the noise variance.

* Work in this paper was prepared through collaborative participation in the Communications and Networks Consortium sponsored by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

2. PROBLEM STATEMENT

We consider the problem of estimating a deterministic scalar parameter θ in the presence of zero-mean AWGN,

$$x(n) = \theta + w(n), \quad n = 0, 1, \dots, N-1, \tag{1}$$

where $w(n) \sim \mathcal{N}(0, \sigma^2)$, and n is the sensor index. Throughout, we will use $p(w) := 1/(\sqrt{2\pi}\sigma)\exp[-w^2/(2\sigma^2)]$ to denote the noise probability density function (pdf).

If all the observations $\{x(n)\}_{n=0}^{N-1}$ were available, the MLE of θ would be the Sample Mean Estimator, $\bar{x} = N^{-1}\sum_{n=0}^{N-1} x(n)$. This can be regarded as a clairvoyant estimator with variance

$$\mathrm{var}(\bar{x}) = \frac{\sigma^2}{N}. \tag{2}$$

Due to bandwidth limitations, the observations $x(n)$ have to be quantized and estimation can only be based on these quantized values. To this end, we will henceforth think of quantization as the construction of a set of indicator variables (referred to, as binary observations)

$$b_k(n) = 1\{x(n) \in (\tau_k, +\infty)\}, \quad k \in \mathbf{Z}, \tag{3}$$

where τ_k is a threshold defining $b_k(n)$ and $\mathbf{Z}$ denotes the set of integers. The bandwidth constraint manifests itself in dictating estimation of θ to be based on the binary observations $\{b_k(n), k \in \mathbf{Z}\}_{n=0}^{N-1}$. The initial goal of this paper is twofold: i) *develop the MLE for estimating θ given a set of binary observations*, and ii) *study the associated CRLB* – a bound achieved by the MLE as $N \to \infty$.

Instrumental to the ensuing derivations is the fact that $b_k(n)$ as defined in (3) is a Bernoulli variable with parameter

$$q_k(\theta) := \Pr\{b_k(n) = 1\} = F(\tau_k - \theta), \quad k \in \mathbf{Z}, \tag{4}$$

where $F(x) := 1/(\sqrt{2\pi}\sigma)\int_x^{+\infty}\exp(-u^2/2\sigma^2)\,du$ denotes the complementary Cumulative Distribution Function (CDF) of the noise.

The problem under consideration bears similarities and differences with quantization. On the one hand, for a fixed n the set of binary observations $\{b_k(n), k \in \mathbf{Z}\}$ specifies uniquely the quantized value of $x(n)$ to one of the pre-specified levels $\{\tau_k, k \in \mathbf{Z}\}$. On the other hand, different from quantization in which the goal is to *reconstruct* $x(n)$ (and the optimum solution is known to be given by Lloyd's quantizer), our goal is to *estimate* θ.

3. MLE BASED ON BINARY OBSERVATIONS: COMMON THRESHOLDS

Let us consider the most stringent bandwidth constraint, requiring sensors to transmit one bit per $x(n)$ observation. And as a simple first approach, let every sensor use the same threshold τ_c to form

$$b(n) = 1\{x(n) \in (\tau_c, +\infty)\}, \quad n = 0, 1, \dots, N-1. \tag{5}$$

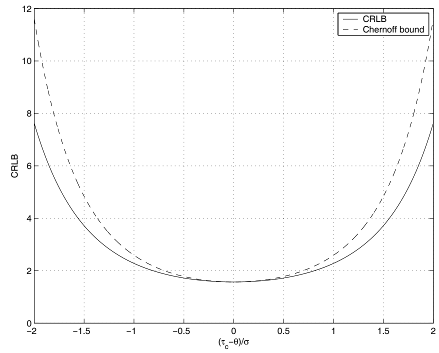

Fig. 1. CRLB and Chernoff bound as a function of the distance between τ_c and θ measured in AWGN standard deviation (σ) units.

Dropping the subscript k, we let $\mathbf{b} := [b(0), \ldots, b(N-1)]^T$, and denote as $q(\theta)$ the parameter of these Bernoulli variables. We are now ready to derive the MLE and the pertinent CRLB; see [5][1].

