

Bandwidth-Constrained MAP Estimation for Wireless Sensor Networks

S. Faisal Ali Shah, Alejandro Ribeiro and Georgios B. Giannakis

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

Abstract—We deal with distributed parameter estimation algorithms for use in wireless sensor networks (WSNs) with a fusion center when only quantized observations are available due to power/bandwidth constraints. The main goal of the paper is to design efficient estimators when the parameter can be modelled as random with a priori information. In particular, we develop maximum a posteriori (MAP) estimators for distributed parameter estimation and formulate the problem under different scenarios. We show that the pertinent objective function is concave and hence, the corresponding MAP estimator can be obtained efficiently through simple numerical maximization algorithms.

I. INTRODUCTION

Characterized by small low-cost, low-power devices equipped with limited sensing, computational and communication capabilities, wireless sensor networks (WSNs) are well motivated for environmental monitoring, industrial instrumentation and surveillance applications [7]. The distributed topology of these networks along with their limited power budget and communication resources gave rise to the area of collaborative signal and information processing [3].

Bandwidth-constrained distributed estimation arises when deploying a WSN for monitoring, in which case the estimation of certain parameters of interest necessitates collection of different sensor estimates. As sensor observations have to be quantized, WSN-based parameter estimators must rely on (perhaps severely) quantized observations [4], [5]. Interestingly, it has been shown in [6] that when the noise variance is comparable to the parameter's dynamic range, even quantization to a single bit per observation leads to a small penalty in estimation variance when comparing maximum likelihood (ML) estimators based on quantized (binary) versus original (analog-amplitude) observations. A characteristic common to all these works is the assumption that some prior information is available, at least to bound the range of possible parameter values. This suggests naturally, a connection with Bayesian estimation.

Building on this observation, the present paper addresses the problem of maximum a posteriori (MAP) estimation based on binary observations, and shows that the graceful

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.

Email at {sfaisal, aribeiro, georgios}@ece.umn.edu

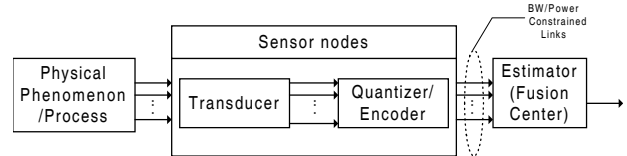


Fig. 1. Block diagram of wireless sensor network

performance degradation of ML estimators extends to their MAP counterparts. Moreover, we establish that while MAP estimators cannot be expressed in closed form, they are found as the maximum of a concave function; thus ensuring the convergence of fast descent algorithms e.g., Newton's method.

II. PROBLEM FORMULATION

Consider a physical phenomenon characterized by a set of P parameters that we lump in vector form as $\theta = [\theta_1, \dots, \theta_P]^T$. From available a priori knowledge, θ is modelled as a random vector parameter with prior probability density function (pdf) $p_\theta(\theta)$ and mean $\langle \theta \rangle = \mu_\theta$.

For measuring θ , we deploy a WSN composed of N sensors $\{S_n\}_{n=0}^{N-1}$, with each sensor observing θ through a linear transformation

$$\mathbf{x}(n) = \mathbf{H}\theta + \mathbf{w}(n), \quad (1)$$

where $\mathbf{x}(n) = [x_1(n), \dots, x_K(n)]^T \in \mathbf{R}^K$ is the measurement vector at sensor S_n , $\mathbf{w}(n) \in \mathbf{R}^K$ is zero-mean additive noise with pdf $p_{\mathbf{w}}(\mathbf{w})$ and the matrix $\mathbf{H} \in \mathbf{R}^{K \times P}$. We denote the vector formed by concatenating all the observations $\{\mathbf{x}(n)\}_{n=0}^{N-1}$ of N sensors as $\mathbf{x}_{0:N-1}$. For simplicity of exposition, we assume that \mathbf{H} and $p_{\mathbf{w}}(\mathbf{w})$ are constant across sensors, and we also assume that $\mathbf{w}(n_1)$ is independent of $\mathbf{w}(n_2)$ for $n_1 \neq n_2$.

A clairvoyant (CV) benchmark for estimators based on binary observations corresponds to having all analog-amplitude observations $\mathbf{x}_{0:N-1}$ be available at the fusion center. In this case, a possible approach to estimate θ is the MAP estimator [2]

$$\theta_{CV} = \underset{\theta}{\text{arg max}} \{p(\theta | \mathbf{x}_{0:N-1})\}, \quad (2)$$

where $p(\theta | \mathbf{x}_{0:N-1})$ is the conditional pdf of θ given $\mathbf{x}_{0:N-1}$.

