Distributed Quantization-Estimation Using Wireless Sensor Networks*

Alejandro Ribeiro and Georgios B. Giannakis
Dept. of ECE, University of Minnesota
200 Union Street, Minneapolis, MN 55455 †

Abstract— **Abstract — Wireless sensor networks 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; and iii) 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.**

Keywords — (5) Comm/information theory aspects of sensor networks

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 [5]. The potential of WSNs for surveillance has by now been well appreciated especially in the context of data fusion and distributed detection; e.g., [15], [16] and references therein. However, except for recent works where spatial correlation is exploited to reduce the amount of information exchanged among nodes [1], [3], [6], [11], [12], use of WSNs for the equally important problem of distributed parameter estimation remains largely uncharted. 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 [7], [8], [10]. 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.

In a companion paper we studied estimation of a scalar mean-location parameter in the presence of zero-mean additive white Gaussian noise [14]. We proved that when the dynamic range of the unknown parameter is comparable to the noise standard deviation, estimation based on sign quantization of the original observations exhibits variance almost equal to the variance of the (clairvoyant) estimator based on unquantized observations. We further established that under signal-to-noise ratio (SNR) conditions encountered with WSNs, even a single bit per sensor can have a variance close to the clairvoyant estimator. In this paper, we derive distributed parameter estimators based on binary observations along with their fundamental error-variance limits for more pragmatic signal models. Interestingly, for this class of models 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.

We begin with mean-location parameter estimation in the presence of known univariate but generally non-Gaussian noise 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). 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 IV). Under relaxed conditions, the resultant Maximum Likelihood Estimator (MLE) turns out to be the maximum of a concave function, thus ensuring convergence of Newton-type iterative algorithms. Moreover, we show that by 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 (Section IV-A). Simulations corroborate our theoretical findings in Section V, where we also test them on a motivating application involving distributed parameter estimation with a WSN for measuring vector flow (Section V-B). We conclude the paper in Section VI.

* 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.

† Tel/fax: (612)626-7781/625-4583; Email: {aribeiro, georgios}@ece.umn.edu.

II. Problem Statement

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

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

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 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

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

taking the value 1 when $\mathbf{x}(n)$ belongs to the region $B_k(n) \subset \mathbf{R}^M$, and 0 otherwise. Estimation of $\boldsymbol{\theta}$ will rely on this 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

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

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 Maximum Likelihood Estimators (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 [14][1]:

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

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 [14]

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

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

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

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 $\mathrm{var}(\bar{x}) = \sigma^2/N$ [4, p.31]. Optimizing $B(\theta)$ with respect to Δ_c, yields the optimum at $\Delta_c = 0$ and

$$B_{\min} = \frac{\pi}{2} \frac{\sigma^2}{N}, \qquad (7)$$

the minimum CRLB. Eq. (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 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.

A closer look at $B(\theta)$ in (5) will confirm that the loss can be huge if $\tau_c - \theta \gg 0$. Indeed, as $\tau_c - \theta \rightarrow \infty$ the denominator in (5) goes to zero faster than its numerator, since F_w is the integral of the non-negative pdf p_w; and thus, $B(\theta) \rightarrow \infty$ as $\tau_c - \theta \rightarrow \infty$. 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

[1] Although related results are derived in [14, Prop.1] for Gaussian noise, it is straightforward to generalize the referred proof to cover also non-Gaussian noise pdfs.

figure of merit for estimators based on binary observations is what we will term quantization signal-to-noise ratio (Q-SNR):

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

Notice that contrary to common wisdom, the smaller Q-SNR is, the easier it becomes to select τ_c judiciously and the better performance can be achieved by the estimator in (4), for a given σ, since its CRLB in (5) can be lower.

B. Known Noise pdf with Unknown Variance

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. 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 $\mathrm{E}[w^2(n)] = \sigma^2$. Introducing the standardized variable $v(n) := w(n)/\sigma$ we 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, $\mathrm{E}[v^2(n)] = 1$, 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.

