Bandwidth-Constrained Distributed Estimation for Wireless Sensor Networks—Part II: Unknown Probability Density Function

Alejandro Ribeiro, *Student Member, IEEE,* and Georgios B. Giannakis, *Fellow, IEEE*

***Abstract*—Wireless sensor networks (WSNs) deployed to perform surveillance and monitoring tasks have to operate under stringent energy and bandwidth limitations. These motivate well distributed estimation scenarios where sensors quantize and transmit only one, or a few bits per observation, for use in forming parameter estimators of interest. In a companion paper, we developed algorithms and studied interesting tradeoffs that emerge even in the simplest distributed setup of estimating a scalar location parameter in the presence of zero-mean additive white Gaussian noise of known variance. Herein, we derive distributed estimators based on binary observations along with their fundamental error-variance limits for more pragmatic signal models: i) known univariate but generally non-Gaussian noise probability density functions (pdfs); ii) known noise pdfs with a finite number of unknown parameters; iii) completely unknown noise pdfs; and iv) practical generalizations to multivariate and possibly correlated pdfs. Estimators utilizing either independent or colored binary observations are developed and analyzed. Corroborating simulations present comparisons with the clairvoyant sample-mean estimator based on unquantized sensor observations, and include a motivating application entailing distributed parameter estimation where a WSN is used for habitat monitoring.**

***Index Terms*—Distributed parameter estimation, wireless sensor networks (WSNs).**

I. Introduction

Wireless sensor networks (WSNs) consist of low-cost energy-limited transceiver nodes spatially deployed in large numbers to accomplish monitoring, surveillance and control tasks through cooperative actions [10]. The potential of WSNs for surveillance has by now been well appreciated especially in the context of data fusion and distributed detection; e.g., [24], [25], and references therein. However, except, e.g., for recent works where spatial correlation is exploited to reduce the amount of information exchanged among nodes [2], [3], [6], [7], [11], [16], [19], [20], use of WSNs for the equally important problem of distributed parameter estimation remains a largely uncharted territory. When sensors have to quantize measurements in order to save energy and bandwidth, estimators based on quantized samples and pertinent tradeoffs have been studied for relatively simple models. Specifically, quantizer designs for mean-location scalar parameter estimation in additive noise of known distribution were studied in [1], [17], and [18], while one bit per sensor quantization in noise of unknown distribution was dealt with in [12]–[14]. In the present paper, we consider estimation based on a single bit per sensor for a number of pragmatic signal models. It is worth stressing that in these contributions as well as in the present work that deals with WSN-based distributed parameter acquisition under bandwidth constraints, the notions of quantization and estimation are intertwined. In fact, quantization becomes an integral part of estimation as it creates a set of *binary observations* based on which the estimator must be formed—a problem distinct from parameter estimation based on the unquantized observations.

In a companion paper we study estimation of a scalar mean-location parameter in the presence of zero-mean additive white Gaussian noise [23]. For this simple model, we define the so called quantization signal-to-noise ratio (Q-SNR) as the ratio of the parameter's dynamic range over the noise standard deviation, and advocated different strategies depending on whether the Q-SNR is low, medium or high. An interesting conclusion from [23] is that in low-medium Q-SNR, estimation based on sign quantization of the original observations exhibits variance almost equal to the variance of the (clairvoyant) estimator based on unquantized observations. Interestingly, for the pragmatic class of models considered here it is still true that transmitting a few bits (or even a single bit) per sensor can approach under realistic conditions the performance of the estimator based on unquantized data. The impact of the latter to WSNs is twofold. On the one hand, we effect energy savings by transmitting a single bit per sensor; and on the other hand, we simplify analog to digital conversion to (inexpensive) signal level comparation. While results in the present paper apply only when the Q-SNR is low-to-medium this is rather typical for WSNs.

We begin with mean-location parameter estimation in the presence of known univariate but generally non-Gaussian noise probability density functions (pdfs) (Section III-A). We next develop mean-location parameter estimators based on binary observations and benchmark their performance when the noise variance is unknown; however, the same approach in principle applies to any noise pdf that is known except for a finite number of unknown parameters (Section III-B). Subsequently, we move to the most challenging case where the noise pdf is

Manuscript received August 13, 2004; revised April 8, 2005. A portion of the results in this paper appeared in [21] and [22]. This work was supported by 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. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Yucel Altunbasak.

The authors are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: aribeiro@ece.umn.edu; georgios@ece.umn.edu).

Digital Object Identifier 10.1109/TSP.2006.874366

completely unknown (Section IV). Finally, we consider vector generalizations where each sensor observes a given (possibly nonlinear) function of the unknown parameter vector in the presence of multivariate and possibly colored noise (Section V). While challenging in general, it will turn out that under relaxed conditions, the resultant maximum likelihood estimator (MLE) is the maximum of a concave function, thus ensuring convergence of Newton-type iterative algorithms. Moreover, in the presence of colored Gaussian noise, we show that judiciously quantizing each sensor's data renders the estimators' variance stunningly close to the variance of the clairvoyant estimator that is based on the unquantized observations; thus, nicely generalizing the results of Sections III-A, III-B, and [23] to the more realistic vector parameter estimation problem (Section V-A). Numerical examples corroborate our theoretical findings in Section VI, where we also test them on a motivating application involving distributed parameter estimation with a WSN for measuring vector flow (Section VI-B). We conclude the paper in Section VII.

II. Problem Statement

Consider a WSN consisting of N sensors deployed to estimate a deterministic $p \times 1$ vector parameter $\boldsymbol{\theta}$. The nth sensor observes an $M \times 1$ vector of noisy observations

$$\mathbf{x}(n) = \mathbf{f}_n(\boldsymbol{\theta}) + \mathbf{w}(n), \quad n = 0, 1, \ldots, N-1 \tag{1}$$

where $\mathbf{f}_n : \mathbf{R}^p \rightarrow \mathbf{R}^M$ is a known (generally nonlinear) function and $\mathbf{w}(n)$ denotes zero-mean noise with pdf $p_{\mathbf{w}}(\mathbf{w})$, that is either unknown or known possibly up to a finite number of unknown parameters. We further assume that $\mathbf{w}(n_1)$ is independent of $\mathbf{w}(n_2)$ for $n_1 \neq n_2$; i.e., noise variables are independent across sensors. We will use $\mathbf{J}_n$ to denote the Jacobian of the differentiable function $\mathbf{f}_n$ whose (i,j)th entry is given by $[\mathbf{J}_n]_{ij} = \partial[\mathbf{f}_n]_i / \partial[\boldsymbol{\theta}]_j$.

Due to bandwidth limitations, the observations $\mathbf{x}(n)$ have to be quantized and estimation of $\boldsymbol{\theta}$ can only be based on these quantized values. We will, henceforth, think of quantization as the construction of a set of indicator variables

$$b_k(n) = \mathbf{1}\{\mathbf{x}(n) \in B_k(n)\}, \quad k = 1, \ldots, K \tag{2}$$

taking the value 1 when $\mathbf{x}(n)$ belongs to the region $B_k(n) \subset \mathbf{R}^M$, and 0 otherwise. Throughout, we suppose that the regions $B_k(n)$ are computed at the fusion center where resources are not at a premium.

Estimation of $\boldsymbol{\theta}$ will rely on the set of *binary* variables $\{b_k(n), k = 1, \ldots, K\}_{n=0}^{N-1}$. The latter are Bernoulli distributed with parameters $q_k(n)$ satisfying

$$q_k(n) := \Pr\{b_k(n) = 1\} = \Pr\{\mathbf{x}(n) \in B_k(n)\}. \tag{3}$$

In the ensuing sections, we will derive the Cramér-Rao Lower Bound (CRLB) to benchmark the variance of all unbiased estimators $\hat{\boldsymbol{\theta}}$ constructed using the binary observations $\{b_k(n), k = 1, \ldots, K\}_{n=0}^{N-1}$. We will further show that it is possible to find MLEs that (at least asymptotically) are known to achieve the CRLB. Finally, we will reveal that the CRLB based on $\{b_k(n), k = 1, \ldots, K\}_{n=0}^{N-1}$ can come surprisingly close to the clairvoyant CRLB based on $\{x(n)\}_{n=0}^{N-1}$ in certain applications of practical interest.

III. Scalar Parameter Estimation—Parametric Approach

Consider the case where $\boldsymbol{\theta} \leftrightarrow \theta$ is a scalar $(p = 1), x(n) = \theta + w(n)$, and $p_w(w) \leftrightarrow p_w(w, \sigma)$ is known, with σ denoting the noise standard deviation. Seeking first estimators $\hat{\theta}$ when the possibly non-Gaussian noise pdf is known, we move on to the case where σ is unknown, and prove that in both cases the variance of $\hat{\theta}$ based on a single bit per sensor can come close to the variance of the sample mean estimator, $\bar{x} := N^{-1} \sum_{n=0}^{N-1} x(n)$.

A. Known Noise Pdf

When the noise pdf is known, we will rely on a single region $B_1(n)$ in (2) to generate a single bit $b_1(n)$ per sensor, using a threshold τ_c common to all N sensors: $B_1(n) := B_c = (\tau_c, \infty), \forall n$. Based on these binary observations, $b_1(n) := \mathbf{1}\{\mathbf{x}(n) \in (\tau_c, \infty)\}$ received from all N sensors, the fusion center seeks estimates of θ.

Let $F_w(u) := \int_u^\infty p_w(w)\, dw$ denote the Complementary Cumulative Distribution Function (CCDF) of the noise. Using (3), we can express the Bernoulli parameter as, $q_1 = \int_{\tau_c - \theta}^\infty p_w(w)\, dw = F_w(\tau_c - \theta)$; and its MLE as $\hat{q}_1 = N^{-1} \sum_{n=0}^{N-1} b_1(n)$. Invoking now the invariance property of MLE, it follows readily that the MLE of θ is given by [23][1]

$$\hat{\theta} = \tau_c - F_w^{-1}\left(\frac{1}{N}\sum_{n=0}^{N-1} b_1(n)\right). \tag{4}$$

Furthermore, it can be shown that the CRLB, that bounds the variance of any unbiased estimator $\hat{\theta}$ based on $b_1(n)_{n=0}^{N-1}$ is [23]

$$\text{var}(\hat{\theta}) \geq \frac{1}{N} \frac{F_w(\tau_c - \theta)[1 - F_w(\tau_c - \theta)]}{p_w^2(\tau_c - \theta)} := B(\theta). \tag{5}$$

If the noise is Gaussian, and we define the σ-*distance* between the threshold τ_c and the (unknown) parameter θ as $\Delta_c := (\tau_c - \theta)/\sigma$, then (5) reduces to

$$B(\theta) = \frac{\sigma^2}{N} \frac{2\pi Q(\Delta_c)[1 - Q(\Delta_c)]}{e^{-\Delta_c}} := \frac{\sigma^2}{N} D(\Delta_c) \tag{6}$$

with $Q(u) := (1/\sqrt{2\pi}) \int_u^\infty e^{-w^2/2}\, dw$ denoting the Gaussian tail probability function.