As discussed earlier, power and bandwidth constraints, dictate the need of a quantizer mapping the analog-amplitude observations $\mathbf{x}(n)$ to a finite set:

$$\mathbf{b}(n) = \mathbf{q}(\mathbf{x}(n)), \quad \mathbf{q} : \mathbf{R}^K \rightarrow \{-1, 1\}^K, \quad (3)$$

with $\mathbf{b}(n) = [b_1(n), \dots, b_K(n)]^T$ a K -component binary message. Note that implicit to (3) is the fact that we are restricting the sensors to transmit one bit per scalar observation. Similar to $\mathbf{x}_{0:N-1}$, we define the binary observation of N sensors as $\mathbf{b}_{0:N-1}$.

The problem addressed in this paper is the MAP estimation of θ based on the binary messages $\mathbf{b}_{0:N-1}$, where the MAP estimator is defined as [2]

$$\theta_{\text{MAP}} = \arg \max_{\theta} \{ p(\theta | \mathbf{b}_{0:N-1}) \} \quad (4)$$

$$= \arg \max_{\theta} \{ p(\mathbf{b}_{0:N-1} | \theta) + p(\theta) \},$$

with $p(\theta | \mathbf{b}_{0:N-1})$ and $p(\mathbf{b}_{0:N-1} | \theta)$ are the conditional pdfs of θ given $\mathbf{b}_{0:N-1}$ and of $\mathbf{b}_{0:N-1}$ given θ , respectively. Note that to obtain the second equality in (4) we used Bayes' rule, the monotonicity of the logarithm and eliminated a normalizing constant that does not depend on θ .

The goals of this paper are to: i) derive quantization functions $\mathbf{q}(\mathbf{x}(n))$; ii) establish that in many cases of interest θ_{MAP} in (4) can be found as the maximum of a concave function; and iii) show that the estimation variance of θ_{MAP} in (4) based on binary observations comes close to the estimation variance of θ_{CV} in (2) based on analog-amplitude observations.

III. MAP ESTIMATOR - SCALAR PARAMETERS

The MAP formulation allows us to tackle parameter estimation problems in different scenarios. First simple situation is when the parameter $\theta \leftrightarrow \theta$ and the observations $x(n) \leftrightarrow \mathbf{x}(n)$ are scalar. In this case, we define the message as

$$b(n) = \text{sgn}(x(n) - h\mu_\theta) \quad (5)$$

where the function $\text{sgn}(\cdot)$ observes the sign of its argument. The estimator θ_{MAP} in (4) can then be written in terms of the prior pdf $p_\theta(\theta)$ and the cumulative distribution function (CDF) of the noise and can be found with robust numerical algorithms as asserted by the following proposition.

Proposition 1 Consider a random scalar parameter θ with prior distribution $p_\theta(\theta)$, an observation model as in (1), and binary messages $\mathbf{b}_{0:N-1}$ defined as in (5). Then,

- (a) if $F_w(w)$ denotes the CDF of the noise variables $w(n)$ the MAP estimator of θ given $\mathbf{b}_{0:N-1}$ is

$$\theta_{\text{MAP}} = \arg \max_{\theta} \left\{ \sum_{n=0}^{N-1} F_w(hb(n)(\theta - \mu_\theta) + p_\theta(\theta)) \right\} \quad (6)$$

$$= \arg \max_{\theta} L(\theta)$$

and;

- (b) if $F_w(\cdot)$ and $p_\theta(\cdot)$ are log-concave functions [1, p. 104], then $L(\theta)$ is a concave function of θ .

Proof: Writing (4) for scalar parameter θ

$$\theta_{\text{MAP}} = \arg \max_{\theta} \{ p(\theta | \mathbf{b}_{0:N-1}) \} \quad (7)$$

$$= \arg \max_{\theta} \{ p(\mathbf{b}_{0:N-1} | \theta) + p(\theta) \},$$

where we ignored the term containing $p(\mathbf{b}_{0:N-1})$ as it is independent of θ . Making use of the independence among the observations, we can write $p(\mathbf{b}_{0:N-1} | \theta) = \prod_{n=0}^{N-1} p(b(n) | \theta)$.

The value of $p(b(n) | \theta)$ can be obtained by explicit enumeration:

$$p(b(n) | \theta) = \begin{cases} F_w(h(\theta - \mu_\theta)) & \text{if } b(n) = 1 \\ 1 - F_w(h(\theta - \mu_\theta)) & \text{if } b(n) = 0 \end{cases}$$

Combining the above two cases, we can write

$$p(b(n) | \theta) = F_w(hb(n)(\theta - \mu_\theta)) \quad (8)$$

Substituting (8) into (7) the result in (6) follows.

