DISTRIBUTED ESTIMATION USING BINARY DATA TRANSMITTED OVER FADING CHANNELS

Onur Ozdemir, Ruixin Niu, and Pramod K. Varshney

Department of Electrical Engineering and Computer Science
335 Link Hall, Syracuse University, Syracuse, NY 13244, USA

ABSTRACT

We study the parametric distributed estimation problem using a wireless sensor network (WSN) where each sensor observes an unknown scalar parameter, quantizes its observation and sends its quantized observation to a fusion center via fading and noisy communication channels. We propose to incorporate channel statistics rather than the instantaneous channel state information (CSI) into the maximum likelihood (ML) formulation and show that the resulting likelihood function is strictly log-concave almost surely with a change of variable provided that at least one of the communication channels between the sensors and the fusion center has nonzero capacity. We also investigate the effects of channel layer on the sensor threshold design and show that the threshold design problem is coupled with the channel layer and the sensor signal-to-noise ratio (SNR) only for nonsymmetric channels. Our formulation is very general in the sense that no assumptions are made about the physical layer in terms of the modulation schemes and the reception techniques.

Index Terms— Distributed estimation, fading channels, maximum likelihood estimation.

1. INTRODUCTION

The problem of parametric distributed estimation in the context of low-cost, low-power and bandlimited wireless sensor networks (WSNs) has recently been studied in the literature [1, 2, 3, 4, 5, 6, 7]. In [1, 2] the authors assume that the analog sensor measurements are transmitted from the sensors to the fusion center via fading and noisy wireless channels. Although their formulation allows to use linear estimators, the assumption of analog sensor measurements is highly demanding for a resource constrained WSN. In [3, 4], the authors use quantized data and assume that the communication channels between the sensors and the fusion center are ideal. The assumption of ideal communication channels may not always be the case for a WSN because of the associated stringent energy and bandwidth constraints. The authors in [5] analyze the distributed estimation problem by using binary data and by modeling the communication channels as additive white Gaussian noise (AWGN) channels. They do not consider the effects of channel fading due to multipath. The concept of fading channels with quantized sensor data has been studied in [6, 7], but they all assume the availability of instantaneous channel state information (CSI), which may be too costly to acquire for a resource constrained WSN. In addition, recent methods are all based on the assumption of coherent reception at the fusion center (more specifically BPSK), which requires a complex receiver at the fusion center and (possibly) training samples from the sensors to estimate the phase of the carrier of each sensor. In this paper, we investigate the distributed estimation problem using binary data transmitted over fading and noisy channels. We propose to incorporate channel statistics into the likelihood function without making any assumptions on either the modulation scheme employed at the sensors or the reception technique employed at the fusion center. More importantly, we show that by including the channel statistics into the likelihood function, the resulting maximum likelihood problem is strictly log-concave almost surely as long as at least one of the channels has nonzero capacity. This result is exciting since our approach only requires channel statistics rather than the instantaneous CSI of the wireless channels between sensors and the fusion center. Furthermore, we propose a sensor threshold design approach by adopting an objective function based on Fisher Information and show that, for symmetric sensor noise distributions, the sensor threshold design problem is coupled with the channel signal-to-noise ratio (SNR) and the sensor SNR only if the channel model is nonsymmetric. In other words, for a fixed sensor background noise level, sensor thresholds should be channel optimized if the channels are nonsymmetric.

2. PROBLEM FORMULATION

Let $\theta \in [-U, U]$ denote the unknown scalar parameter to be estimated. The received signal at each sensor can be written as follows:

$$x_i = \theta + n_i, \quad i = 1, \dots, N, \tag{1}$$

where n_i is the i.i.d. zero-mean sensor noise whose probability density function (pdf) is $p_n(n; \sigma)$ with σ denoting the known noise standard deviation and N is the total number of sensors. We assume that sensor noises are independent across sensors. At each sensor, the received signal x_i is quantized into two levels before being sent to the fusion center. The quantized observation model at sensor i can be written as

$$s_i = \begin{cases} 0, & x_i \le \tau \\ 1, & x_i > \tau, \end{cases} \tag{2}$$

where s_i is the binary observation of the i^{th} sensor and τ represents the predetermined threshold for the binary quantizer at each sensor. Let u_i denote the transmitted symbol (representing s_i) from the ith sensor using a binary modulation scheme. Note that u_i can take one of 2 possible values, i.e., $u_i \in \{u_i^{(s_i)}, s_i = 0, 1\}$. A generic system model is shown in Fig. 1. Each u_i is transmitted to the fusion center via fading and noisy communication channels. The channel gains as well as the channel noises are assumed to be independent across different channels and the fading for each channel is assumed to be slow enough that it can be assumed constant during the transmission of an observation symbol u_i (i=1,...,N). Note that h_i and v_i denote the fading channel gain and the channel noise, respectively, for the ith sensor in Fig. 1. After demodulation, the received observations at the fusion center can be denoted in vector form as $\mathbf{y} = [y_1 \dots y_N]^T$.