The bound $B(\theta)$ is the variance of $\bar{x}$, scaled by the factor $D(\Delta_c)$; recall that $\text{var}(\bar{x}) = \sigma^2/N$ ([8, p. 31]). Optimizing $B(\theta)$ with respect to Δ_c, yields the optimum at $\Delta_c = 0$ and the minimum CRLB as

$$B_{\min} = \frac{\pi}{2} \frac{\sigma^2}{N}. \tag{7}$$

Equation (7) reveals something unexpected: relying on a single bit per $x(n)$, the estimator in (4) incurs a minimal (just a $\pi/2$ factor) increase in its variance relative to the clairvoyant $\bar{x}$ which

[1] Although related results are derived in ([23], Prop. 1) for Gaussian noise, it is straightforward to generalize the referred proof to cover also non-Gaussian noise pdfs.

relies on the unquantized data $x(n)$. But this minimal loss in performance corresponds to the ideal choice $\Delta_c = 0$, which implies $\tau_c = \theta$ and requires perfect knowledge of the unknown θ for selecting the quantization threshold τ_c. How do we select τ_c and how much do we loose when the unknown θ lies anywhere in $(-\infty, \infty)$, or when θ lies in $[\Theta_1, \Theta_2]$, with Θ_1, Θ_2 finite and known a priori? Intuition suggests selecting the threshold as close as possible to the parameter. This can be realized with an iterative estimator $\hat{\theta}^{(i)}$, which can be formed as in (4), using $\tau_c^{(i)} = \hat{\theta}^{(i-1)}$, the parameter estimate from the previous $(i-1)^{\text{st}}$ iteration.

But in the batch formulation considered herein, selecting τ_c is challenging; and a closer look at $B(\theta)$ in (5) will confirm that the loss can be huge if $\tau_c - \theta \gg 0$. Indeed, as $\tau_c - \theta \to \infty$ the denominator in (5) goes to zero faster than its numerator, since F_w is the integral of the nonnegative pdf p_w; and thus, $B(\theta) \to \infty$ as $\tau_c - \theta \to \infty$. The implication of the latter is twofold: i) since it shows up in the CRLB, the potentially high variance of estimators based on quantized observations is inherent to the possibly severe bandwidth limitations of the problem itself and is not unique to a particular estimator; ii) for any choice of τ_c, the fundamental performance limits in (5) are dictated by the end points $\tau_c - \Theta_1$ and $\tau_c - \Theta_2$ when θ is confined to the interval $[\Theta_1, \Theta_2]$. On the other hand, how successful the τ_c selection is depends on the dynamic range $|\Theta_1 - \Theta_2|$ which makes sense because the latter affects the error incurred when quantizing $x(n)$ to $b_1(n)$. Notice that in such joint quantization-estimation problems one faces two sources of error: quantization and noise. To account for both, the proper figure of merit for estimators based on binary observations is what we will term quantization signal-to-noise ratio (Q-SNR) that we define as[2]

$$\gamma := \frac{|\Theta_1 - \Theta_2|^2}{\sigma^2}. \tag{8}$$

Notice that contrary to common wisdom, the smaller Q-SNR is, the easier it becomes to select τ_c judiciously. Furthermore, the variance increase in (5) relative to the variance of the clairvoyant $\bar{x}$ is smaller, for a given σ. This is because as the Q-SNR increases the problem becomes more difficult in general, but the rate at which the variance increases is smaller for the CRLB in (5) than for $\text{var}(\bar{x}) = \sigma^2/N$.

However, no matter how small the variance in (5) can be made by properly selecting τ_c, the estimator $\hat{\theta}$ in (4) requires perfect knowledge of the noise pdf which may not be always justifiable. For example, while assuming that the noise is Gaussian (or follows a known non-Gaussian pdf that accurately fits the problem) is reasonable, assuming that its variance (or any other parameter of the pdf) is known, is not. The search for estimators in more realistic scenarios motivates the next subsection.

B. *Known Noise pdf With Unknown Variance*

A more realistic approach is to assume that the noise pdf is known (e.g., Gaussian) but some of its parameters are unknown. A case frequently encountered in practice is when the noise pdf is known except for its variance $\text{E}[w^2(n)] = \sigma^2$. Introducing the standardized variable $v(n) := w(n)/\sigma$ allows us to write the signal model as

$$x(n) = \theta + \sigma v(n). \tag{9}$$

Let $p_v(v)$ and $F_v(v) := \int_v^\infty p_v(u)\,du$ denote the known pdf and CCDF of $v(n)$. Note that according to its definition, $v(n)$ has zero mean, $\text{E}[v^2(n)] = 1$, and the pdfs of v and w are related by $p_w(w) = (1/\sigma)p_v(w/\sigma)$. Note also that all two parameter pdfs can be standardized likewise. This is even true for a broad class of three-parameter pdfs provided that one parameter is known. Consider as a typical example the generalized Gaussian class of pdfs ([9, p. 384])

$$p_w(w) = \frac{\beta c(\beta)}{2\sigma\Gamma(1/\beta)} e^{-c(\beta)\left|\frac{x}{\sigma}\right|^\beta}, \quad c(\beta) := \left[\frac{\Gamma(3/\beta)}{\Gamma(1/\beta)}\right]^{1/2} \tag{10}$$

with the gamma function defined as $\Gamma(x) := \int_0^\infty t^{x-1} e^t\,dt$ and β a known constant. In this case too, $v(n) = w(n)/\sigma$ has unit variance and (9) applies.

To estimate θ when σ is also unknown while keeping the bandwidth constraint to 1 bit per sensor, we divide the sensors in two groups each using a different region (i.e., threshold) to define the binary observations

$$B_1(n) := \begin{cases} (\tau_1, \infty) := B_1 & \text{for } n = 0, \ldots, (N/2) - 1 \\ (\tau_2, \infty) := B_2 & \text{for } n = (N/2), \ldots, N. \end{cases} \tag{11}$$

That is, the first $N/2$ sensors quantize their observations using the threshold τ_1, while the remaining $N/2$ sensors rely on the threshold τ_2. Without loss of generality, we assume $\tau_2 > \tau_1$.

The Bernoulli parameters of the resultant binary observations can be expressed in terms of the CCDF of $v(n)$ as

$$q_1(n) := \begin{cases} F_v\left[\frac{\tau_1 - \theta}{\sigma}\right] := q_1 & \text{for } n = 0, \ldots, (N/2) - 1 \\ F_v\left[\frac{\tau_2 - \theta}{\sigma}\right] := q_2 & \text{for } n = (N/2), \ldots, N. \end{cases} \tag{12}$$

Given the noise independence across sensors, the MLEs of q_1, q_2 can be found, respectively, as

$$\hat{q}_1 = \frac{2}{N} \sum_{n=0}^{N/2-1} b_1(n), \quad \hat{q}_2 = \frac{2}{N} \sum_{n=N/2}^{N-1} b_1(n). \tag{13}$$

Mimicking (4), we can invert F_v in (12) and invoke the invariance property of MLEs, to obtain the MLE $\hat{\theta}$ in terms of $\hat{q}_1$ and $\hat{q}_2$. This result is stated in the following proposition that also derives the CRLB for this estimation problem.

Proposition 1: Consider estimating θ in (9), when σ is unknown, based on binary observations constructed from the regions defined in (11).

a) The MLE of θ is

$$\hat{\theta} = \frac{F_v{}^{-1}(\hat{q}_2)\tau_1 - F_v{}^{-1}(\hat{q}_1)\tau_2}{F_v{}^{-1}(\hat{q}_2) - F_v{}^{-1}(\hat{q}_1)} \tag{14}$$

[2]Attaching to γ the notion of SNR is justified if we consider θ as random uniformly distributed over $[\Theta_1, \Theta_2]$, in which case the numerator of γ is proportional to the signal's mean square value $\text{E}(\theta^2)$. Likewise, we can view the numerator $[\Theta_1, \Theta_2]^2$ as the root mean-square (rms) value of θ in the deterministic treatment herein.

with ${F_v}^{-1}$ denoting the inverse function of F_v, and $\hat{q}_1, \hat{q}_2$ given by (13).

b) The variance of any unbiased estimator of θ based on $\{b_1(n)\}_{n=0}^{N-1}$ is bounded by

$$\mathrm{var}(\hat{\theta}) \geq \frac{2\sigma^2}{N}\left(\frac{\Delta_1\Delta_2}{\Delta_2-\Delta_1}\right)^2\left[\frac{q_1(1-q_1)}{p_v^2(\Delta_1)\Delta_1^2}+\frac{q_2(1-q_2)}{p_v^2(\Delta_2)\Delta_2^2}\right] := B(\theta) \quad (15)$$

where q_k is given by (12), and

$$\Delta_k := \frac{\tau_k-\theta}{\sigma}, \quad k=1,2 \quad (16)$$

is the σ-distance between θ and the threshold τ_k.

Proof: Using (12), we can express θ in terms of $\mathbf{q} := (q_1, q_2)$, as

$$\theta = \frac{{F_v}^{-1}(q_2)\tau_1 - {F_v}^{-1}(q_1)\tau_2}{{F_v}^{-1}(q_2) - {F_v}^{-1}(q_1)}. \quad (17)$$

Since the MLEs of q_k are available from (13), just recall the invariance property of MLE and replace q_k by $\hat{q}_k$ to arrive at (14).

To prove claim (b), note that because of the noise independence the Fisher Information Matrix (FIM) for the estimation of $\mathbf{q}$ is diagonal and its inverse is given by

$$\mathbf{I}^{-1}(\mathbf{q}) = \begin{pmatrix} \frac{q_1(1-q_1)}{N/2} & 0 \\ 0 & \frac{q_2(1-q_2)}{N/2} \end{pmatrix}. \quad (18)$$

Applying known CRLB expressions for transformations of estimators, we can obtain the CRLB of $\hat{\theta}$ as ([8, p. 45])

$$\mathrm{var}(\hat{\theta}) \geq \left(\frac{\partial\theta}{\partial q_1}, \frac{\partial\theta}{\partial q_2}\right)\mathbf{I}^{-1}(\mathbf{q})\left(\frac{\partial\theta}{\partial q_1}, \frac{\partial\theta}{\partial q_2}\right)^T := B(\theta) \quad (19)$$

where the derivatives involved can be obtained from (17), and are given by

$$\frac{\partial\theta}{\partial q_k} = (-1)^{k-1}\frac{\sigma}{\Delta_k p_v(\Delta_k)}\frac{\Delta_1\Delta_2}{\Delta_2-\Delta_1}, \quad k=1,2. \quad (20)$$

Expanding the quadratic form in (19), and substituting the derivatives for the expressions in (20), the CRLB in (15) follows. □

Equation (15) is reminiscent of (5), suggesting that the variances of the estimators they bound are related. This implies that even when the known noise pdf contains unknown parameters the variance of $\hat{\theta}$ can come close to the variance of the clairvoyant estimator $\bar{x}$, provided that the thresholds τ_1, τ_2 are chosen close to θ relative to the noise standard deviation (so that Δ_1, Δ_2, and $\Delta_2 - \Delta_1$ in (16) are ≈ 1). For the Gaussian pdf, Fig. 1 shows the contour plot of $B(\theta)$ in (15) normalized by $\sigma^2/N := \mathrm{var}(\bar{x})$. It is easy to see that for $\theta \in [\Theta_1, \Theta_2]$, the worst case variance is minimized by setting $\tau_1 \approx \Theta_1$ and $\tau_2 \approx \Theta_2$. With this selection in the low Q-SNR regime $\Delta_1, \Delta_2 \approx 1$, and the relative variance increase $B(\theta)/\mathrm{var}(\bar{x})$ is less than 3.