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:

$$B_1(n) := \begin{cases} (\tau_1, \infty) := B_1, & \text{for } n = 0, \ldots, (N/2) - 1 \\ (\tau_2, \infty) := B_2, & \text{for } n = (N/2), \ldots, N. \end{cases} \tag{10}$$

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:

$$q_1(n) := \begin{cases} F_v\left[\frac{\tau_1 - \theta}{\sigma}\right] & := q_1 \quad \text{for } n = 0, \ldots, (N/2) - 1, \\ F_v\left[\frac{\tau_2 - \theta}{\sigma}\right] & := q_2 \quad \text{for } n = (N/2), \ldots, N. \end{cases} \tag{11}$$

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

$$\hat{q}_1 = \frac{2}{N}\sum_{n=0}^{N/2-1} b_1(n), \qquad \hat{q}_2 = \frac{2}{N}\sum_{n=N/2}^{N-1} b_1(n). \tag{12}$$

Mimicking (4), we can invert F_v in (11) 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[2].

Proposition 1 *Consider estimating θ in* (9)*, based on binary observations constructed from the regions defined in* (10).

[2]Omitted due to space considerations, proofs pertaining to claims in this work can be found in [13]

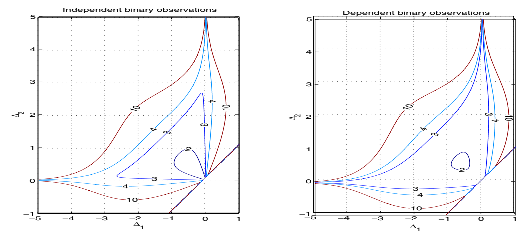

Fig. 1. Per bit CRLB when the binary observations are independent (Section III-B) and dependent (Section III-C), respectively. In both cases, the variance increase with respect to the sample mean estimator is small when the σ-distances are close to 1, being slightly better for the case of dependent binary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

(a) *The MLE of θ is*

$$\hat{\theta} = \frac{F_v^{-1}(\hat{q}_2)\tau_1 - F_v^{-1}(\hat{q}_1)\tau_2}{F_v^{-1}(\hat{q}_2) - F_v^{-1}(\hat{q}_1)}, \tag{13}$$

with F_v^{-1} denoting the inverse function of F_v, and $\hat{q}_1$, $\hat{q}_2$ given by (12).

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

$$B(\theta) := \frac{2\sigma^2}{N}\left(\frac{\Delta_1\Delta_2}{\Delta_2 - \Delta_1}\right)^2 \left[\frac{q_1(1-q_1)}{p_v^2(\Delta_1)\Delta_1^2} + \frac{q_2(1-q_2)}{p_v^2(\Delta_2)\Delta_2^2}\right] \tag{14}$$

where q_k is given by (11)*, and*

$$\Delta_k := \frac{\tau_k - \theta}{\sigma}, \qquad k = 1, 2, \tag{15}$$

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

Eq. (14) 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 (15) are ≈ 1). For the Gaussian pdf, Fig. 1 shows the contour plot of $B(\theta)$ in (14) normalized by $\sigma^2/N := \mathrm{var}(\bar{x})$. Notice that 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

In the previous subsection, we restricted the sensors to transmit only 1 bit 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:

$$\begin{aligned} B_1(n) &:= B_1 = (\tau_1, \infty), \qquad n = 0, 1, \ldots, N-1, \\ B_2(n) &:= B_2 = (\tau_2, \infty), \qquad n = 0, 1, \ldots, N-1, \end{aligned} \tag{16}$$

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 (11).