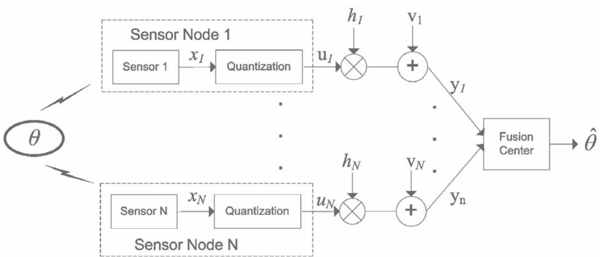

Fig. 1. Generic system model.

The fusion center needs to find an estimate, $\hat{\theta}$, for the unknown parameter based on the received observation vector $\mathbf{y}$.

3. MLE BASED ON CHANNEL STATISTICS

Note that if the instantaneous CSI for every channel is known at the fusion center, the estimation problem can be formulated as in [5], only with a difference of incorporating nonidentical channel SNRs for different channels. Here we consider a more general scenario where only the probability density function (pdf) of the channels, i.e., $p(h_i)$'s, are known. For this scenario, the likelihood of an observation for a given transmitted symbol at sensor i can be written as

$$p(y_i|u_i) = \int_0^\infty p\left(y_i|h_i, u_i\right) p(h_i) dh_i. \tag{3}$$

Note that the channel statistics is incorporated into the likelihood function in (3). The conditional pdf $p\left(y_i|h_i, u_i\right)$ is dependent on the physical layer of the network (e.g. the modulation scheme and the reception technique) and so is $p\left(y_i|u_i\right)$. Using (3), the likelihood function for sensor i is given as

$$\begin{aligned} p(y_i|\theta) &= \sum_{u_i} p(y_i|u_i) p(u_i|\theta) \qquad (4) \\ &= [1 - P_n(\tau - \theta)] p(y_i|u_i = u_i^{(0)}) + [P_n(\tau - \theta)] p(y_i|u_i = u_i^{(1)}), \end{aligned}$$

where $P_n(\cdot)$ is the complementary cumulative distribution function (cdf) of n_i. Then it is straightforward to write the log-likelihood function at the fusion center

$$\Lambda(\theta) = \sum_i^N \ln p(y_i|\theta), \tag{5}$$

where $p(y_i|\theta)$ is given in (4). From (5), the MLE of θ given $\mathbf{y}$ is the solution of the following maximization problem

$$\hat{\theta} = \arg\max_\theta \Lambda(\theta). \tag{6}$$

Proposition 1. *a) $\Lambda(\theta)$ is almost surely strictly concave in $P_n(\tau-\theta)$ for $P_n(\tau-\theta) \in (0,1)$ provided that at least one of the channels has nonzero capacity.*

b) The Fisher Information for the estimation problem in (3)-(6) is given by

$$I(\theta) = \qquad (7)$$
$$\sum_i \int_{y_i} \frac{p_n^2(\tau-\theta)\left[p(y_i|u_i = u_i^{(0)}) - p(y_i|u_i = u_i^{(1)})\right]^2}{p(y_i|\theta)} dyi$$

where $p(y_i|u_i)$ and $p(y_i|\theta)$ can be calculated using (3) and (4), respectively, and $p_n(u) \leftrightarrow p_n(u;\sigma)$.

Proof. a) Following a similar procedure as in [5], the second derivative of (5) with respect to $P_n(\tau-\theta)$ is given as

$$\begin{aligned} \frac{\partial^2 \Lambda(\theta)}{\partial P_n^2(\tau-\theta)} &= -\sum_i \frac{[p(y_i|u_i = u_i^{(0)}) - p(y_i|u_i = u_i^{(1)})]^2}{p^2(y_i|\theta)} \\ &\leq 0. \end{aligned} \tag{8}$$

Note that the equality in (8) is satisfied if all the channels have zero capacity, i.e., when $p(y_i|u_i = u_i^{(0)}) = p(y_i|u_i = u_i^{(1)})$ for all y_i and for all $i = 1, \ldots, N$, or if $p(y_i|u_i = u_i^{(0)}) = p(y_i|u_i = u_i^{(1)})$ for some y_i. However, the probability of the latter condition being satisfied is zero. Therefore, $\Lambda(\theta)$ is almost surely strictly concave in $P_n(\tau-\theta)$ if at least one of the channels has nonzero capacity.