The claim in Proposition 1-[b] follows from the fact that the sum of concave functions is also concave. ■

The value of Proposition 1 is not as much in giving an expression for θ_{MAP} (Proposition 1-[a]) as in establishing that the function $L(\theta)$ is concave (Proposition 1-[b]). The latter assures efficient numerical implementation of our algorithm as discussed in the following remark.

Remark 1 The numerical search needed to obtain θ_{MAP} could be challenged either by the multimodal nature of $L(\theta)$ or by numerical ill-conditioning caused by e.g., saddle points. But when the log-concavity conditions in Proposition 1-[b] are satisfied, computationally efficient search algorithms like e.g. Newton's method are guaranteed to converge to the global maximum [1, Chap. 2].

A special case of a log-concave pdf is the Gaussian one. Consequently, in the frequently encountered case of a Gaussian prior $p_\theta(\theta) = \frac{1}{\sqrt{\pi}\sigma_\theta} \exp(-(\theta - \mu_\theta)^2 / \sigma_\theta^2)$ and Gaussian noise $p_w(w) = \frac{1}{\sqrt{\pi}\sigma_w} \exp(-w^2 / \sigma_w^2)$ the MAP estimator can be expressed as

$$\theta_{\text{MAP}} = \arg \max_{\theta} \left\{ \sum_{n=0}^{N-1} Q\left(\frac{hb(n)(\mu_\theta - \theta)}{\sigma_w}\right) - \frac{(\theta - \mu_\theta)^2}{2\sigma_\theta^2} \right\}, \quad (9)$$

where $Q(x) = \int_x^\infty \frac{1}{\sqrt{\pi}} \exp(-u^2) du$ is the Gaussian tail function. Moreover, descent methods e.g., Newton's algorithm are guaranteed to converge to the global maximum.

Other log-concave pdfs are the uniform distribution in a convex set and the members of the generalized Gaussian family.

IV. MAP ESTIMATOR - VECTOR PARAMETERS

Results in Section III can be extended to cover the general vector parameter - vector observation model in (1). Start by considering the case of white Gaussian noise; i.e., $\mathbf{w}(n) \sim \mathcal{N}(\mathbf{0}, \sigma_w^2 \mathbf{I})$. In this case, we write $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_K]^T$ and define the components of the message $\mathbf{b}(n)$ as

$$b_k(n) = \text{sgn}(x_k(n) - \mathbf{h}_k^T \boldsymbol{\mu}_\theta), \quad (10)$$

for $k \in \{1, \dots, K\}$. The resemblance with the problem of Section III is clear and not surprisingly the following proposition holds true.

Proposition 2 Consider a vector parameter $\boldsymbol{\theta}$, with log-concave prior distribution $p_\theta(\boldsymbol{\theta})$; an observation model as in (1) with $p_w(\mathbf{w})$ white Gaussian with variance $\sigma_w^2 \mathbf{I}$; and binary messages $\mathbf{b}_{0:N-1}$ as

in (10). Then, if we define the per sensor likelihood $L_n(\boldsymbol{\theta})$ as

$$L_n(\boldsymbol{\theta}) = \sum_{k=1}^K Q\left(\frac{b_k(n)\mathbf{h}_k^T \boldsymbol{\mu}_\theta - \boldsymbol{\theta}}{\sigma_w}\right), \quad (11)$$

we have that:

(a) The MAP estimator of $\boldsymbol{\theta}$ based on $\mathbf{b}_{0:N-1}$ is given by

$$\boldsymbol{\theta}_{\text{MAP}} = \underset{\boldsymbol{\theta}}{\text{argmax}} \left[\sum_{n=0}^{N-1} L_n(\boldsymbol{\theta}) \right] + p_\theta(\boldsymbol{\theta}). \quad (12)$$

(b) The likelihood $L(\boldsymbol{\theta})$ is a concave function of $\boldsymbol{\theta}$.

Let us note that the multivariate Gaussian belongs to the log-concave class of pdfs. Indeed, if $\boldsymbol{\theta}$ has a Gaussian prior distribution with covariance matrix $\boldsymbol{\theta}\boldsymbol{\theta}^T \quad \mathbf{C}_\theta$ then

$$p_\theta(\boldsymbol{\theta}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta)^T \mathbf{C}_\theta^{-1}(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta)\right), \quad (13)$$

with $C = \frac{1}{2} (\pi)^P (\mathbf{C}_\theta)$. The expression in (13) is a quadratic expression on $\boldsymbol{\theta}$ which is concave for positive semidefinite \mathbf{C}_θ .

Proposition 2 establishes that at least for white Gaussian noise the comments in Remark 1 also hold true for vector parameters and observations. In the coming section we show that the same result can be obtained for colored noise.