Note the subtle differences between (10) and (16). While each of the N sensors generates 1 binary observation according to (10), each sensor creates 2 binary observations as per (16). 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 (16), 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* (16) *is given by*

$$\hat{\mathbf{q}} = \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{b}(n). \tag{17}$$

Interestingly, (17) coincides with (12), proving that the corresponding estimators of θ are identical; i.e., (13) 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* (16)*. The variance of any unbiased estimator of θ,* $\mathrm{var}(\hat{\theta})$*, based on* $\{b_1(n), b_2(n)\}_{n=0}^{N-1}$ *is bounded by*

$$B_D(\theta) := \frac{\sigma^2}{N} \left(\frac{\Delta_1 \Delta_2}{\Delta_2 - \Delta_1} \right)^2 \left[\frac{q_1\,(1-q_1)}{p_v^2(\Delta_1)\Delta_1^2} + \frac{q_2\,(1-q_2)}{p_v^2(\Delta_2)\Delta_2^2} - \frac{q_2\,(1-q_1)}{p_v(\Delta_1)p(\Delta_2)\Delta_1\Delta_2} \right], \tag{18}$$

where the subscript D in $B_D(\theta)$ is used as a mnemonic for the dependent binary observations this estimator relies on [c.f. (14)*].*

Unexpectedly, (18) is similar to (14). 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,

$$C(\theta) = NB(\theta), \qquad C_D(\theta) = N \log_2(3) B_D(\theta)\ , \tag{19}$$

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 (14), (18) and (19) 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.

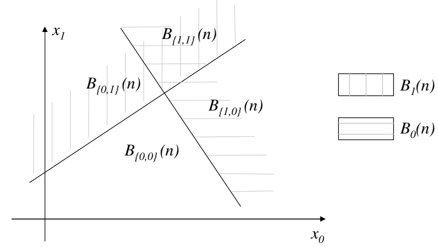

Fig. 2. The vector of binary observations $\mathbf{b}$ takes on the value $\{y_1, y_2\}$ if and only if $x(n)$ belongs to the region $B_{\{y_1,y_2\}}$.

IV. 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

$$\mathcal{B} := \{\boldsymbol{\beta} \in \mathbf{R}^K \mid [\boldsymbol{\beta}]_k \in \{0,1\},\ k = 1, \ldots, K\}, \tag{20}$$

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

$$\mathbf{B}_\beta(n) := \bigcap_{[\boldsymbol{\beta}]_k = 1} B_k(n) \bigcap_{[\boldsymbol{\beta}]_k = 0} \bar{B}_k(n), \tag{21}$$

where $\bar{B}_k(n)$ denotes the set-complement of $B_k(n)$ in $\mathbf{R}^M$. Note that the definition in (21) implies that $x(n) \in \mathbf{B}_\beta(n)$ if and only if $\mathbf{b}(n) = \boldsymbol{\beta}$; see also Fig. 2 for an illustration in $\mathbf{R}^2$ ($M = 2$). The corresponding probabilities are:

$$\begin{aligned} q_\beta(n) \quad &:= \quad \Pr\{\mathbf{b}(n) = \boldsymbol{\beta}\} \;=\; \Pr\{\mathbf{x}(n) \in \mathbf{B}_\beta(n)\} \\ &= \quad \int_{\mathbf{B}_\beta(n)} p_{\mathbf{w}}[\mathbf{u} - \mathbf{f}_n(\boldsymbol{\theta}); \boldsymbol{\psi}]\ d\mathbf{u}, \end{aligned} \tag{22}$$

with $\mathbf{f}_n$ as in (1), and $\boldsymbol{\psi}$ containing the unknown parameters of the known noise pdf. Using definitions (22) and (20), we can write the pertinent log-likelihood function as

$$L(\boldsymbol{\theta}, \boldsymbol{\psi}) = \sum_{n=0}^{N-1} \sum_{y \in \mathcal{B}} \delta(\mathbf{b}(n) - \boldsymbol{\beta}) \ln q_\beta(n), \tag{23}$$

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

$$\hat{\boldsymbol{\theta}} = \arg\max_{(\boldsymbol{\theta}, \boldsymbol{\psi})} L(\boldsymbol{\theta}, \boldsymbol{\psi})\ . \tag{24}$$

The nonlinear search needed to obtain $\hat{\boldsymbol{\theta}}$ could be challenging. Fortunately, as the following proposition asserts, 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.