C. Dependent Binary Observations

As aforementioned, we restricted the sensors to transmit only 1 bit (binary observation) per $x(n)$ datum, and divided the sensors in two classes each quantizing $x(n)$ using a different threshold. A related approach is to let each sensor use two thresholds, thus providing information as to whether $x(n)$ falls in two different regions

$$B_1(n) := B_1 = (\tau_1, \infty), \quad n = 0, 1, \ldots, N-1$$
$$B_2(n) := B_2 = (\tau_2, \infty), \quad n = 0, 1, \ldots, N-1 \quad (21)$$

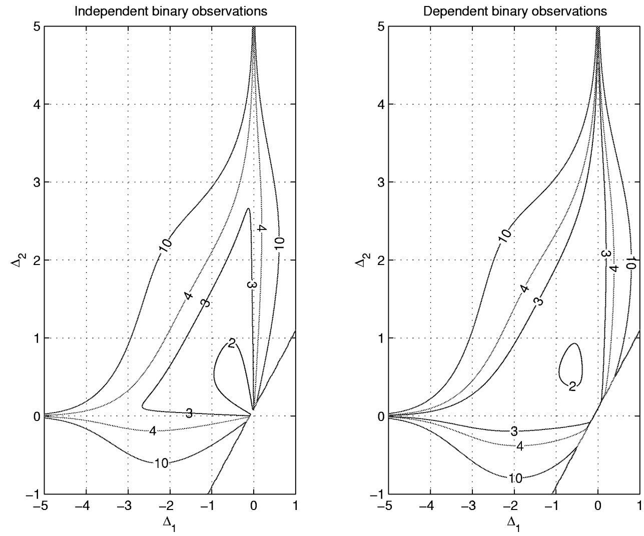

Fig. 1. Per bit CRLB when the binary observations are independent (Section III-B) and dependent (Section III-C), respectively. In both cases, the variance increase with respect to the sample mean estimator is small when the σ-distances are close to 1, being slightly better for the case of dependent binary observations (Gaussian noise).

where $\tau_2 > \tau_1$. We define the per sensor vector of binary observations $\mathbf{b}(n) := [b_1(n), b_2(n)]^T$, and the vector Bernoulli parameter $\mathbf{q} := [q_1(n), q_2(n)]^T$, whose components are as in (12). Surprisingly, estimation performance based on these dependent observations will turn out to improve that of independent observations.

Note the subtle differences between (11) and (21). While each of the N sensors generates 1 binary observation according to (11), each sensor creates 2 binary observations as per (21). The total number of bits from all sensors in the former case is N, but in the latter $N \log_2 3$, since our constraint $\tau_2 > \tau_1$ implies that the realization $\mathbf{b} = (0, 1)$ is impossible. In addition, all bits in the former case are independent, whereas correlation is present in the latter since $b_1(n)$ and $b_2(n)$ come from the same $x(n)$. Even though one would expect this correlation to complicate matters, a property of the binary observations defined as per (21), summarized in the next lemma, renders estimation of θ based on them feasible.

Lemma 1: The MLE of $\mathbf{q} := (q_1(n), q_2(n))^T$ based on the binary observations $\{\mathbf{b}(n)\}_{n=0}^{N-1}$ constructed according to (21) is given by

$$\hat{\mathbf{q}} = \frac{1}{N}\sum_{n=0}^{N-1}\mathbf{b}(n). \quad (22)$$

Proof: See Appendix A.1.

Interestingly, (22) coincides with (13), proving that the corresponding estimators of θ are identical; i.e., (14) yields also the

MLE $\hat{\theta}$ even in the correlated case. However, as the following proposition asserts, correlation affects the estimator's variance and the corresponding CRLB.

Proposition 2: Consider estimating θ in (9), when σ is unknown, based on binary observations constructed from the regions defined in (21). The variance of any unbiased estimator of θ based on $\{b_1(n), b_2(n)\}_{n=0}^{N-1}$ is bounded by

$$\text{var}(\hat{\theta}) \geq B_D(\theta) := \frac{\sigma^2}{N}\left(\frac{\Delta_1\Delta_2}{\Delta_2 - \Delta_1}\right)^2 \times \left[\frac{q_1(1-q_1)}{p_v^2(\Delta_1)\Delta_1^2} + \frac{q_2(1-q_2)}{p_v^2(\Delta_2)\Delta_2^2} - \frac{q_2(1-q_1)}{p_v(\Delta_1)p(\Delta_2)\Delta_1\Delta_2}\right] \quad (23)$$

where the subscript D in $B_D(\theta)$ is used as a mnemonic for the dependent binary observations this estimator relies on [cf. (15)].

Proof: See Appendix A.2.

Unexpectedly, (23) is similar to (15). Actually, a fair comparison between the two requires compensating for the difference in the total number of bits used in each case. This can be accomplished by introducing the per-bit CRLBs for the independent and correlated cases, respectively

$$C(\theta) = NB(\theta), \quad C_D(\theta) = N\log_2(3)B_D(\theta) \quad (24)$$

which lower bound the corresponding variances achievable by the transmission of 1 bit.

Evaluation of $C(\theta)/\sigma^2$ and $C_D(\theta)/\sigma^2$ follows from (15), (23) and (24) and is depicted in Fig. 1 for Gaussian noise and σ-distances Δ_1, Δ_2 having amplitude as large as 5. Somewhat surprisingly, both approaches yield very similar bounds with the one relying on dependent binary observations being slightly better in the achievable variance; or correspondingly, in requiring a smaller number of sensors to achieve the same CRLB.

IV. Scalar Parameter Estimation—Unknown Noise pdf

When the noise pdf is known, we estimated θ by setting up a common region $B_1(n) := B_c$ for the N sensors to obtain their binary observations; for one unknown (θ) we required one region. For a known pdf with unknown variance, we set up two regions B_1, B_2 and had either half of the sensors use B_1 to construct their binary observations and the other half use B_2; or, let each sensor transmit two binary observations. In either case, for two unknowns (θ and σ) we utilized two regions.

Proceeding similarly, we can keep relaxing the required knowledge about the noise pdf by setting up additional regions to obtain similar $\hat{\theta}$ estimators in the presence of noise with known pdf, but with a finite number of unknown parameters. Instead of this more or less straightforward parametric extension, we will pursue in this section a *nonparametric* approach in order to address the more challenging extreme case where the pdf is completely unknown, except obviously for its mean that will be assumed to be zero so that θ in (9) is identifiable.

To this end, let $p_x(x)$ and $F_x(x)$ denote the pdf and CCDF of the observations $x(n)$. As θ is the mean of $x(n)$, we can write

$$\theta := \int_{-\infty}^{+\infty} x p_x(x)\,dx = -\int_{-\infty}^{+\infty} x \frac{\partial F_x(x)}{\partial x}\,dx$$

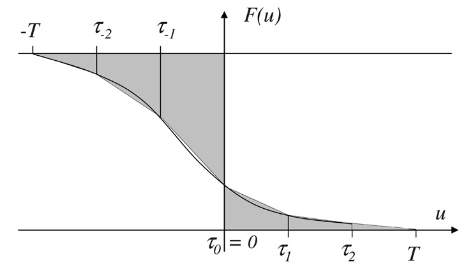

Fig. 2. When the noise pdf is unknown numerically integrating the CCDF using the trapezoidal rule yields an approximation of the mean.

$$= \int_0^1 F_x^{-1}(v)\,dv \quad (25)$$

where in establishing the second equality we used the fact that the pdf is the negative derivative of the CCDF, and in the last equality we introduced the change of variables $v = F_x(x)$. But note that the integral of the inverse CCDF can be written in terms of the integral of the CCDF as (see also Fig. 2)

$$\theta = -\int_{-\infty}^{0} [1 - F_x(u)]\,du + \int_0^{+\infty} F_x(u)\,du \quad (26)$$

allowing one to express the mean θ of $x(n)$ in terms of its CCDF. To avoid carrying out integrals with infinite range, let us assume that $x(n) \in (-T, T)$ which is always practically satisfied for T sufficiently large, so that we can rewrite (26) as

$$\theta = \int_{-T}^{T} F_x(u)\,du - T. \quad (27)$$

Numerical evaluation of the integral in (27) can be performed using a number of known techniques. Let us consider an ordered set of interior points $\{\tau_k\}_{k=1}^{K}$ along with end-points $\tau_0 = -T$ and $\tau_{K+1} = T$. Relying on the fact that $F_x(\tau_0) = F_x(-T) = 1$ and $F_x(\tau_{K+1}) = F_x(T) = 0$, application of the trapezoidal rule for numerical integration yields (see also Fig. 2)

$$\theta = \frac{1}{2}\sum_{k=1}^{K}(\tau_{k+1} - \tau_{k-1})F_x(\tau_k) - T + e_a \quad (28)$$

with e_a denoting the approximation error. Certainly, other methods like Simpson's rule, or the broader class of Newton-Cotes formulas, can be used to further reduce e_a.

Whichever the choice, the key is that binary observations constructed from the region $B_k := (\tau_k, \infty)$ have Bernoulli parameters q_k satisfying

$$q_k := \Pr\{x(n) > \tau_k\} = F_x(\tau_k). \quad (29)$$

Inserting the nonparametric estimators $\hat{F}_x(\tau_k) = \hat{q}_k$ in (28), our parameter estimator when the noise pdf is unknown takes the form

$$\hat{\theta} = \frac{1}{2}\sum_{k=1}^{K}\hat{q}_k(\tau_{k+1} - \tau_{k-1}) - T. \quad (30)$$

Since $\hat{q}_k$s are unbiased, (28) and (30) imply that $\mathrm{E}(\hat{\theta}) = \theta + e_a$. Being biased, the proper performance indicator for $\hat{\theta}$ in (30) is the mean squared error (mse), not the variance. In order to evaluate this mse let us, as we did in Section III-C, consider the cases of independent and dependent binary observations.

A. Independent Binary Observations

Divide the N sensors in K subgroups containing N/K sensors each, and define the regions[3]

$$B_1(n) := B_k = (\tau_k, \infty),\quad n = (k-1)(N/K), \ldots, k(N/K) - 1 \quad (31)$$

the region $B_1(n)$ will be used by sensor n to construct and transmit the binary observation $b_1(n)$. Herein, the unbiased estimators of the Bernoulli parameters q_k are

$$\hat{q}_k = \frac{1}{(N/K)} \sum_{n=(k-1)(N/K)}^{k(N/K)-1} b_1(n), \quad k = 1, \ldots, K \quad (32)$$

and are used in (30) to estimate θ. It is easy to verify that $\text{var}(\hat{q}_k) = q_k(1-q_k)/(N/K)$, and that $\hat{q}_{k_1}$ and $\hat{q}_{k_2}$ are independent for $k_1 \neq k_2$.

The resultant MSE, $\text{E}[(\theta - \hat{\theta})^2]$, will be bounded as stated in the following proposition.

Proposition 3: Consider the estimator $\hat{\theta}$ in (30), with $\hat{q}_k$ given by (32). Assume that for T sufficiently large and known $p_x(x) = 0$, for $|x| \geq T$; and that the noise pdf has bounded derivative $\dot{p}_w(u) := \partial p_w(w)/\partial w$, and define $\tau_{\max} := \max_k\{\tau_{k+1} - \tau_k\}$ and $\dot{p}_{\max} := \max_{u \in (-T,T)}\{\dot{p}_w(u)\}$. The mse is given by

$$\text{E}[(\theta - \hat{\theta})^2] = |e_a|^2 + \text{var}(\hat{\theta}) \quad (33)$$

with the approximation error e_a and $\text{var}(\hat{\theta})$, satisfying

$$|e_a| \leq \frac{T\dot{p}_{\max}}{6}\tau_{\max}^2 \quad (34)$$

$$\text{var}(\hat{\theta}) = \sum_{k=1}^{K} \frac{(\tau_{k+1} - \tau_{k-1})^2}{4} \frac{q_k(1-q_k)}{N/K} \quad (35)$$

where $\{\tau_k\}_{k=1}^K$ is a grid of thresholds in $(-T, T)$ and $\{q_k\}_{k=1}^K$ as in (29).