Proposition 1 *The MLE $\hat{\theta}$ based on the binary observations $\mathbf{b}$ is:*

$$\hat{\theta} = \tau_c - F^{-1}\left(\frac{1}{N}\sum_{n=0}^{N-1} b(n)\right). \quad (6)$$

The CRLB for any unbiased estimator $\hat{\theta}$ based on $\mathbf{b}$ is given by

$$\mathrm{var}(\hat{\theta}) \geq \frac{1}{N}\left[\frac{p^2(\tau_c-\theta)}{F(\tau_c-\theta)[1-F(\tau_c-\theta)]}\right]^{-1} := B(\theta). \quad (7)$$

Proposition 1 asserts that θ can be consistently estimated from a single binary observation per sensor, with variance as small as $B(\theta)$. Minimizing $B(\theta)$ reveals that $B_{\min}$ is achieved when $\tau_c = \theta$, and

$$B_{\min} = \frac{2\pi\sigma^2}{4N} \approx 1.57\frac{\sigma^2}{N}\ . \quad (8)$$

In words, if we place τ_c optimally, the variance increases only by a factor of $\pi/2$ with respect to the clairvoyant estimator $\bar{x}$ that relies on unquantized observations. If we use the (tight) Chernoff bound for the complementary CDF, a simple bound on $B(\theta)$ can be obtained:

$$B(\theta) \leq \frac{\pi\sigma^2}{2N} e^{+\frac{1}{2}[(\tau_c-\theta)/\sigma]^2}. \quad (9)$$

Fig. 1 depicts $B(\theta)$ and its Chernoff bound, from where it becomes apparent that for $|\tau_c - \theta|/\sigma \leq 1$ the increase in variance will be around 2 [c.f. (7), and (9)]. Roughly speaking, to achieve a variance close to $\mathrm{var}(\bar{x})$ in (2), it suffices to place τ_c "$\sigma-$close" to θ.

One can envision an iterative algorithm in which the threshold is iteratively adjusted over time. Call $\tau_c^{(j)}$ the threshold used at time j, and $\hat{\theta}^{(j)}$ the corresponding estimate obtained as in (6). Having this estimate, we can now set $\tau_c^{(j+1)} = \hat{\theta}^{(j)}$, for subsequent estimates to not only benefit from the increased number of observations but also from improved binary observations.

4. MLE BASED ON BINARY OBSERVATIONS: NON-IDENTICAL THRESHOLDS

When the dynamic range of θ is in the order of σ (i.e., the possible values of θ are restricted to an interval of size comparable to σ) the variance of the estimator introduced in Section 3 will be close to $\mathrm{var}(\bar{x})$. When the dynamic range of θ is large relative to σ, a different approach must be pursued using binary observations $b_k(n)$, generated from different regions $(\tau_k, +\infty)$ in order to assure that there will always be a threshold τ_k close to the true parameter.

Let N_k be the total number of sensors transmitting binary observations based on the threshold τ_k, and define $\rho_k := N_k/N$ as the corresponding fraction of sensors. We further suppose that the index k_n chosen by sensor n, is known at the destination (the fusion center or peer sensors in an ad-hoc WSN). Algorithmically, we can summarize our approach in three steps:

[S1] Define a set of thresholds $\boldsymbol{\tau} = \{\tau_k, k \in \mathbf{Z}\}$ and associated frequencies $\boldsymbol{\rho} = \{\rho_k, k \in \mathbf{Z}\}$.

[S2] Convene the index k_n to be used by sensor n; i.e., sensor n generates the binary observation $b_{k_n}(n)$ using the threshold τ_{k_n}. Define $\mathbf{b} := [b_{k_0}(0), \ldots, b_{k_{N-1}}(N-1)]^T$.

[S3] Transmit the corresponding binary observations to find the MLE, as we describe next.

The log-likelihood function is given by

$$L(\theta) = \sum_{n=0}^{N-1} b_{k_n}(n)\ln(q_{k_n}(\theta)) + (1-b_{k_n}(n))\ln(1-q_{k_n}(\theta)), \quad (10)$$

from where the MLE of θ given $\{b_{k_n}\}_{n=0}^{N-1}$, is

$$\hat{\theta} = \arg\max_\theta\{L(\theta)\}. \quad (11)$$

As $\hat{\theta}$ in (11) cannot be found in closed-form, we resort to a numerical search, such as Newton's algorithm based on the iteration

$$\hat{\theta}^{(i+1)} = \hat{\theta}^{(i)} - \frac{\dot{L}(\hat{\theta}^{(i)})}{\ddot{L}(\hat{\theta}^{(i)})}, \quad (12)$$

where $\dot{L}(\theta) := \partial L(\theta)/\partial\theta$, and $\ddot{L}(\theta) := \partial^2 L(\theta)/\partial\theta^2$ are the first and second derivatives of the log-likelihood function. Albeit numerically found, the Newton iteration (12) is guaranteed to converge to the global optimum of $L(\theta)$ thanks to its concavity:

Proposition 2 *$L(\theta)$ given in-* (10) *is concave on θ.*

Furthermore, the CRLB for this problem is:

Proposition 3 *The CRLB for any unbiased estimator $\hat{\theta}$ based on $\mathbf{b}$ is*

$$\begin{aligned} B(\theta, \boldsymbol{\tau}, \boldsymbol{\rho}) &= \frac{1}{N}\left[\sum_k \frac{\rho_k p^2(\tau_k-\theta)}{F(\tau_k-\theta)[1-F(\tau_k-\theta)]}\right]^{-1} \\ &:= \frac{1}{N} S^{-1}(\theta, \boldsymbol{\tau}, \boldsymbol{\rho}). \end{aligned} \quad (13)$$

Using non-identical thresholds across sensors provides an additional degree of freedom, that we exploit in the ensuing subsection.

4.1. Selection of the parameters $(\boldsymbol{\tau}, \boldsymbol{\rho})$

Since the CRLB depends on θ the selection of $(\boldsymbol{\tau}, \boldsymbol{\rho})$ depends not only on the estimator variance for a specific value of θ, but also on how confident we are that the actual parameter will take on this value. To incorporate this confidence we introduce a weighting function, $W(\theta)$, which accounts for the relative importance of different values of θ. For instance, if we know a priori that $\theta \in (\Theta_1, \Theta_2)$, we can choose $W(\theta) = u(\theta-\Theta_1) - u(\theta-\Theta_2)$, where $u(\cdot)$ is the unit step function.

Given this weighting function, a reasonable performance indicator is the weighted variance,

$$\mathcal{C}_W := \int_{-\infty}^{+\infty} W(\theta)\mathrm{var}(\hat{\theta})\, d\theta. \quad (14)$$

Although we only have an expression for the CRLB (13), we know that the MLE will approach this bound as $N \to \infty$. Consequently,

[1] Omitted due to space considerations, proofs pertaining to claims in this paper can be found in [7]

selecting the best possible $(\boldsymbol{\tau}, \boldsymbol{\rho})$ amounts to finding the set $(\boldsymbol{\tau}, \boldsymbol{\rho})$ that minimizes the weighted asymptotic CRLB [c.f (13) and (14)],

$$\lim_{N \to +\infty} N\mathcal{C}_W = N\mathcal{B}_W(\boldsymbol{\tau}, \boldsymbol{\rho}) = \int_{-\infty}^{+\infty} \frac{W(\theta)}{S(\theta, \boldsymbol{\tau}, \boldsymbol{\rho})}\, d\theta. \tag{15}$$

Thus, the optimum set $(\boldsymbol{\tau}^*, \boldsymbol{\rho}^*)$, should be selected as,

$$\begin{aligned}(\boldsymbol{\tau}^*, \boldsymbol{\rho}^*) = \quad & \underset{(\boldsymbol{\tau}, \boldsymbol{\rho})}{\arg\min} \quad \int_{-\infty}^{+\infty} \frac{W(\theta)}{S(\theta, \boldsymbol{\tau}, \boldsymbol{\rho})}\, d\theta, \\ & \text{s.t.} \qquad \sum_k \rho_k = 1, \rho_k \geq 0\ \forall k. \end{aligned} \tag{16}$$

Solving (16) is complex, but introducing a proper relaxation we have been able to obtain the following theorem.