Proposition 3 *If the MLE problem in* (24) *satisfies the conditions:*

- [c1] *The noise pdf* $p_{\mathbf{w}}(\mathbf{w}; \boldsymbol{\psi}) \leftrightarrow p_{\mathbf{w}}(\mathbf{w})$ *is log-concave [2, 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.*

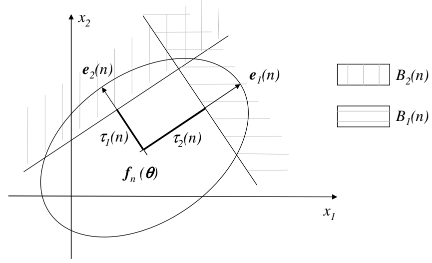

Fig. 3. Selecting the regions $B_k(n)$ perpendicular to the covariance matrix eigenvectors results in independent binary observations.

then $L(\boldsymbol{\theta})$ in (23) *is a concave function of $\boldsymbol{\theta}$.*

Note that [c1] is satisfied by common noise pdfs, including the multivariate Gaussian [2, 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.

A. Colored Gaussian Noise

Analyzing the performance of the MLE in (24) 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:

$$\mathbf{C}(n) = \sum_{m=1}^{M} \sigma_m^2(n)\mathbf{e}_m(n)\mathbf{e}_m^T(n). \tag{25}$$

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

$$B_k(n) = \{\mathbf{x} \in \mathbf{R}^M \mid \mathbf{e}_k^T(n)\mathbf{x} \geq \tau_k(n)\}, \quad k = 1, \ldots, K = M, \tag{26}$$

Fig (3) depicts the regions $B_k(n)$ in (26) 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 $\boldsymbol{\theta}$.

Herein, the Bernoulli parameters $q_k(n)$ take on a particularly simple form in terms of the Gaussian tail function,

$$\begin{aligned} q_k(n) &= \int_{\mathbf{e}_k^T(n)\mathbf{u} \geq \tau_k(n)} p_{\mathbf{w}}(\mathbf{u} - \mathbf{f}_n(\boldsymbol{\theta}))\ d\mathbf{u} \\ &= Q\left(\frac{\tau_k(n) - \mathbf{e}_k^T(n)\mathbf{f}_n(\boldsymbol{\theta})}{\sigma_k(n)}\right) := Q(\Delta_k(n)), \end{aligned} \tag{27}$$

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

Due to the independence among binary observations $p(\mathbf{b}(n)) = \prod_{k=1}^{K}[q_k(n)]^{b_k(n)}[1 - q_k(n)]^{1-b_k(n)}$, leading to

$$L(\boldsymbol{\theta}) = \sum_{n=0}^{N-1}\sum_{k=1}^{K} b_k(n)\ln q_k(n) + [1 - b_k(n)]\ln[1 - q_k(n)], \tag{28}$$

whose NK *independent* summands replace the $N2^K$ *dependent* summands in (23).

Since the regions $B_k(n)$ are half-spaces, Proposition 3 applies to the maximization of (28) and guarantees that the numerical search for the $\hat{\boldsymbol{\theta}}$ estimator in (28) 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 (24), tractable. In particular, it is possible to derive a closed-form expression for the Fisher Information Matrix (FIM) [4, p.44], as we establish next.