A. Colored Gaussian Noise

Consider again the model in (1) with $\mathbf{w}(n)$ colored Gaussian with covariance matrix $\mathbf{w}(n)\mathbf{w}^T(n) \quad \mathbf{C}(n)$ at sensor S_n . In this case the components of the observation $\mathbf{x}(n)$ are not independent; hence, the components of $\mathbf{b}(n)$ are not and the conditional probabilities in (4) are difficult to obtain in general.

Nonetheless, there is a particular definition of $\mathbf{b}(n)$ that results in independent components. Consider the eigenvector decomposition of $\mathbf{C}(n)$,

$$\mathbf{C}(n) = \mathbf{V}(n) \text{diag}\{\sigma_1^2(n), \dots, \sigma_K^2(n)\} \mathbf{V}^T(n), \quad (14)$$

where $\{\sigma_k^2(n)\}_{k=1}^K$ are the eigenvalues of $\mathbf{C}(n)$ and $\mathbf{V}(n) = [\mathbf{v}_1(n) \dots \mathbf{v}_K(n)]$ is the matrix of eigenvectors such that $\mathbf{v}_i^T(n)\mathbf{C}(n)\mathbf{v}_j(n) = \sigma_i^2(n)\delta_{ij}$. If we define the k^{th} component of $b_k(n)$ as

$$b_k(n) = \text{sgn}\left[\mathbf{v}_k^T(n)\{\mathbf{x}(n) - \mathbf{H}\boldsymbol{\mu}_\theta\}\right], \quad (15)$$

the resultant binary observations are independent when $\boldsymbol{\theta}$ is given as we assert in the following lemma:

Lemma 1 Consider the model in (1) with $\mathbf{w}(n)$ Gaussian with covariance matrix $\mathbf{w}(n)\mathbf{w}^T(n) \quad \mathbf{C}(n)$. Then, binary observations defined as in (15) are independent for $k_1 \neq k_2$:

$$\{b_{k_1}(n), b_{k_2}(n) | \boldsymbol{\theta}\} = \{b_{k_1}(n) | \boldsymbol{\theta}\} \{b_{k_2}(n) | \boldsymbol{\theta}\} \quad (16)$$

Proof: Define $u_k(n) = \mathbf{v}_k^T(n)\mathbf{w}(n)$ and note that the distribution of $b_k(n)$ is given by [c.f (1), (15)]

$$b_k(n) = \begin{cases} 1 & \text{if } u_k(n) \geq \mathbf{v}_k^T(n)\mathbf{H}(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta) \\ -1 & \text{otherwise} \end{cases} \quad (17)$$

On the other hand, note that $u_{k_1}(n)$ and $u_{k_2}(n)$ are independent for $k_1 \neq k_2$. Indeed, since $u_k(n)$ is normally distributed it suffices to prove that $u_{k_1}(n)$ and $u_{k_2}(n)$ are uncorrelated:

$$\begin{aligned} u_{k_1}(n)u_{k_2}(n) &= \mathbf{v}_{k_1}^T(n) \mathbf{w}(n)\mathbf{w}^T(n) \mathbf{v}_{k_2}(n) \\ &= \mathbf{v}_{k_1}^T(n)\mathbf{C}(n)\mathbf{v}_{k_2}(n) \\ &= \sigma_{k_1}^2(n)\delta(k_1 - k_2) \end{aligned} \quad (18)$$

Lemma 1 follows because $b_{k_1}(n)$ and $b_{k_2}(n)$ are functions of the independent random variables $u_{k_1}(n)$ and $u_{k_2}(n)$ [c.f. (16) and (18)]. ■

The independence of the binary observations dictates that the per-sensor log-likelihood can be written as

$$L_n(\boldsymbol{\theta}) = \sum_{k=1}^K Q\left(\frac{b_k(n)\mathbf{v}_k^T \mathbf{H} \boldsymbol{\mu}_\theta - \boldsymbol{\theta}}{\sigma_k(n)}\right). \quad (19)$$

From where a proposition analogous to Proposition 2 follows readily.

Proposition 3 Consider the setting of Proposition 2 with $p_w(\mathbf{w})$ colored Gaussian with covariance matrix $\mathbf{w}(n)\mathbf{w}^T(n) \quad \mathbf{C}(n)$ at sensor S_n ; binary messages $\mathbf{b}_{0:N-1}$ as in (15); and per-sensor log-likelihood $L_n(\boldsymbol{\theta})$ as in (19). Then, results [a] and [b] of Proposition 2 hold true.

Propositions 1-3 establish that for many pragmatic signal models MAP estimation from binary observations can be posed as a convex optimization problem with virtues summarized in Remark 1.