Theorem 1 *The weighted CRLB of any estimator $\hat{\theta}$ based on binary observations must satisfy,*

$$\mathcal{B}_W(\boldsymbol{\tau}, \boldsymbol{\rho}) \geq B_{\min} := \frac{1}{N} \frac{\left[\int_{-\infty}^{+\infty} [W(\theta)]^{\frac{1}{2}}\, d\theta\right]^2}{\int_{-\infty}^{+\infty} \frac{p^2(u)}{F(u)[1-F(u)]}\, du} \tag{17}$$

The bound is attained if and only if there exist a set $(\boldsymbol{\tau}, \boldsymbol{\rho})$ such that

$$S(\theta, \boldsymbol{\tau}, \boldsymbol{\rho}) = \mathcal{K}[W(\theta)]^{\frac{1}{2}}, \quad \mathcal{K} := \frac{\int_{-\infty}^{+\infty} \frac{p^2(u)}{F(u)[1-F(u)]}\, du}{\int_{-\infty}^{+\infty} [W(\theta)]^{\frac{1}{2}}\, d\theta}. \tag{18}$$

The claims of Theorem 1, are reminiscent of Cramer-Rao's Theorem in the sense that (17) establishes a bound, and (18) offers a condition for this bound to be attained. Specializing (17) to a Gaussian-shaped $W(\theta)$, with variance σ_θ^2, we find

$$B_{\min} = \frac{\pi\sqrt{2}}{1.6} \frac{\sigma\sigma_\theta}{N}\ . \tag{19}$$

The best possible weighted variance for any estimator based on a single binary observation per sensor can only be close to the clairvoyant variance when $\sigma_\theta \approx \sigma$, a condition valid in small to medium SNR scenarios. When the SNR is high ($\sigma_\theta \gg \sigma$), the performance gap between (2) and (19) is significant requiring a different approach.

Although we cannot assure that there always exists a set $(\boldsymbol{\tau}, \boldsymbol{\rho})$ such that $S(\theta, \boldsymbol{\tau}, \boldsymbol{\rho}) = \mathcal{K}[W(\theta)]^{\frac{1}{2}}$, we can adopt as a relaxed optimal solution the set $(\boldsymbol{\tau}^\dagger, \boldsymbol{\rho}^\dagger)$ that minimizes the distance between $S(\theta, \boldsymbol{\tau}, \boldsymbol{\rho})$ and $\mathcal{K}[W(\theta)]^{\frac{1}{2}}$, so that

$$\begin{aligned}(\boldsymbol{\tau}^\dagger, \boldsymbol{\rho}^\dagger) = \quad & \underset{(\boldsymbol{\tau}, \boldsymbol{\rho})}{\arg\min} \quad \left\| \mathcal{K}[W(\theta)]^{\frac{1}{2}} - \sum_k s_k(\theta) \right\|_2 \\ & \text{s.t.} \qquad \rho_k > 0 \end{aligned} \tag{20}$$

with,

$$s_k(\theta) = \frac{\rho_k p^2(\tau_k - \theta)}{F(\tau_k - \theta)[1 - F(\tau_k - \theta)]}\ . \tag{21}$$

We emphasize that $(\boldsymbol{\tau}^\dagger, \boldsymbol{\rho}^\dagger)$ obtained as the solution of (20) will in general be different from the optimum $(\boldsymbol{\tau}^*, \boldsymbol{\rho}^*)$ obtained as the solution of (16). Nonetheless numerically solving (20) yields a small minimum distance, illustrating that the estimator (11) based on $(\boldsymbol{\tau}^\dagger, \boldsymbol{\rho}^\dagger)$ is nearly optimal (see Section 7).

5. RELAXING THE BANDWIDTH CONSTRAINT

Variances of the estimators in Sections 3 and 4 are close to $\operatorname{var}(\bar{x})$ when the parameter's range is small, or, in the order of the noise variance. If for a Gaussian weight we define the SNR as $\gamma := \sigma_\theta^2/\sigma^2$, the variance of the estimator defined by (11) is [c.f. (2) and (19)],

$$B_{\min} = \frac{\pi\sqrt{2}}{1.6}\sqrt{\gamma}\, \operatorname{var}(\bar{x}) \tag{22}$$

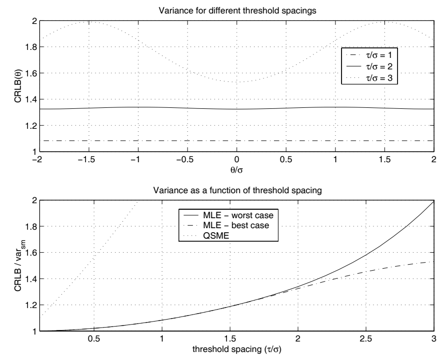

Fig. 2. Variance of the estimator relying on a sequence of binary observations. Room for decreasing variance once $\tau < \sigma$ is small.