Proposition 4 *The FIM, $\mathbf{I}$, for estimating $\boldsymbol{\theta}$ based on the binary observations obtained from the regions defined in* (26)*, is given by*

$$\mathbf{I} = \sum_{n=0}^{N-1}\mathbf{J}_n^T\left[\sum_{k=1}^{K}\frac{e^{-\Delta_k^2(n)}\mathbf{e}_k(n)\mathbf{e}_k^T(n)}{2\pi\sigma_k^2(n)Q(\Delta_k(n))[1 - Q(\Delta_k(n))]}\right]\mathbf{J}_n, \tag{29}$$

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

Inspection of (29) shows that the variance of the MLE in (24) 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 [13, Apx. D]),

$$\mathbf{I}_x = \sum_{n=0}^{N-1}\mathbf{J}_n^T\left[\sum_{m=1}^{M}\frac{\mathbf{e}_m(n)\mathbf{e}_m^T(n)}{\sigma_m^2(n)}\right]\mathbf{J}_n^T. \tag{30}$$

If we define the equivalent noise powers as

$$\rho_k^2(n) := \frac{2\pi Q(\Delta_k(n))[1 - Q(\Delta_k(n))]}{e^{-\Delta_k^2(n)}}\sigma_k^2(n), \tag{31}$$

we can rewrite (29) in the form

$$\mathbf{I} = \sum_{n=0}^{N-1}\mathbf{J}_n^T\left[\sum_{k=1}^{K}\frac{\mathbf{e}_k(n)\mathbf{e}_k^T(n)}{\rho_k^2(n)}\right]\mathbf{J}_n^T, \tag{32}$$

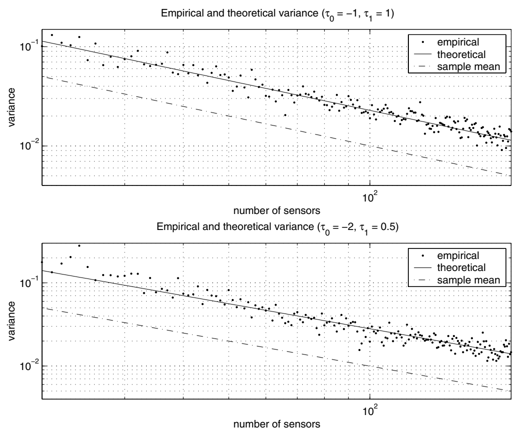

Fig. 4. Noise of unknown power estimator. The simulation corroborates the close to clairvoyant variance prediction of (14) ($\sigma = 1$, $\theta = 0$, Gaussian noise)

which except for the noise powers has form identical to (30). Thus, comparison of (32) with (30) 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 (31) 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

$$\rho_k^2(n) = \frac{\pi}{2}\sigma_k^2(n). \tag{33}$$

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 (24), 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.

V. Simulations

A. Scalar parameter estimation

We begin by simulating the estimator in (13) for scalar parameter estimation in the presence of AWGN with unknown variance. Results are shown in Fig. 4 for two different sets of σ-distances, Δ_1, Δ_2, corroborating the values predicted by (14) and the fact that the performance loss with respect to the clairvoyant sample mean estimator, $\bar{x}$, is indeed small.

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 IV-A. Suppose we wish to estimate a vector flow using incidence observations. With reference to Fig. 5, consider the flow vector $\mathbf{v} := (v_0, v_1)^T$, and a sensor positioned at an angle $\phi(n)$

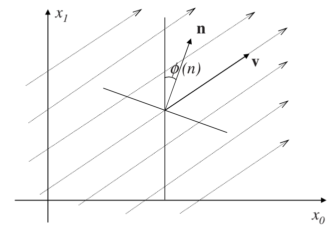

Fig. 5. The vector flow $\mathbf{v}$ incises over a certain sensor capable of measuring the normal component of $\mathbf{v}$.

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

$$x(n) := \langle \mathbf{v}, \mathbf{n}\rangle + w(n) = v_0 \sin[\phi(n)] + v_1 \cos[\phi(n)] + w(n), \tag{34}$$

where $\langle , \rangle$ denotes inner product, $w(n)$ is zero-mean AWGN, and $n = 0, 1, \ldots, N-1$ is the sensor index. The model (34) applies to the measurement of hydraulic fields, pressure variations induced by wind and radiation from a distant source [9].