V. MEAN-SQUARE ERROR (MSE) ANALYSIS

In this section, we study the increase in mean-square error (MSE) when binary observations are used in lieu of the analog-amplitude observations. For estimation of random parameters, bounds on the MSE can be obtained by computing the pertinent Fisher Information Matrix (FIM) \mathbf{J} that we write as the sum of two parts [8, p. 84]:

$$\mathbf{J} = \mathbf{J}_D + \mathbf{J}_P, \quad (20)$$

where \mathbf{J}_D represents information obtained from the data, and \mathbf{J}_P captures *a priori* information. The MSE of the i^{th} component of $\boldsymbol{\theta}$ is bounded by the i^{th} diagonal element of \mathbf{J} ,

$$(\theta_i) \geq [\mathbf{J}^{-1}]_{ii}. \quad (21)$$

Also, note that for any FIM, $[\mathbf{J}^{-1}]_{ii} \geq 1/[\mathbf{J}]_{ii}$ [2]. This property yields a different bound on (θ_i) ,

$$(\theta_i) \geq \frac{1}{[\mathbf{J}]_{ii}}, \quad (22)$$

which is easier to compute although not tight in general.

The following proposition states a bound (exact value) on $[\mathbf{J}]_{ii}$ when binary (analog-amplitude) observations are used.

Proposition 4 Consider the signal model in (1) with $\mathbf{w}(n)$ white Gaussian with covariance matrix $\mathbf{w}(n)\mathbf{w}^T(n) \quad \sigma_w^2 \mathbf{I}$ and Gaussian prior distribution with covariance $\boldsymbol{\theta}\boldsymbol{\theta}^T \quad \mathbf{C}_\theta$. Write (1) componentwise as $\mathbf{x}(n)$ is $x_k(n) = \mathbf{h}_k^T \boldsymbol{\theta} +$

$w_k(n)$. Then, the i^{th} diagonal element of the FIM \mathbf{J} in (20) satisfies:

(a) when binary observations as in (10) are used,

$$\mathbf{J}_{ii} \geq \frac{N}{\pi\sigma_w} \sum_{k=1}^K \frac{h_{ki}^2}{\sqrt{\sigma_w^2 + \mathbf{h}_k^T \mathbf{C}_\theta \mathbf{h}_k}} + [\mathbf{C}_\theta^{-1}]_{ii}, \quad (23)$$

(b) when analog-amplitude observations are used,

$$\mathbf{J}_{CV\ ii} = \frac{N}{\sigma_w^2} \sum_{k=1}^K h_{ki}^2 + [\mathbf{C}_\theta^{-1}]_{ii}. \quad (24)$$

Proof: See Appendix A.

For the clairvoyant FIM, \mathbf{J}_{CV} , the bound derived above is essentially the same as the one obtained for MMSE estimators in [2]. Indeed, the MSE for the Gaussian prior - Gaussian noise considered in Proposition 4 is

$$(\theta_i) \left[\left(\frac{\mathbf{H}^T \mathbf{H} + \mathbf{C}_\theta^{-1}}{\sigma_w^2} \right)^{-1} \right]_{ii}. \quad (25)$$

Comparing (25) with (21), we see that the term inside the parenthesis in (25) is the FIM in (24).

For the bound when binary observations are used, corroborating simulations in Section VI prove that it is actually tight. Consequently, it provides a good theoretical means to characterize the performance of estimators based on binary observations.

Remark 2 As a figure of merit, consider the MSEs when estimating θ_i from a set of quantized observations $\mathbf{b}_{0:N-1}$ and when using analog-amplitude observations $\mathbf{x}_{0:N-1}$. We define the average information loss \mathcal{L}_i as

$$\mathcal{L}_i = \frac{(\theta_i)}{(\theta_{i,CV})} \approx \frac{\mathbf{J}_{CV\ ii}}{\mathbf{J}_{ii}}. \quad (26)$$

For large N , the second terms in (23) and (24) can be neglected and the information loss can be approximated as:

$$\mathcal{L}_i \approx \frac{\pi}{\sqrt{\frac{\mathbf{h}^T \mathbf{C}_\theta \mathbf{h}}{\sigma_w^2} + \gamma}} = \frac{\pi}{\sqrt{1 + \gamma}}, \quad (27)$$

where we defined $\gamma = \frac{\mathbf{h}^T \mathbf{C}_\theta \mathbf{h}}{\sigma_w^2}$ which is the average signal to noise ratio (SNR) of the observations $\mathbf{x}(n)$. Note that as $\gamma \rightarrow \infty$ the information loss $\mathcal{L}_i \rightarrow \pi/2$, corroborating results in [6] for deterministic parameter estimation. In any event, we stress the remarkable fact that for low to medium SNR γ , the average information loss \mathcal{L}_i is a small factor.