Proof: Since the estimators $\hat{q}_k$ are unbiased and $\hat{\theta}$ is linear in each $\hat{q}_k$, it follows from (30) that $\text{E}(\hat{\theta}) = \theta + e_a$. Thus, we can write

$$\text{E}[(\theta - \hat{\theta})^2] = |\theta - \text{E}(\hat{\theta})|^2 + \text{var}(\hat{\theta}) = |e_a|^2 + \text{var}(\hat{\theta}) \quad (36)$$

which expresses the mse in terms of the numerical integration error and the estimator variance.

To bound e_a, simply recall that the absolute error of the trapezoidal rule is given by ([5, sec. 7.4.2])

$$|e_a| = \frac{1}{12} \sum_{k=1}^{K} \left| \frac{\partial^2 F_x(\xi_k)}{\partial x^2} \right| (\tau_{k+1} - \tau_k)^3 \quad (37)$$

where $\partial^2 F_x(\xi_k)/\partial x^2$ is the second derivative of the noise CCDF evaluated at some point $\xi_k \in (\tau_k, \tau_{k+1})$. By noting that $\partial^2 F_x(\xi_k)/\partial x^2 = \dot{p}_w(\xi_k - \theta)$, and using the extreme values $\tau_{\max}$ and $\dot{p}_{\max}$, (37) can be readily bounded as in (34).

Equation (35) follows after recalling how the variance of a linear combination $(\hat{\theta})$ of independent random variables $(\hat{q}_k)$ is related to the sum of the variances of the summands

$$\text{var}(\hat{\theta}) = \sum_{k=1}^{K} \frac{(\tau_{k+1} - \tau_{k-1})^2}{4} \text{var}(\hat{q}_k) \quad (38)$$

and using the fact that $\text{var}(\hat{q}_k) = q_k(1-q_k)/(N/K)$. □

A number of interesting remarks can be made about (34)–(35).

First note from (38) that the larger contributions to $\text{var}(\hat{\theta})$ occur when $q_k \approx 1/2$, since this value maximizes the coefficients $\text{var}(\hat{q}_k)$; equivalently, this happens when the thresholds satisfy $\tau_k \approx \theta$ [cf. (29)]. Thus, as with the case where the noise pdf is known, when θ belongs to an a priori known interval $[\Theta_1, \Theta_2]$, this knowledge must be exploited in selecting thresholds around the likeliest values of θ.

On the other hand, note that the $\text{var}(\hat{\theta})$ term in (33) will dominate $|e_a|^2$, because $|e_a|^2 \propto \tau_{\max}^4$ as per (34). To clarify this point, consider an equispaced grid of thresholds with $\tau_{k+1} - \tau_k = \tau = \tau_{\max}, \forall k$, such that $\tau_{\max} = 2T/(K+1) < 2T/K$. Using the (loose) bound $q_k(1-q_k) \leq 1/4$, the MSE is bounded by [cf. (33)–(35)]

$$\text{E}[(\theta - \hat{\theta})^2] < \frac{4T^6\dot{p}_{\max}^2}{9K^4} + \frac{T^2}{N}. \quad (39)$$

The bound in (39) is minimized by selecting $K = N$, which amounts to having *each sensor use a different region* to construct its binary observation. In this case, $|e_a|^2 \propto N^{-4}$ and its effect becomes practically negligible. Moreover, most pdfs have relatively small derivatives; e.g., for the Gaussian pdf we have $\dot{p}_{\max} = (2\pi e \sigma^4)^{-1/2}$. The integration error can be further reduced by resorting to a more powerful numerical integration method, although its difference with respect to the trapezoidal rule will not have any impact in practice.

Since $K = N$, the selection $\tau_{k+1} - \tau_k = \tau, \forall k$, reduces the estimator (30) to

$$\hat{\theta} = \tau \sum_{n=0}^{N-1} b_1(n) - T = T\left[\frac{1}{N+1} \sum_{n=0}^{N-1} b_1(n) - 1\right] \quad (40)$$

that *does not require knowledge of the threshold* used to construct the binary observation at the fusion center of a WSN. This feature allows for each sensor to randomly select its threshold without using values preassigned by the fusion center; see also [13] and [14] for related random quantization algorithms.

Remark 1: While ${e_a}^2 \propto T^6$ seems to dominate $\text{var}(\hat{\theta}) \propto T^2$ in (39), this is not true for the operational low-to-medium Q-SNR range for distributed estimators based on binary observations. This is because the support $2T$ over which $F_x(x)$ in (27) is nonzero depends on σ and the dynamic range $|\Theta_1 - \Theta_2|$ of the parameter θ. And as the Q-SNR decreases, $T \propto \sigma$. But since $\dot{p}_{\max} \propto \sigma^{-2}$ the integration error is ${e_a}^2 \propto \sigma^2/N^4$ which is negligible when compared to the term $\text{var}(\hat{\theta}) \propto \sigma^2/N$.

[3]We recall that in the notation $B_k(n)$, the argument n denotes the sensor and the subscript k a region used by this sensor. In this sense, $B_1(n)$ signifies that each sensor is using only one threshold.

B. Dependent Binary Observations

Similar to Section III-C, the second possible approach is to let each sensor form more than one binary observation per $x(n)$, Different from Section III-C, the performance advantage will lie on the side of independent binary observations. Define

$$B_k(n) := B_k = (\tau_k, \infty),\ n = 0, \ldots, N-1, \quad k = 1, \ldots, K \tag{41}$$

and let each sensor transmit the vector of binary observations $\mathbf{b} := (b_1(n), \ldots, b_K(n))^T$. As before, let $\mathbf{q} := (q_1, \ldots, q_K)^T$ denote the vector of Bernoulli parameters.

Since by definition $\mathbf{b}$ can only take on values of the form $\mathbf{b} = (1, \ldots 1, 0, \ldots, 0)^T$, we deduce that the number of bits required by this approach is $N_b = N \log_2(K)$.

Surprisingly, Lemma 1, extends to this case as well.

Lemma 2: The MLE of $\mathbf{q}$ based on the binary observations $\{\mathbf{b}(n)\}_{n=0}^{N-1}$ is given by

$$\hat{\mathbf{q}} = \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{b}(n) \tag{42}$$

with covariance between elements $l \geq k$

$$\text{cov}(\hat{q}_k, \hat{q}_l) = \frac{q_l(1-q_k)}{N}. \tag{43}$$

Proof: See Appendix B.1

When the binary observations come from the same $x(n)$, the optimum estimators for q_k are exactly the same as when they come from independent observations. Furthermore, the variance of $\hat{q}_k$ is identical for both cases as can be seen by setting $k = l$ in (43).

The variance of $\hat{\theta}$, alas, will be different when we rely on dependent binary observations. While e_a will remain the same, $\text{var}(\hat{\theta})$ will turn out to be bounded as stated in the ensuing proposition.

Proposition 4: Let $\hat{\theta}$ be the estimator in (30), with $\hat{q}_k$ denoting the kth component of $\hat{\mathbf{q}}$ in (42). The mse is given as in (33), with e_a bounded as in (34), and variance bounded as,

$$\text{var}(\hat{\theta}) \leq \frac{\tau_{\max}^2 K^2}{4N}. \tag{44}$$

Proof: See Appendix B.2.

As we did in Section IV-A, let us consider equally spaced thresholds $\tau_{k+1} - \tau_k := \tau = 2T/(K+1) \forall k$, to obtain [cf. (34), (35), (44)]

$$\text{E}[(\theta - \hat{\theta})^2] < \frac{4T^6 \dot{p}_{\max}^2}{9\,K^4} + \frac{T^2}{N}. \tag{45}$$

Notice that the MSE bound in (45) coincides with (39). Considering the extra bandwidth required by the estimator based on correlated binary observations, the one relying on independent ones is preferable when the noise pdf is unknown. However, one has to be careful when comparing bounds (as opposed to exact performance metrics); this is particularly true for this problem since the penalty in the required number of bits is small, namely a factor $\log_2(K)$. A fair statement is that in general both estimators will have comparable variance, and the selection would better be based on other criteria such as sensor complexity or the cost of collecting the $x(n)$ observations.

Apart from providing useful bounds on the finite-sample performance, (34), (35), (39), and (44), establish asymptotic optimality of the $\hat{\theta}$ estimators in (30) and (40) as summarized in the following:

Corollary 1: Under the assumptions of Propositions 3 and 4, and the conditions: i) $\tau_{\max} \propto K^{-1}$; and ii) $T^2/N, T^6/K^4 \to 0$ as $T, K, N \to \infty$, the estimators $\hat{\theta}$ in (30) and (40) are asymptotically (as $K, N \to \infty$) unbiased and consistent in the mean-square sense.

Proof: Notice that i) ensures that the MSE bounds in (39) and (45) hold true; and let $T, K, N \to \infty$ with the convergence rates satisfying ii) to conclude that $\text{E}[(\theta - \hat{\theta})^2] \to 0$. □

The estimators in (30) and (40) are consistent even if the support of the data *pdf* is infinite, as long as we guarantee a proper rate of convergence relative to the number of sensors and thresholds.

Remark 2: pdf-unaware bandwidth-constrained distributed estimation was introduced in [13], where it was referred to as universal. While the approach here is different, implicitly utilizing the data pdf (through the numerical approximation of the CCDF) to construct the consistent estimator of (30); the mse bound (39) for the simplified estimator (40) coincides with the mse bound for the universal estimator in (13). Note though, that the general mse expression of Proposition 3 can be used to optimize the placement and allocation of thresholds across sensors to lower the mse. Also different from [13], our estimators can afford noise pdfs with unbounded support as asserted by Corollary 1; and as we will see in Section V, the approach herein can be readily generalized to vector parameter estimation—a practical scenario where universal estimators like [13] are yet to be found.

C. Practical Considerations

At this point, it is interesting to compare the estimators in (4), (14), and (40). For that matter, consider that $\theta \in [\Theta_1, \Theta_2] = [-\sigma, \sigma]$, and that the noise is Gaussian with variance σ^2, yielding a Q-SNR $\gamma = 4$. None of these estimators can have variance smaller than $\text{var}(\bar{x}) = \sigma^2/N$; however, for the (medium) $\gamma = 4$ Q-SNR value they can come close. For the known pdf estimator in (4), the variance is $\text{var}(\hat{\theta}) \approx 2\sigma^2/N$. For the known pdf, unknown variance estimator in (14) we find $\text{var}(\hat{\theta}) \approx 3\sigma^2/N$. The unknown pdf estimator in (40) requires an assumption about the essentially nonzero support of the Gaussian pdf. If we suppose that the noise pdf is nonzero over $[-2\sigma, 2\sigma]$, the corresponding variance becomes $\text{var}(\hat{\theta}) \approx 9\sigma^2/N$. Respectively, the penalties due to the transmission of a single bit per sensor with respect to $\bar{x}$ are approximately 2, 3, and 9. While the increasing penalty is expected as the uncertainty about the noise pdf increases, the relatively small loss is rather unexpected.

All the estimators discussed so far rely on certain thresholds τ_k. Either this threshold has to be communicated to the nodes by the fusion center, or, one can resort to the iterative approach discussed in Section III-A. These two approaches are different in terms of transmission cost and estimation accuracy. Assuming a resource-rich fusion center, the cost of transmitting the thresholds is indeed negligible, and the batch approach incurs an overall small transmission cost. However, it relies on rough *a priori* knowledge that can quickly become outdated. The iterative estimator, on the other hand, is always using the best available information for threshold positioning but requires continuous updates which increase transmission cost. A hybrid of these two approaches may offer a desirable tradeoff between estimation accuracy and transmission cost, and constitutes an interesting direction for future research.