Estimating $\mathbf{v}$ fits the framework of Section IV-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 (32) is easily found to be

$$\mathbf{I} = \sum_{n=0}^{N-1} \frac{1}{\rho_1^2(n)} \begin{pmatrix} \sin^2[\phi(n)] & \sin[\phi(n)]\cos[\phi(n)] \\ \sin[\phi(n)]\cos[\phi(n)] & \cos^2[\phi(n)] \end{pmatrix}. \tag{35}$$

Furthermore, since $x(n)$ in (34) is linear in $\mathbf{v}$ and the noise pdf is log-concave (Gaussian) the log-likelihood function is concave as asserted by Proposition 3.

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:

$$\bar{\mathbf{I}} = \frac{2}{\pi\sigma^2} \begin{pmatrix} N/2 & 0 \\ 0 & N/2 \end{pmatrix}. \tag{36}$$

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

$$\mathrm{var}(v_0) = \mathrm{var}(v_1) = \frac{\pi\sigma^2}{N}. \tag{37}$$

Fig. 6 depicts the bound (37), 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.

VI. Conclusions

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

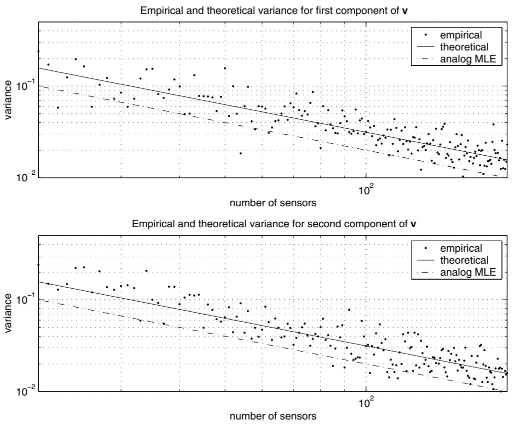

Fig. 6. Average variance for the components of $\mathbf{v}$. The empirical as well as the bound (37) are compared with the analog observations based MLE ($\mathbf{v} = (1, 1)$, $\sigma = 1$).

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 maximum likelihood estimators (MLEs) for scalar mean-location parameters in the presence of generally non-Gaussian noise when the noise pdf is completely known, and when the pdf is known except for a number of unknown parameters.

In both 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[3].

[3] 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.

REFERENCES

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

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

[3] E. Ertin, R. Moses, and L. Potter, "Network parameter estimation with detection failures," in *Proc. of the Intnl. Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 273–276, Montreal, Canada, May 2004.

[4] S. M. Kay, *Fundamentals of Statistical Signal Processing - Estimation Theory*. Prentice Hall, 1993.

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

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

[7] 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.

[8] Z.-Q. Luo, "Universal Decentralized Estimation in a Bandwidth Constrained Sensor Network," *IEEE Transactions on Information Theory*, November 2003 (submitted). available at http://www.ece.umn.edu/users/luozq/recent_work.html.

[9] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, "Wireless sensor networks for habitat monitoring," in *Proc. of the 1st ACM International Workshop on Wireless Sensor Networks and Applications*, vol. 3, pp. 88–97, Atlanta, Georgia, 2002.

[10] 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.

[11] S. S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed compression in a dense microsensor network," *IEEE Signal Processing Magazine*, vol. 19, pp. 51–60, March 2002.

[12] M. G. Rabbat and R. D. Nowak, "Decentralized source localization and tracking," in *Proc. of the 2004 IEEE Intnl. Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 921–924, Montreal, Canada, May 2004.

[13] A. Ribeiro and G. B. Giannakis, "Bandwidth-Constrained Distributed Estimation for Wireless Sensor Networks, Part II: Unknown pdf," *IEEE Transactions on Signal Processing*, submitted August 2004. available at http://www.ece.umn.edu/users/aribeiro/Research/Publications.html.

[14] 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/Publications.html.

[15] Y. Sung, L. Tong, and A. Swami, "Asymptotic locally optimal detector for large-scale sensor networks under the Poisson regime," in *Proc. of the International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 1077–1080, Montreal, Canada, May 2004.

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