VI. NUMERICAL RESULTS

Consider using the MAP estimator in (12) to estimate a two-dimensional vector parameter. The random vector $\boldsymbol{\theta} = [\theta_1 \ \theta_2]^T$ is assumed to have zero mean and diagonal covariance matrix $\mathbf{C}_\theta = \sigma_\theta^2 \mathbf{I}$. Our goal is to find the MSE of $\boldsymbol{\theta}$ as $(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^2$ through simulations and compare it with the analytical results of Section V.

We first study the effect of the number of sensors N on the MSE with results shown in Fig. 2. Simulated MSE values are scattered around analytical corroborating our approximate

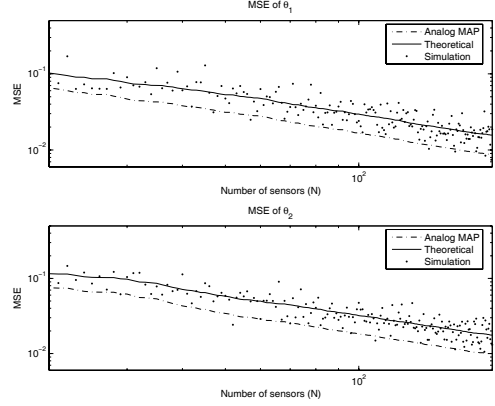


Fig. 2. MSE of MAP estimator vs. number of sensors ($\gamma = 0.3$; $\mathbf{h} = [\cos \psi \ \sin \psi]$ with $\psi \in [-\pi, \pi]$)

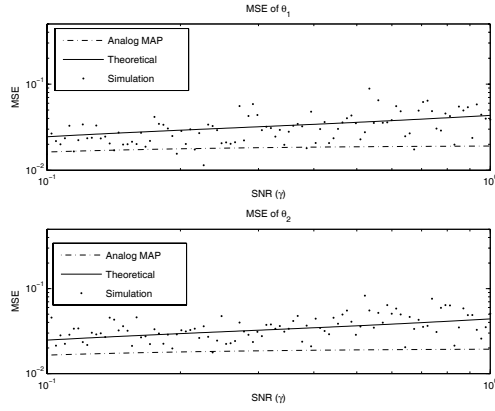


Fig. 3. MSE of MAP estimator vs. SNR ($N = 100$; $\mathbf{h} = [\cos \psi \ \sin \psi]$ with $\psi \in [-\pi, \pi]$)

analysis. We also compare the results with the analog MAP estimator finding that the performance loss in terms of increased MSE is relatively small, consistent with the comments in Remark 2.

For fixed N , results for different values of SNR $\gamma = \sigma_\theta^2 / \sigma_w^2$ are shown in Fig. 3. Note that different from the MAP based on analog observations, the MSE of the MAP based on binary observations increases with γ (it is constant for the former). Figs. 2 and 3 show that the simulation results follow the analytical results closely.

VII. CONCLUSIONS

We investigated distributed maximum a posteriori (MAP) estimation of random parameters using wireless sensor networks (WSN) with severely quantized observations. For different pragmatic signal scenarios considered, we established that the MAP estimators can be obtained as the maximum of concave functions, thus ensuring convergence of e.g., interior point methods. We also compared the MSE performance of our estimators based on severely quantized data with estimators based on analog-amplitude observations. While the MSE performance penalty increases with the observation signal

to noise ratio (SNR) we showed that, quite surprisingly, the MSE increase is rather small. Future research directions include consideration of estimation based on quantized – non-necessarily binary – data and application of MAP to state estimation of dynamic stochastic processes based on quantized data.

APPENDIX

A. Proof of Proposition 3

From (20), the i th diagonal element of FIM has two parts: $\mathbf{J}_{D \ ii} = \mathbf{J}_{D \ ii} + \mathbf{J}_{P \ ii}$ and each part is defined as :

$$\mathbf{J}_{D \ ii} = \mathbf{\theta, b} \left[\frac{\partial^2}{\partial \theta_i^2} p(\mathbf{b}_{0:N-1} | \boldsymbol{\theta}) \right] \quad (28)$$

$$\mathbf{J}_{P \ ii} = \mathbf{\theta, b} \left[\frac{\partial^2}{\partial \theta_i^2} p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \right]. \quad (29)$$