V. Vector Parameter Generalization

Let us now return to the general problem we started with in Section II. We begin by defining the per sensor vector of binary observations $\mathbf{b}(n) := (b_1(n), \ldots, b_K(n))^T$, and note that since its entries are binary, realizations $\boldsymbol{\beta}$ of $\mathbf{b}(n)$ belong to the set

$$\mathcal{B} := \{\boldsymbol{\beta} \in \mathbf{R}^K \mid [\boldsymbol{\beta}]_k \in \{0, 1\}, k = 1, \ldots, K\} \tag{46}$$

where $[\boldsymbol{\beta}]_k$ denotes the kth component of $\boldsymbol{\beta}$. With each $\boldsymbol{\beta} \in \mathcal{B}$ and each sensor we now associate the region

$$\mathbf{B}_{\beta}(n) := \bigcap_{[\boldsymbol{\beta}]_k=1} B_k(n) \bigcap_{[\boldsymbol{\beta}]_k=0} \bar{B}_k(n) \tag{47}$$

where $\bar{B}_k(n)$ denotes the set-complement of $B_k(n)$ in $\mathbf{R}^M$. Note that the definition in (47) implies that $\mathbf{x}(n) \in \mathbf{B}_{\beta}(n)$ if and only if $\mathbf{b}(n) = \boldsymbol{\beta}$; see also Fig. 3 for an illustration in $\mathbf{R}^2$ ($M = 2$). The corresponding probabilities are

$$\begin{aligned} q_{\beta}(n) &:= \Pr\{\mathbf{b}(n) = \boldsymbol{\beta}\} \\ &= \Pr\{\mathbf{x}(n) \in \mathbf{B}_{\beta}(n)\} \\ &= \int_{\mathbf{B}_{\beta}(n)} p_{\mathbf{w}}[\mathbf{u} - \mathbf{f}_n(\boldsymbol{\theta}); \boldsymbol{\psi}]\, d\mathbf{u} \end{aligned} \tag{48}$$

with $\mathbf{f}_n$ as in (1), and $\boldsymbol{\psi}$ containing the unknown parameters of the known noise pdf. Using (48) and (46), we can write the pertinent log-likelihood function as

$$L(\boldsymbol{\theta}, \boldsymbol{\psi}) = \sum_{n=0}^{N-1} \sum_{\boldsymbol{\beta} \in \mathcal{B}} \delta(\mathbf{b}(n) - \boldsymbol{\beta}) \ln q_{\beta}(n) \tag{49}$$

and the MLE of $\boldsymbol{\theta}$ as

$$\hat{\boldsymbol{\theta}} = \arg\max_{(\boldsymbol{\theta}, \boldsymbol{\psi})} L(\boldsymbol{\theta}, \boldsymbol{\psi}). \tag{50}$$

The nonlinear search needed to obtain $\hat{\boldsymbol{\theta}}$ could be challenged either by the multimodal nature of $L(\boldsymbol{\theta}, \boldsymbol{\psi})$ or by numerical ill-conditioning caused by, e.g., saddle points or by $q(n)$ values close to zero for which $L(\boldsymbol{\theta}, \boldsymbol{\psi})$ becomes unbounded. While this is true in general, under certain conditions that are usually met in practice, $L(\boldsymbol{\theta}, \boldsymbol{\psi})$ is concave which implies that computationally efficient search algorithms can be invoked to find its global maximum. This subclass is defined in the following proposition.

Proposition 5: If the MLE problem in (50) satisfies the conditions:

c1) The noise pdf $p_{\mathbf{w}}(\mathbf{w}; \boldsymbol{\psi}) \leftrightarrow p_{\mathbf{w}}(\mathbf{w})$ is log-concave ([4, p. 104]), and $\boldsymbol{\psi}$ is known.
c2) The functions $\mathbf{f}_n(\boldsymbol{\theta})$ are linear; i.e., $\mathbf{f}_n(\boldsymbol{\theta}) = \mathbf{H}_n \boldsymbol{\theta}$, with $\mathbf{H}_n \in \mathbf{R}^{(M \times p)}$.
c3) The regions $B_k(n)$ are chosen as half-spaces.

then $L(\boldsymbol{\theta})$ in (49) is a concave function of $\boldsymbol{\theta}$.

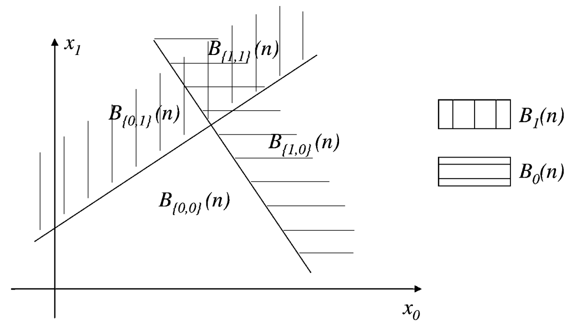

Fig. 3. The vector of binary observations $\mathbf{b}$ takes on the value $\{\beta_1, \beta_2\}$ if and only if $\mathbf{x}(n)$ belongs to the region $B_{\{\beta_1, \beta_2\}}$.

Proof: See Appendix C.

Note that c1) is satisfied by common noise pdfs, including the multivariate Gaussian ([4, p. 104]); and also that c2) is typical in parameter estimation. Moreover, even when c2) is not satisfied, linearizing $\mathbf{f}_n(\boldsymbol{\theta})$ using Taylor's expansion is a common first step, typical in, e.g., parameter tracking applications. On the other hand, c3) places a constraint in the regions defining the binary observations, which is simply up to the designer's choice.

The importance of Proposition 5 is that maximization of a concave function is a well-behaved numerical problem safely solvable by standard descent methods such as Newton's algorithm. Proposition 5 nicely generalizes our earlier results on scalar parameter estimators in [23] to the more practical case of vector parameters and vector observations.

A. Colored Gaussian Noise

Analyzing the performance of the MLE in (50) is only possible asymptotically (as N or SNR go to infinity). Notwithstanding, when the noise is Gaussian, simplifications render variance analysis tractable and lead to interesting guidelines for constructing the estimator $\hat{\boldsymbol{\theta}}$.

Restrict $p_{\mathbf{w}}(\mathbf{w}; \boldsymbol{\psi}) \leftrightarrow p_{\mathbf{w}}(\mathbf{w})$ to the class of multivariate Gaussian pdfs, and let $\mathbf{C}(n)$ denote the noise covariance matrix at sensor n. Assume that $\{\mathbf{C}(n)\}_{n=0}^{N-1}$ are known and let $\{(\mathbf{e}_m(n), \sigma_m^2(n))\}_{m=1}^{M}$ be the set of eigenvectors and associated eigenvalues

$$\mathbf{C}(n) = \sum_{m=1}^{M} \sigma_m^2(n) \mathbf{e}_m(n) \mathbf{e}_m^T(n). \tag{51}$$

For each sensor, we define a set of $K = M$ regions $B_k(n)$ as half-spaces whose borders are hyperplanes perpendicular to the covariance matrix eigenvectors; i.e.,

$$B_k(n) = \left\{\mathbf{x} \in \mathbf{R}^M \;\middle|\; \mathbf{e}_k^T(n)\mathbf{x} \geq \tau_k(n)\right\} \quad k = 1, \ldots, K = M. \tag{52}$$

Fig. 4 depicts the regions $B_k(n)$ in (52) for $M = 2$. Note that since each entry of $\mathbf{x}(n)$ offers a distinct scalar observation, the selection $K = M$ amounts to a bandwidth constraint of 1 *bit per sensor per dimension*.

The rationale behind this selection of regions is that the resultant binary observations $b_k(n)$ are independent, meaning that $\Pr\{b_{k_1}(n) b_{k_2}(n)\} = \Pr\{b_{k_1}(n)\} \Pr\{b_{k_2}(n)\}$ for $k_1 \neq k_2$. As a result, we have a total of MN independent binary observations to estimate $\boldsymbol{\theta}$.

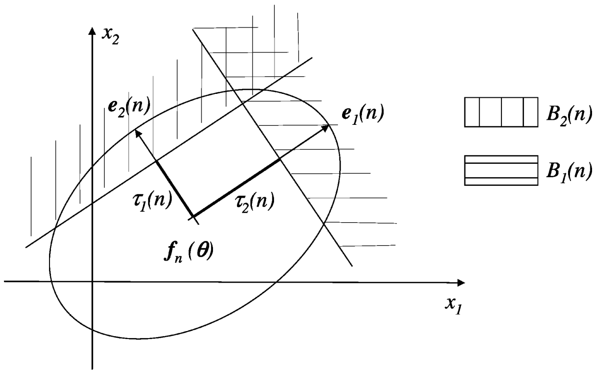

Fig. 4. Selecting the regions $B_k(n)$ perpendicular to the covariance matrix eigenvectors results in independent binary observations.

Herein, the Bernoulli parameters $q_k(n)$ take on a particularly simple form in terms of the Gaussian tail function $Q(u) := (1/\sqrt{2\pi})\int_u^\infty e^{-u^2/2}\,du$

$$\begin{aligned} q_k(n) &= \int\limits_{\mathbf{e}_k^T(n)\mathbf{u}\geq\tau_k(n)} p_{\mathbf{w}}(\mathbf{u}-\mathbf{f}_n(\boldsymbol{\theta}))\,d\mathbf{u} \\ &= Q\left(\frac{\tau_k(n)-\mathbf{e}_k^T(n)\mathbf{f}_n(\boldsymbol{\theta})}{\sigma_k(n)}\right) \\ &:= Q(\Delta_k(n)) \end{aligned} \tag{53}$$

where we introduced the *σ-distance* between $\mathbf{f}_n(\boldsymbol{\theta})$ and the corresponding threshold $\Delta_k(n) := [\tau_k(n)-\mathbf{e}_k^T(n)\mathbf{f}_n(\boldsymbol{\theta})]/\sigma_k(n)$.

The independence among binary observations implies that $p(\mathbf{b}(n)) = \prod_{k=1}^K [q_k(n)]^{b_k(n)}[1-q_k(n)]^{1-b_k(n)}$, and leads to a simple log-likelihood function

$$L(\boldsymbol{\theta}) = \sum_{n=0}^{N-1}\sum_{k=1}^{K} b_k(n)\ln q_k(n) + [1-b_k(n)]\ln[1-q_k(n)] \tag{54}$$

whose NK *independent* summands replace the $N2^K$ *dependent* summands in (49).

Since the regions $B_k(n)$ are half-spaces, Proposition 5 applies to the maximization of (54) and guarantees that the numerical search for the $\hat{\boldsymbol{\theta}}$ estimator in (54) is well conditioned and will converge to the global maximum, at least when the functions $\mathbf{f}_n$ are linear. More important, it will turn out that these regions render finite sample performance analysis of the MLE in (50), tractable. In particular, it is possible to derive a closed-form expression for the Fisher Information Matrix (FIM) ([8, p. 44]), as we establish next.

Proposition 6: The FIM, $\mathbf{I}$, for estimating $\boldsymbol{\theta}$ based on the binary observations obtained from the regions defined in (52), is given by

$$\mathbf{I} = \sum_{n=0}^{N-1}\mathbf{J}_n^T\left[\sum_{k=1}^{K}\frac{e^{-\Delta_k^2(n)}\mathbf{e}_k(n)\mathbf{e}_k^T(n)}{2\pi\sigma_k^2(n)Q(\Delta_k(n))[1-Q(\Delta_k(n))]}\right]\mathbf{J}_n \tag{55}$$

where $\mathbf{J}_n$ denotes the Jacobian of $\mathbf{f}_n(\boldsymbol{\theta})$.