indicating that the bound on the average variance grows as $\sqrt{\gamma}$ with respect to the clairvoyant estimator's variance. Recalling that with the optimal set $(\boldsymbol{\tau}^*, \boldsymbol{\rho}^*)$, (11) is the best possible estimator when the bandwidth is constrained to one bit per sensor, the poor performance at high SNR is not a problem of the estimator itself, but is inherent to the harsh bandwidth constraint (just 1 bit per sensor). In order to accommodate high-SNR scenarios, we will allow for transmissions of a larger number of binary observations. Using a sequence of thresholds $\boldsymbol{\tau} := \{\tau_k, k \in \mathbf{Z}\}$, we will rely on multiple binary observations per sensor, $\mathbf{b}(n) := \{b_k(n), k \in \mathbf{Z}\}$, with corresponding Bernoulli parameters $\mathbf{q} := \{q_k = \Pr\{x(n) > \tau_k\}, k \in \mathbf{Z}\}$.

Since $x(n)$ cannot be at the same time smaller than τ_{k_1} and larger than τ_{k_2} for $k_1 < k_2$, $\mathbf{b}$ can only take on realizations of the form

$$\boldsymbol{\beta}_l = \{\beta_k, k \in Z | y_k = 1,\ \text{for}\ k \leq l,\ \text{and}\ y_k = 0,\ \text{for}\ k > l\}. \tag{23}$$

The event $\mathbf{b}(n) = \boldsymbol{\beta}_l$ corresponds to the event $\{x(n) \in (\tau_l, \tau_{l+1})\}$, reinforcing the fact that creating multiple binary observations is just a different way of looking at quantization.

Given these definitions, the per-sensor log-likelihood is:

$$L_n(\theta) = \sum_{k=-\infty}^{+\infty} \delta[\boldsymbol{\beta}_k - \mathbf{b}(n)] \ln[q_{k+1}(\theta) - q_k(\theta)], \tag{24}$$

where $\delta[\boldsymbol{\beta}_k - \mathbf{b}(n)] := 1$, if $\boldsymbol{\beta}_k = \mathbf{b}(n)$; and 0 otherwise. Independence across sensors implies

$$L(\theta) = \sum_{n=0}^{N-1} L_n(\theta), \tag{25}$$

and yields the MLE of θ given $\{\mathbf{b}(n)\}_{n=0}^{N-1}$ as

$$\hat{\theta} = \operatorname{argmax}_\theta\{L(\theta)\}. \tag{26}$$

Two important features of $\hat{\theta}$ in (26) are summarized next.

Proposition 4 *(a) $L(\theta)$ in (25) is a concave function of θ; and (b) The CRLB of any unbiased estimator of θ based on $\{\mathbf{b}(n)\}_{n=0}^{N-1}$ is*

$$B(\theta) = \frac{1}{N}\left[\sum_{k=-\infty}^{+\infty} \frac{[p(\tau_{k+1} - \theta) - p(\tau_k - \theta)]^2}{F(\tau_{k+1} - \theta) - F(\tau_k - \theta)}\right]^{-1}. \tag{27}$$

By asserting that $L(\theta)$ in (25) is concave, Proposition 4-(a) implies that $\hat{\theta}$ in (26) can be reliably implemented. To appreciate the value of Proposition 4-(b) notice that for an infinite set of equally spaced thresholds (with spacing $\tau := \tau_{k+1} - \tau_k$), $B(\theta)$ is periodic with period τ. Fig. 2 depicts $B(\theta)$ parameterized by τ/σ, along with the maximum and minimum values of $B(\theta)$ as functions of τ/σ. An immediate observation is that for a given τ the worst and best variances are almost equal for $\tau \leq 2\sigma$, being for all practical purposes constant when $\tau \leq \sigma$. Also important, when $\tau \leq \sigma$, $B(\theta)$ is almost equal to the clairvoyant estimator's variance.