Note that the expectation in (28) and (29) is joint with respect to $\boldsymbol{\theta}$ and \mathbf{b} . It can be expressed as: $\mathbf{\theta, b} \cdot \mathbf{\theta} \{ \mathbf{b} | \boldsymbol{\theta} \}$. Assuming independence among the sensors $\{S_n\}_{n=0}^{N-1}$ and each observation $\{b_k(n)\}_{k=1}^K$ of sensor S_n , we can write the probability in (28) as:

$$p(\mathbf{b}_{0:N-1} | \boldsymbol{\theta}) = \sum_{n=0}^{N-1} \sum_{k=1}^K p(b_k(n) | \boldsymbol{\theta}) \mathcal{F}(\boldsymbol{\theta}), \quad (30)$$

with $p(b_k(n) | \boldsymbol{\theta}) = Q\left(\frac{b_k(n) \mathbf{h}_k^T [\boldsymbol{\mu}_{\boldsymbol{\theta}} - \boldsymbol{\theta}]}{\sigma_w}\right)$ since $w(n)$ is Gaussian. To facilitate the analysis, we define an observation independent quantity $q_k = Q\left(\frac{\mathbf{h}_k^T [\boldsymbol{\mu}_{\boldsymbol{\theta}} - \boldsymbol{\theta}]}{\sigma_w}\right)$. Since $b_k(n) \in \{\pm 1\}$ and $Q(-a) = 1 - Q(a)$, an equivalent form of the log-likelihood in (30) is given by

$$\mathcal{F}(\boldsymbol{\theta}) = \sum_{n=0}^{N-1} \sum_{k=1}^K \left[\frac{+b_k(n)}{q_k + 1 - q_k} \{ -q_k \} \right]. \quad (31)$$

Differentiating (31) twice with respect to (wrt) θ_i and evaluating the expectation $\mathbf{b} | \boldsymbol{\theta}$, we obtain

$$\mathbf{J}_{D \ ii} = \mathbf{\theta} \left[\sum_{n=0}^{N-1} \sum_{k=1}^K \frac{1}{q_k(1 - q_k)} \left(\frac{\partial q_k}{\partial \theta_i} \right)^2 \right], \quad (32)$$

where we have used the fact that $\mathbf{b} | \boldsymbol{\theta} \ b_k(n) = q_k - 1$. Now, differentiating q_k wrt to θ_i , we find

$$\frac{\partial q_k}{\partial \theta_i} = -\frac{1}{\sqrt{\pi} \sigma_w} e^{-[\mathbf{h}_k^T (\boldsymbol{\mu}_{\boldsymbol{\theta}} - \boldsymbol{\theta})]^2 / 2\sigma_w^2} \mathbf{h}_{ki}. \quad (33)$$

Using the chernoff bound and (33), eq. (32) becomes

$$\begin{aligned} \mathbf{J}_{D \ ii} &\geq \frac{N}{\pi \sigma_w^2} \mathbf{\theta} \left[\sum_{n=0}^{N-1} \sum_{k=1}^K \frac{h_{ki}^2 e^{-[\mathbf{h}_k^T (\boldsymbol{\mu}_{\boldsymbol{\theta}} - \boldsymbol{\theta})]^2 / 2\sigma_w^2}}{e^{-[\mathbf{h}_k^T (\boldsymbol{\mu}_{\boldsymbol{\theta}} - \boldsymbol{\theta})]^2 / 2\sigma_w^2}} \right] \\ &\geq \frac{N}{\pi \sigma_w^2} \sum_{k=1}^K h_{ki}^2 \mathbf{\theta} \left[e^{-[\mathbf{h}_k^T (\boldsymbol{\mu}_{\boldsymbol{\theta}} - \boldsymbol{\theta})]^2 / 2\sigma_w^2} \right] \\ &\geq \frac{N}{\pi \sigma_w^2} \sum_{k=1}^K h_{ki}^2 \int_{-\infty}^{\infty} e^{-[\mathbf{h}_k^T (\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}})]^2 / 2\sigma_w^2} p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) d\boldsymbol{\theta}. \end{aligned}$$

Since the prior distribution of $\boldsymbol{\theta}$ is Gaussian, we reach

$$\mathbf{J}_{D \ ii} \geq \frac{N(\mathbf{C}_{\boldsymbol{\theta}})^{-1/2}}{\pi \sigma_w^2 (\pi)^{K/2}} \sum_{k=1}^K h_{ki}^2 \int_{-\infty}^{\infty} e^{-\mathcal{G}_k(\boldsymbol{\theta})} d\boldsymbol{\theta}, \quad (34)$$

where the exponent in the integrand is given by

$$\begin{aligned} \mathcal{G}_k(\boldsymbol{\theta}) &= \frac{1}{\sigma_w^2} \mathbf{h}_k^T (\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}})^2 + -(\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}})^T \mathbf{C}_{\boldsymbol{\theta}}^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}}) \\ &\quad - \boldsymbol{\alpha}^T [\mathbf{g}_k \mathbf{g}_k^T + \mathbf{I}] \boldsymbol{\alpha} \end{aligned}$$

where we define $\boldsymbol{\alpha} = \mathbf{C}_{\boldsymbol{\theta}}^{-1/2} (\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}})$ and $\mathbf{g}_k = \sigma_w^{-1} \mathbf{C}_{\boldsymbol{\theta}}^{1/2} \mathbf{h}_k$. With this, the integral in (34) becomes