Proof: We just have to consider the second derivative of the log-likelihood function in (54)

$$\begin{aligned} \frac{\partial^2 L(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^2} = \sum_{n=0}^{N-1}\sum_{k=1}^{K}\frac{\partial^2 q_k}{\partial\boldsymbol{\theta}^2}\left[\frac{b_k(n)}{q_k(n)}-\frac{1-b_k(n)}{1-q_k(n)}\right] \\ -\frac{\partial q_k}{\partial\boldsymbol{\theta}}\frac{\partial^T q_k}{\partial\boldsymbol{\theta}}\left[\frac{b_k(n)}{[q_k(n)]^2}+\frac{1-b_k(n)}{[1-q_k(n)]^2}\right] \end{aligned} \tag{56}$$

and take expected value with respect to the binary observations $b_k(n)$ to obtain

$$\mathrm{E}\left[\frac{\partial^2 L(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^2}\right] = \sum_{n=0}^{N-1}\sum_{k=1}^{K} -\frac{\partial q_k}{\partial\boldsymbol{\theta}}\frac{\partial^T q_k}{\partial\boldsymbol{\theta}}\left[\frac{1}{q_k(n)}+\frac{1}{1-q_k(n)}\right] \tag{57}$$

where we used the fact that $\mathrm{E}[b_k(n)] = q_k(n)$.

On the other hand, differentiating q_k with respect to $\boldsymbol{\theta}$ yields [cf. (53)]

$$\frac{\partial q_k}{\partial\boldsymbol{\theta}} = \frac{e^{-\Delta_k^2(n)/2}}{\sqrt{2\pi}\sigma_k(n)}\mathbf{J}_n^T\mathbf{e}_k(n). \tag{58}$$

The FIM is obtained as the negative of the expected value in (57) if we also substitute (58) into (57), we obtain

$$\mathbf{I} = \sum_{n=0}^{N-1}\sum_{k=1}^{K}\frac{e^{-\Delta_k^2(n)}}{2\pi\sigma_k^2(n)}\mathbf{J}_n^T\mathbf{e}_k(n)\mathbf{e}_k^T(n)\mathbf{J}_n\frac{1}{q_k(n)[1-q_k(n)]}. \tag{59}$$

Moving the common factor $\mathbf{J}_n$ outside the innermost summation and substituting $q_k(n)$ by its value in (53), we obtain (55).□

The FIM places a lower bound in the achievable variance of unbiased estimators since the covariance of any estimator must satisfy

$$\mathrm{cov}(\hat{\boldsymbol{\theta}}) - \mathbf{I}^{-1} \succeq \mathbf{0} \tag{60}$$

where the notation $\succeq \mathbf{0}$ stands for positive semidefiniteness of a matrix; the variances in particular are bounded by $\mathrm{var}(\hat{\boldsymbol{\theta}}_k) \geq [\mathbf{I}^{-1}]_{kk}$.

Inspection of (55) shows that the variance of the MLE in (50) depends on the signal function containing the parameter of interest (via the Jacobians), the noise structure and power (via the eigenvalues and eigenvectors), and the selection of the regions $B_k(n)$ (via the σ-distances). Among these three factors, only the last one is inherent to the bandwidth constraint, the other two being common to the estimator that is based on the original $\mathbf{x}(n)$ observations.

The last point is clarified if we consider the FIM $\mathbf{I}_x$ for estimating $\boldsymbol{\theta}$ given the unquantized vector observations $\mathbf{x}(n)$. This matrix can be shown to be (see Appendix D)

$$\mathbf{I}_x = \sum_{n=0}^{N-1}\mathbf{J}_n^T\left[\sum_{m=1}^{M}\frac{\mathbf{e}_m(n)\mathbf{e}_m^T(n)}{\sigma_m^2(n)}\right]\mathbf{J}_n^T. \tag{61}$$

If we define the equivalent noise powers as

$$\rho_k^2(n) := \frac{2\pi Q(\Delta_k(n))[1-Q(\Delta_k(n))]}{e^{-\Delta_k^2(n)}}\sigma_k^2(n) \tag{62}$$

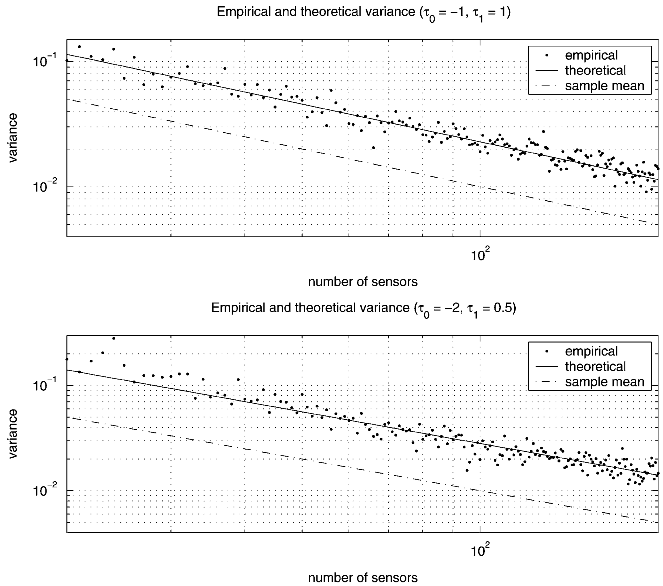

Fig. 5. Noise of unknown power estimator. The CRLB in (15) is an accurate prediction of the variance of the MLE estimator (14) moreover, its variance is close to the clairvoyant sample mean estimator based on the analog observations ($\sigma = 1, \theta = 0$, Gaussian noise).

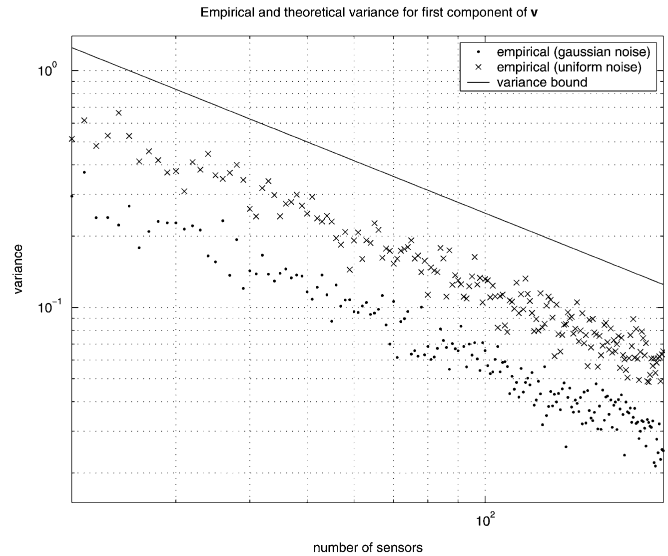

Fig. 6. Universal estimator introduced in Section IV. The bound in (39) overestimates the real variance by a factor that depends on the noise pdf ($\sigma = 1, T = 5$, θ chosen randomly in $[-2, 2]$).

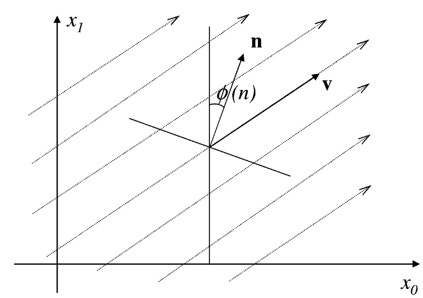

Fig. 7. The vector flow $\mathbf{v}$ incises over a certain sensor capable of measuring the normal component of $\mathbf{v}$.

we can rewrite (55) in the form

$$\mathbf{I} = \sum_{n=0}^{N-1} \mathbf{J}_n^T \left[\sum_{k=1}^{K} \frac{\mathbf{e}_k(n)\mathbf{e}_k^T(n)}{\rho_k^2(n)} \right] \mathbf{J}_n^T \tag{63}$$

which, except for the noise powers, has form identical to (61). Thus, comparison of (63) with (61) reveals that from a performance perspective, *the use of binary observations is equivalent to an increase in the noise variance* from $\sigma_k^2(n)$ to $\rho_k^2(n)$, while the rest of the problem structure remains unchanged.

Since we certainly want the equivalent noise increase to be as small as possible, minimizing (62) over $\Delta_k(n)$ calls for this distance to be set to zero, or equivalently, to select thresholds $\tau_k(n) = \mathbf{e}_k^T(n)\mathbf{f}_n(\boldsymbol{\theta})$. In this case, the equivalent noise power is

$$\rho_k^2(n) = \frac{\pi}{2}\sigma_k^2(n). \tag{64}$$

Surprisingly, even in the vector case a judicious selection of the regions $B_k(n)$ results in a very small penalty $(\pi/2)$ in terms of the equivalent noise increase. Similar to Sections III-A and III-B, we can, thus, claim that while requiring the transmission of 1 bit per sensor per dimension, the variance of the MLE in (50), based on $\{\mathbf{b}(n)\}_{n=0}^{N-1}$, yields a variance close to the clairvoyant estimator's variance—based on $\{\mathbf{x}(n)\}_{n=0}^{N-1}$—for low-to-medium Q-SNR problems.

VI. Simulations

A. Scalar Parameter Estimation

We begin by simulating the estimator in (14) for scalar parameter estimation in the presence of AWGN with unknown variance. Results are shown in Fig. 5 for two different sets of σ-distances, Δ_1, Δ_2, corroborating the values predicted by (15) and the fact that the performance loss with respect to the clairvoyant sample mean estimator $\bar{x}$ is indeed small.

Without invoking assumptions on the noise pdf, we also tested the simplified estimator in (40). Fig. 6 shows one such test, depicting the bound in (39), as well as simulated variances for uniform and Gaussian noise pdfs. Note that the bound overestimates the variance by a factor of roughly 2 for the uniform case and roughly 4 for the Gaussian case. Note that having unbounded derivative, the uniform pdf is not covered by Proposition 3; however, for piecewise linear CCDFs of which uniform noise is a special case, (34) does not hold true but the error of the trapezoidal rule e_a is small anyways, as testified by the corresponding points in Fig. 6.

B. Vector Parameter Estimation—A Motivating Application

In this section, we illustrate how a problem involving vector parameters can be solved using the estimators of Section V-A. Suppose we wish to estimate a vector flow using incidence observations. With reference to Fig. 7, consider the flow vector $\mathbf{v} := (v_0, v_1)^T$, and a sensor positioned at an angle $\phi(n)$ with respect to a known reference direction. We will rely on a set of so-called incidence observations $\{x(n)\}_{n=0}^{N-1}$ measuring the component of the flow normal to the corresponding sensor

$$x(n) := \langle \mathbf{v}, \mathbf{n} \rangle + w(n) = v_0 \sin[\phi(n)] + v_1 \cos[\phi(n)] + w(n) \tag{65}$$

where $\langle , \rangle$ denotes inner product, $w(n)$ is zero-mean AWGN, and the equation holds for $n = 0, 1, \ldots, N-1$. The model (65) applies to the measurement of hydraulic fields, pressure variations induced by wind and radiation from a distant source [15].

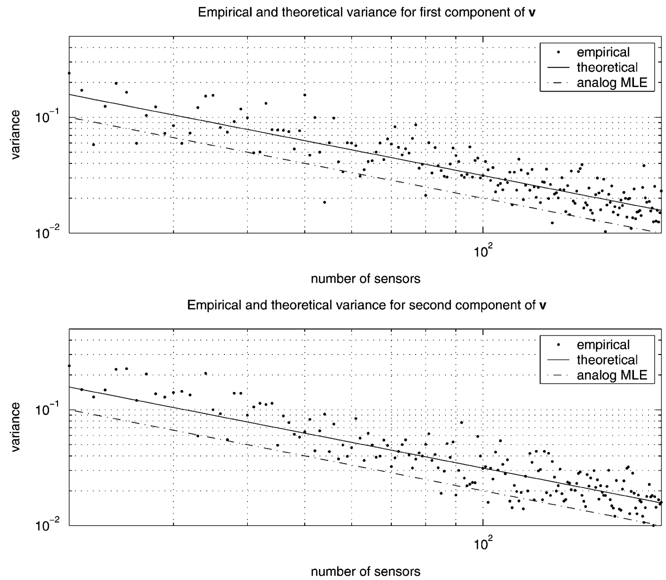

Fig. 8. Average variance for the components of $\mathbf{v}$. The empirical as well as the bound (68) are compared with the analog observations based MLE ($\mathbf{v} = (1,1), \sigma = 1$).