To transmit the binary observations note that if $b_k(n) = 1$, then $b_{k'}(n) = 1$ for $k' < k$; and likewise if $b_k(n) = 0$, then $b_{k'}(n) = 0$ for $k' > k$. Loosely speaking, we can say that each binary observation transmitted provides information about half of the thresholds, and the required number of bits N_t to be transmitted per sensor should grow logarithmically with the allowable parameter range. Precisely, the actual value of N_t will depend on the parameter's range; e.g., for $\theta \in [-U, U]$, it will be $N_t \approx \log_2[(\sigma + 2U)/\tau]$. When the priori knowledge about θ dictates a Gaussian weighting function $W(\theta)$ the result can be summarized as follows.

Proposition 5 *When $W(\theta)$ is a Gaussian bell with variance σ_θ^2, the infinite set $\mathbf{b}(n)$ can be transmitted using N_t bits, such that:*

$$\mathrm{E}(N_t) < 3 + \left[\log_2 Q^{-1}(1/4) + \frac{1}{2}\log_2\left(\frac{\sigma_\theta^2 + \sigma^2}{\tau^2}\right)\right]_+ , \quad (28)$$

where $\tau := \tau_{k+1} - \tau_k \; \forall k$, $Q(x) := 1/(\sqrt{2\pi}) \int_x^{+\infty} \exp(-u^2/2)\, du$, and $[x]_+ = \max(0, x)$.

Combining Propositions 4-(b) and 5 yields a benchmark on the performance of estimators based on binary observations. For a given bandwidth constraint, we determine τ from (28), and from there the benchmark variance from (27).

Eq. (28) can be written more intuitively in terms of γ as

$$\mathrm{E}(N_t) < 2.43 + \frac{1}{2}\log_2(1 + \gamma) + \log_2\left(\frac{\sigma}{\tau}\right), \quad (29)$$

where we substituted the constants in (28) by their explicit values, and assumed for simplicity that the argument inside the $[\cdot]_+$ operator is positive (valid if $\tau^2 < 0.45(\sigma_\theta^2 + \sigma^2)$). The first logarithmic term in (29) can be viewed as quantifying the information that each observation $x(n)$ carries about the underlying parameter, while the second can be thought of as quantifying our confidence on the observations.

5.1. Optimum threshold spacing

Different from classical estimation problems were the number of measurements is given, in bandwidth-constrained problems the total number of available bits, N_b, is given. Thus, a convenient metric for a bandwidth-constrained estimation problem is the following.

Definition 1 *Suppose that for a given estimator based on binary observations, the transmission of binary observations requires an average of $\bar{N}_t$ bits. Define the per-bit worst case CRLB as:*

$$C_b = \bar{N}_t \max_\theta\{B(\theta)\}. \quad (30)$$

For a given bandwidth constraint, N_b, the variance will be bounded by $\mathrm{var}(\hat{\theta}) \geq C_b/N_b$. Applying Definition 1 to the CRLB in (27), we deduce that C_b is a function of the spacing τ

$$C_b(\tau) = \bar{N}_t(\tau) \max_\theta\{CRLB(\theta, \tau)\}, \quad (31)$$

what leads to considering the optimum threshold spacing $\tau^*(\gamma)$ as

$$\tau^* = \mathrm{argmin}_\tau\{C_b(\tau)\}. \quad (32)$$

By accounting for the bandwidth constraint, we proved the existence of an optimum quantization step τ^* and a corresponding *optimum number of bits per observation*. Fig. 3 shows $C_b(\tau)$, and $\tau^*(\gamma)$. It is apparent from these curves, that $\tau^*(\gamma)$ is quite insensitive to variations of γ. When γ varies from 0 dB to 50 dB (a 10^5 range), τ^* moves from 2σ to σ. Furthermore, the curves $C_b(\tau)$ are very flat around the optimum, implying that we can adopt $\tau = \sigma$ as a working compromise for the optimum threshold spacing (i.e., quantization step).

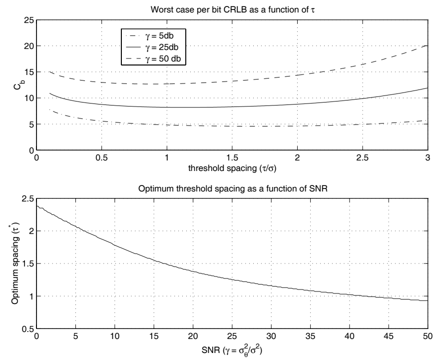

Fig. 3. Threshold spacing that minimizes worst case per bit CRLB as a function of SNR. $C_b(\tau)$ is very flat around the optimum and τ^* has a small change when the SNR moves over a range of 50 dB.