$$\int_{-\infty}^{\infty} e^{-\mathcal{G}_k(\boldsymbol{\theta})} d\boldsymbol{\theta} = \int_{-\infty}^{\infty} e^{-\frac{1}{2} \boldsymbol{\alpha}^T (\mathbf{g}_k \mathbf{g}_k^T + \mathbf{I}) \boldsymbol{\alpha}} | \mathbf{C}_{\boldsymbol{\theta}} |^{1/2} d\boldsymbol{\alpha} = (\pi)^{K/2} | \mathbf{C}_{\boldsymbol{\theta}} |^{1/2} (\mathbf{g}_k \mathbf{g}_k^T + \mathbf{I})^{-1/2}.$$

The determinant can be evaluated as a product of the eigenvalues of $\mathbf{g}_k \mathbf{g}_k^T + \mathbf{I}$. Since $\mathbf{g}_k \mathbf{g}_k^T / (\sigma_w^{-2} \mathbf{h}_k^T \mathbf{C}_{\boldsymbol{\theta}}^{1/2} \mathbf{h}_k)$ is an idempotent matrix with eigenvalues 1 or 0, the eigenvalues of $\mathbf{g}_k \mathbf{g}_k^T$ are $\sigma_w^{-2} \mathbf{h}_k^T \mathbf{C}_{\boldsymbol{\theta}}^{1/2} \mathbf{h}_k$ or 0. Hence, the eigenvalues of $\mathbf{g}_k \mathbf{g}_k^T + \mathbf{I}$ will be $1 + \sigma_w^{-2} \mathbf{h}_k^T \mathbf{C}_{\boldsymbol{\theta}}^{1/2} \mathbf{h}_k$ or 1. Thus, $(\mathbf{g}_k \mathbf{g}_k^T + \mathbf{I})^{-1/2} = \sigma_w^{-1} \mathbf{h}_k^T \mathbf{C}_{\boldsymbol{\theta}}^{-1/2} \mathbf{h}_k$. This finally leads to

$$\mathbf{J}_{D \ ii} \geq \frac{N}{\pi \sigma_w^2} \sum_{k=1}^K \frac{h_{ki}^2}{\sqrt{\sigma_w^2 + \mathbf{h}_k^T \mathbf{C}_{\boldsymbol{\theta}} \mathbf{h}_k}}. \quad (35)$$

Similarly, the FIM for the a priori information is given by

$$\mathbf{J}_P = \mathbf{C}_{\boldsymbol{\theta}}^{-1}. \quad (36)$$

From (35) and (36), we find the bound on the i th diagonal element of FIM as given by (23).

For the Clairvoyant estimator, the conditional probability in (28) is Gaussian such that

$$p(\mathbf{x}_{0:N-1} | \boldsymbol{\theta}) = \frac{1}{\sigma_w^2} \sum_{n=0}^{N-1} \sum_{k=1}^K [x_k(n) - \mathbf{h}_k^T \boldsymbol{\theta}]^2. \quad (37)$$

Differentiating (37) twice wrt to θ_i and computing the joint expectation wrt \mathbf{b} and $\boldsymbol{\theta}$ we get the result as in (24).

REFERENCES

- [1] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [2] S. M. Kay, *Fundamentals of Statistical Signal Processing - Estimation Theory*. Prentice Hall, 1993.
- [3] S. Kumar, F. Zhao, and D. Shepherd, "Collaborative signal and information processing in microsensor networks," *IEEE Signal Processing Magazine*, vol. 19, pp. 13–14, March 2002.
- [4] Z.-Q. Luo, "An isotropic universal decentralized estimation scheme for a bandwidth constrained ad hoc sensor network," *IEEE Journal on Selected Areas in Communications*, vol. 23, pp. 735–744, April 2005.
- [5] H. Papadopoulos, G. Wornell, and A. Oppenheim, "Sequential signal encoding from noisy measurements using quantizers with dynamic bias control," *IEEE Transactions on Information Theory*, vol. 47, pp. 978–1002, 2001.
- [6] A. Ribeiro and G. B. Giannakis, "Bandwidth-Constrained Distributed Estimation for Wireless Sensor Networks, Part II: Unknown pdf," *IEEE Transactions on Signal Processing*, 2006 (to appear).
- [7] S. Servetto, R. Knopp, A. Ephremides, S. Verdu, S. Wicker, and L. Cimini, "Guest Editorial: Fundamental performance limits of wireless sensor networks," *IEEE Journal on Selected Areas on Communications*, vol. 22, pp. 961–965, August 2004.
- [8] H. L. V. Trees, *Detection, Estimation, and Modulation Theory*. J. Wiley and Sons Inc., first ed., 1968.