Estimating $\mathbf{v}$ fits the framework of Section V.A requiring the transmission of a single binary observation per sensor, $b_1(n) = \mathbf{1}\{x(n) \geq \tau_1(n)\}$. The FIM in (63) is easily found to be

$$\mathbf{I} = \sum_{n=0}^{N-1} \frac{1}{\rho_1^2(n)} \begin{pmatrix} \sin^2[\phi(n)] & \sin[\phi(n)]\cos[\phi(n)] \\ \sin[\phi(n)]\cos[\phi(n)] & \cos^2[\phi(n)] \end{pmatrix}. \tag{66}$$

Furthermore, since $x(n)$ in (65) is linear in $\mathbf{v}$ and the noise pdf is log-concave (Gaussian) the log-likelihood function is concave as asserted by Proposition 5.

Suppose that we are able to place the thresholds optimally at $\tau_1(n) = v_0 \sin[\phi(n)] + v_1 \cos[\phi(n)]$, so that $\rho_1^2(n) = (\pi/2)\sigma^2$. If we also make the reasonable assumption that the angles are random and uniformly distributed $\phi(n) \sim U[-\pi, \pi]$ then the average FIM turns out to be

$$\bar{\mathbf{I}} = \frac{2}{\pi\sigma^2} \begin{pmatrix} N/2 & 0 \\ 0 & N/2 \end{pmatrix}. \tag{67}$$

But according to the law of large numbers $\mathbf{I} \approx \bar{\mathbf{I}}$, and the estimation variance will be approximately given by

$$\mathrm{var}(v_0) = \mathrm{var}(v_1) = \frac{\pi\sigma^2}{N}. \tag{68}$$

Fig. 8 depicts the bound (68), as well as the simulated variances $\mathrm{var}(\hat{v}_0)$ and $\mathrm{var}(\hat{v}_1)$ in comparison with the clairvoyant MLE based on $\{x(n)\}_{n=0}^{N-1}$, corroborating our analytical expressions. While this excellent performance is obtained under ideal threshold placement, recalling the harsh bandwidth constraint (1 bit per sensor) justifies the potential of our approach for bandwidth-constrained distributed parameter estimation in this WSN-based context.

VII. Conclusion

We were motivated by the need to effect energy savings in a wireless sensor network deployed to estimate parameters of interest in a decentralized fashion. To this end, we developed parameter estimators for realistic signal models and derived their fundamental variance limits under bandwidth constraints. The latter were adhered to by quantizing each sensor's observation to one or a few bits. By jointly accounting for the unique quantization-estimation tradeoffs present, these bit(s) per sensor were first used to derive distributed MLEs for scalar mean-location parameters in the presence of generally non-Gaussian noise when the noise pdf is completely known; subsequently, when the pdf is known except for a number of unknown parameters; and finally, when the noise pdf is unknown. The unknown pdf case was tackled through a nonparametric estimator of the unknown complementary cumulative distribution function based on quantized (binary) observations.

In all three cases, the resulting estimators turned out to exhibit comparable variances that can come surprisingly close to the variance of the clairvoyant estimator which relies on unquantized observations. This happens when the SNR capturing both quantization and noise effects assumes low-to-moderate values. Analogous claims were established for practical generalizations that were pursued in the multivariate and colored noise cases for distributed estimation of vector parameters under bandwidth constraints. Therein, MLEs were formed via numerical search but the log-likelihoods were proved to be concave thus ensuring fast convergence to the unique global maximum.

A motivating application was also considered reinforcing the conclusion that in low-cost-per-node wireless sensor networks, distributed parameter estimation based even on a single bit per observation is possible with minimal increase in estimation variance[4].

Appendix

A. Proofs of Lemma 1 and Proposition 2

1) Lemma 1: That $\hat{\mathbf{q}}$ is unbiased follows from the linearity of expectation and the fact that $\mathbf{q} = \mathrm{E}[\mathbf{b}(n)]$. We will also establish that $\hat{\mathbf{q}}$ is the MLE using Cramér-Rao's theorem; the result will actually be stronger since $\hat{\mathbf{q}}$ is in fact the minimum variance unbiased estimator (MVUE) of $\mathbf{q}$. To this end, using that $\tau_2 > \tau_1$, the log-likelihood function takes on the form

$$\begin{aligned} L(\mathbf{b}, \mathbf{q}) = \sum_{n=0}^{N-1} & [1 - b_1(n)][1 - b_2(n)] \ln \Pr\{x(n) < \tau_1\} \\ & + b_1(n)[1 - b_2(n)] \ln \Pr\{\tau_1 < x(n) < \tau_2\} \\ & + b_1(n) b_2(n) \ln \Pr\{\tau_2 < x(n)\}. \end{aligned} \tag{69}$$

Note that we can rewrite the probabilities in terms of the components of $\mathbf{q}$, since e.g., $\Pr\{\tau_1 < x(n) < \tau_2\} = q_1 - q_2$. Moreover, since the binary observations are either 0 or 1 and *the combination* $\mathbf{b}(n) = (0,1)$ *is impossible*, we can simplify the products of binary observations; e.g., $b_1(n)[1 - b_2(n)] = b_2(n) - b_1(n)$. Enacting these simplifications in (69), we obtain

$$\begin{aligned} L(\mathbf{b}, \mathbf{q}) = \sum_{n=0}^{N-1} & [1 - b_1(n)] \ln[1 - q_1] \\ & + [b_1(n) - b_2(n)] \ln[q_1 - q_2] + b_2(n) \ln q_2. \end{aligned} \tag{70}$$

From (70), we can obtain the gradient of $L(\mathbf{b}, \mathbf{q})$ as

$$\frac{\partial L}{\partial \mathbf{q}} = \sum_{n=0}^{N-1} \begin{pmatrix} \frac{1-b_1(n)}{1-q_1} & + & \frac{b_1(n)-b_2(n)}{q_1-q_1} \\ -\frac{b_1(n)-b_2(n)}{q_1-q_1} & + & \frac{b_2(n)}{q_2} \end{pmatrix}$$

[4]The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U. S. Government.

$$= N \begin{pmatrix} \frac{1-\hat{q}_1(n)}{1-q_1} & + & \frac{\hat{q}_1(n)-\hat{q}_2(n)}{q_1-q_1} \\ -\frac{\hat{q}_1(n)-\hat{q}_2(n)}{q_1-q_1} & + & \frac{\hat{q}_2(n)}{q_2} \end{pmatrix}. \quad (71)$$

Differentiating once more yields the Hessian, and taking expected value over $\mathbf{b}$ yields the FIM

$$\mathbf{I}(\mathbf{q}) = \frac{N}{q_1 - q_2} \begin{pmatrix} \frac{1-q_2}{1-q_1} & -1 \\ -1 & \frac{q_1}{q_2} \end{pmatrix}. \quad (72)$$

It is a matter of simple algebra to verify that $\partial L/\partial \mathbf{q} = \mathbf{I}(\mathbf{q})[\mathbf{q} - \hat{\mathbf{q}}]$, from where application of Cramer-Rao's theorem concludes the Proof of Lemma 1. □

2) Proposition 2: The proof is analogous to the one of Proposition 1. Let us start by inverting $\mathbf{I}(\mathbf{q})$ [cf. (72)]

$$\mathbf{I}^{-1}(\mathbf{q}) = \frac{1}{N} \begin{pmatrix} (1-q_1)q_1 & (1-q_1)q_2 \\ (1-q_1)q_2 & (1-q_2)q_2 \end{pmatrix}. \quad (73)$$

We can now apply the property stated in (19), with the inverse FIM given by (73), to obtain

$$\mathrm{var}(\hat{\theta}) \geq \left(\frac{\partial\theta}{\partial q_1}\right)^2 \frac{q_1(1-q_1)}{N} + \left(\frac{\partial\theta}{\partial q_2}\right)^2 \frac{q_2(1-q_2)}{N} + \left(\frac{\partial\theta}{\partial q_1}\right)\left(\frac{\partial\theta}{\partial q_2}\right) \frac{q_2(1-q_1)}{N}. \quad (74)$$

Substituting the derivatives from (20) into (74) completes the Proof of Proposition 2. □

B. Proofs of Lemma 2 and Proposition 4

1) Lemma 2: As in the Proof of Lemma 1, the key property is that only some combinations of binary observations are possible; hence, we have

$$b_1(n)\dots b_k(n)[1-b_{k+1}(n)]\dots[1-b_K(n)] = b_k - b_{k+1} \quad (75)$$

and the log-likelihood function takes the form

$$L(\mathbf{q}, \mathbf{b}) = \sum_{n=0}^{N-1}\sum_{k=0}^{K}[b_k(n) - b_{k+1}(n)]\ln(q_k - q_{k-1}) = N\sum_{k=0}^{K}[\hat{q}_k(n) - \hat{q}_{k+1}(n)]\ln(q_k - q_{k-1}) \quad (76)$$

where for the last equality we interchanged summations and substituted $\hat{q}_k$ from (42).

Applying the Neyman-Fisher factorization theorem to (76), we deduce that $\{\hat{q}_k\}_{k=1}^K$ are sufficient statistics for estimating $\mathbf{q}$ ([8, p. 104]). Furthermore, noting that $\mathrm{E}(\hat{\mathbf{q}}) = \mathbf{q}$ and that $\hat{\mathbf{q}}$ is a function of sufficient statistics, application of Rao-Blackwell-Lehmann-Scheffe theorem proves that $\hat{\mathbf{q}}$ is the MVUE (and consequently the MLE) of $\mathbf{q}$.