6. QUANTIZED SAMPLE MEAN ESTIMATOR

Consider the observations $\{x(n)\}_{n=0}^{N-1}$ and quantize them with a uniform quantizer at resolution τ to obtain,

$$x_Q(n) = \tau \,\mathrm{round}[x(n)/\tau], \quad (33)$$

where x_Q denotes the quantized observations and $\mathrm{round}(x)$ is the integer closest to x.

The Quantized Sample Mean Estimator (QSME) is just the sample mean of the quantized observations

$$\bar{x}_Q(n) := \frac{1}{N}\sum_{n=0}^{N-1} x_Q(n), \quad (34)$$

which is a desirable estimator if one just ignores the bandwidth constraint. Interestingly, this simple estimator is not very far from the MLE in (26) as stated in the following proposition.

Proposition 6 *The variance of the QSME in* (34) *is bounded by*

$$\mathrm{E}[(\bar{x}_Q - \theta)^2] \leq \left(1 + \frac{\tau}{\sigma} + \frac{\tau^2}{4\sigma^2}\right)\frac{\sigma^2}{N}. \quad (35)$$

Since $\bar{x}_Q$ is biased, the pertinent performance metric is the Mean Square Error (MSE), not the variance. Fig. 2 shows that the MSE of the MLE for a threshold spacing $\tau = 2\sigma$ is roughly comparable to the MSE of the QSME for a spacing $\tau = \sigma/2$; or equivalently, for comparable variances the QSME requires 2 extra bits per observation. While 2 extra bits is a rather poor solution in low SNR problems requiring transmission of a few bits, for large SNR problems the (slight) bandwidth increase is worthwhile because of the reduced complexity.

7. NUMERICAL RESULTS

We implement here the estimator introduced in Section 4. For a given threshold spacing τ, the set of frequencies $\boldsymbol{\rho}$ is obtained as the (numerical) solution of the least-squares problem in (20). Figs. 4 and 5 show the result of computing $\boldsymbol{\rho}$ for the case of Gaussian and Uniform weighting functions, respectively. In both cases, it is apparent that a threshold spacing $\tau = 2\sigma$ suffices to achieve a small MSE. This is evident in the uniform case where reducing the spacing results in nulling some of the ρ_k. Particularly interesting are the error curves depicting the difference between $[W(\theta)]^{\frac{1}{2}}$ and $S(\theta, \boldsymbol{\tau}, \boldsymbol{\rho})$. When the threshold spacing is reduced from $\tau = 2\sigma$ to $\tau = \sigma$, the error is almost unchanged. We hence deduce that *choosing the thresholds with a spacing smaller than 2σ is of no practical value*. Once the thresholds are designed, the estimation problem itself can be solved using Newton's

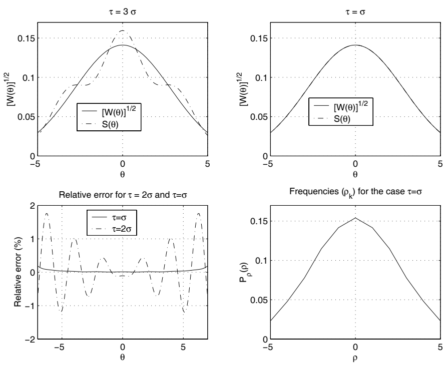

Fig. 4. Gaussian noise and Gaussian-shaped weight. Though a threshold spacing $\tau = \sigma$ reduces the approximation error to almost zero, a spacing $\tau = 2\sigma$ is good enough ($\sigma = 1$, and $\sigma_\theta = 2$).

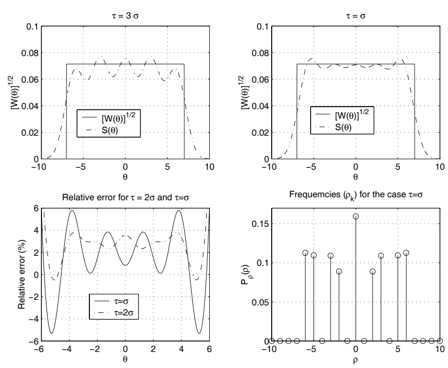

Fig. 5. Gaussian noise and Uniform weight function. A threshold spacing $\tau = \sigma$ has smaller MSE but a spacing $\tau = 2\sigma$ is better in most of the non-zero probability interval ($\sigma = 1$, and prior U[-7,7]).