To compute the covariance, recall that $\mathrm{E}(b_k(n)) = q_k$ to find

$$\begin{aligned}\mathrm{cov}[b_k(n), b_l(n)] &:= \mathrm{E}[(b_k(n) - q_k)(b_l(n) - q_l)] \\ &= \mathrm{E}[b_k(n)b_l(n)] - q_k q_l \\ &= q_l - q_k q_l \end{aligned} \quad (77)$$

where for the last equality we used that $\mathrm{E}[b_k(n)b_l(n)] = \Pr\{b_k(n) = 1, b_l(n) = 1\} = q_l$, for $l \geq k$. The proof follows form the independence of binary observations across sensors

$$\mathrm{cov}[\hat{q}_k(n), \hat{q}_l(n)] = \frac{1}{N^2}\sum_{n=0}^{N-1}\mathrm{cov}[b_k(n), b_l(n)] = \frac{q_l(1-q_k)}{N}. \quad (78)$$

QED. □

2) Proposition 4: To compute the variance of $\hat{\theta}$, use the linearity of expectation to write

$$\mathrm{var}(\hat{\theta}) = \sum_{k=1}^{K}\sum_{l=1}^{K}\frac{(\tau_{k+1} - \tau_{k-1})(\tau_{l+1} - \tau_{l-1})}{4} \times \mathrm{E}[(\hat{q}_k(n) - q_k)(\hat{q}_l(n) - q_l)]. \quad (79)$$

Note that the expected value is by definition $\mathrm{E}[(\hat{q}_k(n) - q_k)(\hat{q}_l(n) - q_l)] = \mathrm{cov}[b_k(n), b_l(n)]$, and is, thus, given by (43) when $l \geq k$. Substituting these values into (79), we obtain

$$\mathrm{var}(\hat{\theta}) = \sum_{k=1}^{K}\frac{(\tau_{k+1} - \tau_{k-1})^2}{4}\frac{q_k(1-q_k)}{N} + 2\sum_{k=1}^{K}\sum_{l>k}\frac{(\tau_{k+1} - \tau_{k-1})(\tau_{l+1} - \tau_{l-1})}{4}\frac{q_l(1-q_k)}{N}. \quad (80)$$

Finally, note that $q_l(1-q_k) \leq 1/4$ and that $\tau_{k+1} - \tau_{k-1} < 2\tau_{\max}$, to arrive at

$$\mathrm{var}(\hat{\theta}) \leq \sum_{k=1}^{K}\frac{\tau_{\max}^2}{4N} + \sum_{k=1}^{K}\sum_{l>k}\frac{\tau_{\max}^2}{2N}. \quad (81)$$

Since the first sum contains K terms and the second $K(K-1)/2$, (44) follows. □

C. Proof of Proposition 5

Consider the indicator function associated with $\mathbf{B}_\beta(n)$

$$\mathbf{g}(\mathbf{w}) = \begin{cases} 1 & \mathbf{w} \in \mathbf{B}_\beta(n) \\ 0 & \text{else} \end{cases}. \quad (82)$$

Equation (47) and c3), imply that since $\mathbf{B}_\beta(n)$ is an intersection of half-spaces, it is convex (in fact c3) is both sufficient and necessary for the convexity of $\mathbf{B}_\beta(n), \forall\, \boldsymbol{\beta}$). Since $\mathbf{g}(\mathbf{w})$ is the indicator function of a convex set it is log-concave (and concave too).

Now, let us rewrite (48) as

$$q_\beta(n) = \int_{\mathbf{R}^M} g(\mathbf{w} + \mathbf{H}_n\boldsymbol{\theta})p_{\mathbf{w}}(\mathbf{w})\,d\mathbf{w} \quad (83)$$

and use the fact that $g(\mathbf{w}+\mathbf{H}_n\boldsymbol{\theta})$ is log-concave in its argument. Moreover, since c2) makes this argument affine in $(\mathbf{w}, \boldsymbol{\theta})$, it follows that $g(\mathbf{w} + \mathbf{H}_n\boldsymbol{\theta})$ is log-concave in $(\mathbf{w}, \boldsymbol{\theta})$. Since $p_{\mathbf{w}}(\mathbf{w})$ is log-concave under c1), the product $g(\mathbf{w} + \mathbf{H}_n\boldsymbol{\theta})p_{\mathbf{w}}(\mathbf{w})$ is log-concave too.

At this point, we can apply the integration property of log-concave functions to claim that $q_\beta(n)$ is log-concave, ([4, p. 104]). Finally, note that $L(\boldsymbol{\theta})$ comprises the sum of logarithms of log-concave functions; thus, each term is concave and so is their sum. □

D. FIM for Estimation of $\boldsymbol{\theta}$ Based on $\{\mathbf{x}(n)\}_{n=0}^{N-1}$

Let $\mathbf{x} := (\mathbf{x}^T(0), \mathbf{x}^T(1), \dots, \mathbf{x}^T(N-1))^T$ and consider the log-likelihood

$$\ln p_{\mathbf{x}}(\mathbf{x}, \boldsymbol{\theta}) = \sum_{n=0}^{N-1}\frac{1}{2}\ln[(2\pi)^M \det(\mathbf{C}(n))] - \frac{1}{2}[\mathbf{x}(n) - \mathbf{f}_n(\boldsymbol{\theta})]^T\mathbf{C}^{-1}(n)[\mathbf{x}(n) - \mathbf{f}_n(\boldsymbol{\theta})]. \quad (84)$$

Differentiating twice with respect to $\boldsymbol{\theta}$, we obtain the first

$$\frac{\partial \ln p_x(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{n=0}^{N-1}\mathbf{J}_n^T\mathbf{C}^{-1}(n)[\mathbf{x}(n) - \mathbf{f}_n(\boldsymbol{\theta})] \quad (85)$$

and second derivative

$$\frac{\partial^2 \ln p_{\mathbf{x}}(\mathbf{x},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} = \sum_{n=0}^{N-1} \frac{\partial \mathbf{J}_n^T}{\partial \boldsymbol{\theta}} \mathbf{C}^{-1}(n)[\mathbf{x}(n) - \mathbf{f}_n(\boldsymbol{\theta})] + \mathbf{J}_n^T \mathbf{C}^{-1}(n) \mathbf{J}_n. \quad (86)$$

Since $\mathrm{E}[\mathbf{x}(n)] = \mathbf{f}_n(\boldsymbol{\theta})$, taking the negative of the expected value in (86) yields the FIM

$$\mathbf{I}_{\mathbf{x}} = \sum_{n=0}^{N-1} \mathbf{J}_n^T \mathbf{C}^{-1}(n) \mathbf{J}_n. \quad (87)$$

Now, recall that the eigenvalues of $\mathbf{C}^{-1}(n)$ are the inverses of the eigenvalues of $\mathbf{C}(n)$, and the eigenvectors are equal. Finally, use $\mathbf{C}^{-1}(n) = \sum_{m=1}^{M} \sigma_m^{-2}(n)\mathbf{e}_m(n)\mathbf{e}_m^T(n)$ to obtain (61). $\square$

ACKNOWLEDGMENT

The authors would like to thank Dr. A. Swami of the ARL for interesting discussions and Prof. Z.-Q. Luo of the University of Minnesota for his valuable feedback and for generously providing preprints [12]–[14].

REFERENCES

[1] M. Abdallah and H. Papadopoulos, "Sequential signal encoding and estimation for distributed sensor networks," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, vol. 4, Salt Lake City, UT, May 2001, pp. 2577–2580.

[2] B. Beferull-Lozano, R. L. Konsbruck, and M. Vetterli, "Rate-distortion problem for physics based distributed sensing," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, vol. 3, Montreal, QC, Canada, May 2004, pp. 913–916.

[3] D. Blatt and A. Hero, "Distributed maximum likelihood estimation for sensor networks," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, vol. 3, Montreal, QC, Canada, May 2004, pp. 929–932.

[4] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[5] G. Dahlquist and A. Bjorck, *Numerical Methods*. New York: Prentice-Hall Series in Automatic Computation, 1974.

[6] E. Ertin, R. Moses, and L. Potter, "Network parameter estimation with detection failures," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, vol. 2, Montreal, QC, Canada, May 2004, pp. 273–276.

[7] J. Gubner, "Distributed estimation and quantization," *IEEE Trans. Inf. Theory*, vol. 39, pp. 1456–1459, 1993.

[8] S. M. Kay, *Fundamentals of Statistical Signal Processing—Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[9] ——, *Fundamentals of Statistical Signal Processing—Detection Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1998.

[10] *IEEE Signal Process. Mag. (Special Issue on Collaborative Information Processing)*, vol. 19, Mar. 2002.

[11] W. Lam and A. Reibman, "Quantizer design for decentralized systems with communication constraints," *IEEE Trans. Commun.*, vol. 41, no. 8, pp. 1602–1605, Aug. 1993.

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

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

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

[15] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, "Wireless sensor networks for habitat monitoring," in *Proc. 1st ACM Int. Workshop on Wireless Sensor Netw. Applicat.*, vol. 3, Atlanta, GA, 2002, pp. 88–97.

[16] R. D. Nowak, "Distributed EM algorithms for density estimation and clustering in sensor networks," *IEEE Trans. Signal Process.*, vol. 51, no. 8, pp. 2245–2253, Aug. 2002.

[17] H. Papadopoulos, G. Wornell, and A. 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.

[18] H. C. Papadopoulos, "Efficient digital encoding and estimation of noisy signals," Ph.D. thesis, ECE Dept., MIT, Cambridge, MA, May 1998.

[19] S. S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed compression in a dense microsensor network," *IEEE Signal Process. Mag.*, vol. 19, no. 2, pp. 51–60, Mar. 2002.

[20] M. G. Rabbat and R. D. Nowak, "Decentralized source localization and tracking," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, vol. 3, Montreal, QC, Canada, May 2004, pp. 921–924.

[21] A. Ribeiro and G. B. Giannakis, "Non-parametric distributed quantization-estimation using wireless sensor networks," in *Proc. Int. Conf. Acoust., Speech, Signal Process.* , vol. 4, Philadelphia, PA, Mar. 18–23, 2005, pp. 61–64.

[22] A. Ribeiro and G. B. Giannakis, "Distributed quantization-estimation using wireless sensor networks," in *Proc. Int. Conf. Commun.*, vol. 2, Seoul, Korea, May 16–20, 2005, pp. 730–736.

[23] ——, "Bandwidth-constrained distributed estimation for wireless sensor networks—Part I: Gaussian case," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1131–1143, Mar. 2005.

[24] Y. Sung, L. Tong, and A. Swami, "Asymptotic locally optimal detector for large-scale sensor networks under the Poisson regime," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, vol. 2, Montreal, QC, Canada, May 2004, pp. 1077–1080.

[25] P. K. Varshney, *Distributed Detection and Data Fusion*. New York: Springer-Verlag, 1997.

Alejandro Ribeiro (S'05) received the B.Sc. degree in electrical engineering from the Universidad de la Republica Oriental del Uruguay, Montevideo, in 1998.

Since May 2003, he has been working toward the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis. From 1998 to 2003, he was a member of the Technical Staff at Bellsouth Montevideo. His research interests lie in the areas of communication theory, signal processing, and networking. His current research focuses on wireless cooperative communications, random access, and wireless *ad hoc* and sensor networks.

Mr. Ribeiro is a Fulbright Scholar.

Georgios B. Giannakis (F'97) received the Diploma in electrical engineering from the National Technical University of Athens, Greece, in 1981. He received the M.Sc. degree in electrical engineering in 1983, the M.Sc. degree in mathematics in 1986, and the Ph.D. degree in electrical engineering in 1986, all from the University of Southern California (USC), Los Angeles.

After lecturing for one year at USC, he joined the University of Virginia in 1987, where he became a Professor of Electrical Engineering in 1997. Since 1999, he has been a professor with the Department of Electrical and Computer Engineering, University of Minnesota, where he now holds an ADC Chair in Wireless Telecommunications. His general interests span the areas of communications and signal processing, estimation and detection theory, time-series analysis, and system identification—subjects on which he has published more than 200 journal papers, 350 conference papers, and two edited books. Current research focuses on transmitter and receiver diversity techniques for single- and multiuser fading communication channels, complex-field and space-time coding, multicarrier, ultrawide band wireless communication systems, cross-layer designs, and sensor networks.

Dr. Giannakis is the corecipient of six paper awards from the IEEE Signal Processing (SP) and Communications Societies (1992, 1998, 2000, 2001, 2003, 2004). He also received the SP Society's Technical Achievement Award in 2000. He served as Editor in Chief for the IEEE SIGNAL PROCESSING LETTERS, as Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE SIGNAL PROCESSING LETTERS, as secretary of the SP Conference Board, as a member of the SP Publications Board, as a member and Vice Chair of the Statistical Signal and Array Processing Technical Committee, as Chair of the SP for Communications Technical Committee, and as a member of the IEEE Fellows Election Committee. He has also served as a member of the IEEE-SP Society's Board of Governors, the Editorial Board for the PROCEEDINGS OF THE IEEE, and the steering committee of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.