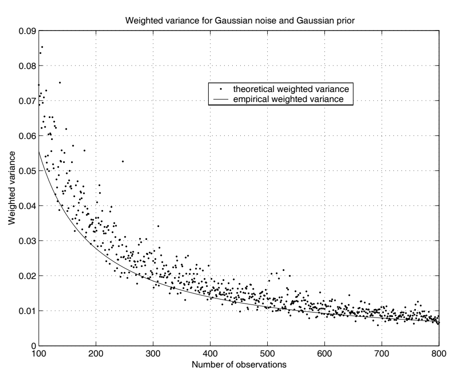

Fig. 6. Gaussian noise and Gaussian weight function. With a threshold spacing $\tau = 2\sigma$ we achieve a good approximation to the minimum asymptotic average variance ($\sigma = 1$, $\tau = 2$, and $\sigma_\theta = 2$).

algorithm, based on the iteration (12). The results are shown on Fig. 6 for the case of a Gaussian weight function with 2σ spacing between thresholds. For each value of N, the experiment is repeated 200 times and the average variance is plotted against the theoretical threshold, which appears to reasonably predict its value. It also confirms that a threshold spacing $\tau = 2\sigma$, is good enough.

8. CONCLUDING REMARKS

We were motivated by the observation that an estimator based on the transmission of a single binary observation per sensor can have variance as small as $\pi/2$ times that of the clairvoyant sample mean estimator (Section 3). By noting that this excellent performance can only be achieved under careful design choices, we introduced a class of estimators establishing our first major result: *in the low-to-medium SNR range, this class of MLE performs close to the clairvoyant estimator's variance* (Section 3). We then tackled high SNR problems, and showed that a quantization step roughly equal to the noise's standard deviation is nearly optimal in the sense of minimizing a properly defined per-bit CRLB (Section 5), establishing our second major conclusion, on the *optimal number of bits per sensor to be transmitted*. The quantized sample mean estimator was introduced showing that at high SNR even a simple-minded estimator requires transmission of only a small number of extra bits than the MLE. This allowed us to establish analytically that *bandwidth-constrained distributed estimation is not a relevant problem in high SNR scenarios*. For such cases, we advocate using the sample mean estimator based on the quantized observations for its low complexity (Section 6). The last major conclusion of the present paper is that *numerical maximization required by our MLE can be posed as a convex optimization problem*, thus ensuring convergence by e.g., Newton-type iterative algorithms.

9. REFERENCES

[1] D. Blatt and A. Hero, "Distributed maximum likelihood estimation for sensor networks," in *Proc. of the ICASSP*, vol. 3, pp. 929–932, Montreal, Canada, May 2004.

[2] S. Kumar, F. Zao, and D. Shepherd, eds., *Special issue on collaborative information processing*, vol. 19 of *IEEE Signal Processing Magazine*, March 2002.

[3] W. Lam and A. Reibman, "Quantizer design for decentralized systems with communication constraints," *IEEE Transactions on Communications*, vol. 41, pp. 1602–1605, Aug. 1993.

[4] Z.-Q. Luo, "An isotropic universal decentralized estimation scheme for a bandwidth constrained ad hoc sensor network," *IEEE Journal on Selected Areas in Communications*, June 2003 (submitted). Available at http://www.ece.umn.edu/users/luozq/recent_work.html.

[5] H. Papadopoulos, G. Wornell, and A. Oppenheim, "Sequential signal encoding from noisy measurements using quantizers with dynamic bias control," *IEEE Transactions on Information Theory*, vol. 47, pp. 978–1002, 2001.

[6] M. G. Rabbat and R. D. Nowak, "Decentralized source localization and tracking," in *Proc. of the ICASSP*, vol. 3, pp. 921–924, Montreal, Canada, May 2004.

[7] A. Ribeiro and G. B. Giannakis, "Bandwidth-Constrained Distributed Estimation for Wireless Sensor Networks, Part I: Gaussian Case," *IEEE Transactions on Signal Processing*, submitted July 2004. Available at http://www.ece.umn.edu/users/aribeiro/Research/Pubs.html.

[8] P. K. Varshney, *Distributed Detection and Data Fusion*. Springer-Verlag, 1997.