b) Using the expression for the Fisher Information which is given by $I(\theta) = -E\left[\frac{\partial^2\Lambda(\theta)}{\partial\theta^2}\right]$, the proof can be carried out by taking the second derivative of $\Lambda(\theta)$ with respect to θ and calculating the expectation. The rigorous proof is omitted here for the sake of brevity. □

The result in Proposition 1.a guarantees the convergence of the ML problem to the global optimum in $P_n(\tau-\theta)$. This result is useful when the sensor noise distribution has the property such that the complementary cdf $P_n(\tau-\theta)$ is a strictly increasing function of θ (or equivalently one-to-one function of θ). Note that the Gaussian cdf is an example of this type. In this case, the invariance property can be used to find $\hat{\theta}$ which is given by

$$\hat{\theta} = \tau - P_n^{-1}(\hat{\alpha}), \; where \quad \alpha \triangleq P_n(\tau-\theta). \tag{9}$$

Let us define $F(\theta) \triangleq P_n(\tau-\theta)$. Assuming identical channel statistics and identical fixed threshold value at each sensor, the Cramér-Rao lower bound (CRLB) for this estimation problem reduces to

$$CRLB(\theta) = I(\theta)^{-1} = \frac{1}{N} I_s(\theta)^{-1}, \tag{10}$$

where $I_s(\theta)$ is defined as the per sensor Fisher Information given as

$$I_s(\theta) = \int_y \frac{p_n^2(\tau-\theta)\left[p(y|u = u^{(0)}) - p(y|u = u(1))\right]^2}{[1 - F(\theta)] p(y|u = u^{(0)}) + F(\theta) p(y|u = u^{(1)})} dy. \tag{11}$$

4. LOCAL SENSOR THRESHOLDS

In order to design optimal sensor thresholds, one needs to define an objective function in terms of which the optimality is sought. It is known that the optimal sensor threshold that minimizes the CRLB under perfect communication channel case is given as $\tau = \theta$ [3]. Since $\theta \in [-U, U]$ is unknown, a cost function which is independent of θ can be defined as the expected Fisher Information, $E_\theta[I(\theta)]$, based on which the optimal sensor thresholds can be designed. With the same assumptions as in (10), optimum sensor threshold is defined to be the solution of the following maximization problem:

$$\tau^{opt} = \arg\max_\tau E_\theta[I(\theta,\tau)] = \arg\max_\tau E_\theta[I_s(\theta,\tau)]. \tag{12}$$

Note from (12) that by using $E_\theta[I(\theta,\tau)]$ as the objective function, the problem is decoupled among sensors. In other words, a threshold that maximizes per sensor Fisher Information is the optimum threshold for all sensors. However, we need to point out that the optimal threshold design procedure given in (12) will only work for low to medium sensor SNR regimes. It is due to the fact that for high SNR regimes ($\sigma \ll U$) assigning an identical threshold value for each sensor will result in severe degradation in the performance for θ values which are approximately in the range $\theta \in [-U, \tau - 3\sigma] \cup [\tau + 3\sigma, U]$. In that case, it is natural to utilize different approaches such as using nonidentical thresholds at different sensors. Here we focus on the scenario where the sensor SNR is in the low or the medium range. Assuming that the sensor noise distribution is symmetric, the followings hold.

Conjecture 1. *For $\theta \in [-U, U]$, if the channel is symmetric, i.e., if $p(y|u^{(0)}) = p(-y + \mu_0 + \mu_1|u^{(1)})$ where $\mu_0 = E[y|u^{(0)}]$ and $\mu_1 = E[y|u^{(1)}]$, then $\tau^{opt} = \arg\max_\tau E_\theta[I_s(\theta,\tau)] = 0$.*

Let us now denote the channel SNR in dB as ρ.

Conjecture 2. *For $\theta \in [-U, U]$, if the channel is not symmetric, i.e., if $p(y|u^{(0)}) \neq p(-y + \mu_0 + \mu_1|u^{(1)})$, then*

a) $\tau^{opt} = \arg\max_\tau E_\theta[I_s(\theta,\tau)] \approx 0$ only when $\rho \to -\infty$.

b) for $\rho \nrightarrow -\infty$, τ^{opt} is a function of both the sensor SNR and the channel SNR, i.e, $\tau^{opt} = \arg\max_\tau E_\theta[I_s(\theta,\tau)] = f(\sigma, \rho)$.

Consider a nonsymmetric channel. Note that if $\rho \to -\infty$, then $p(y|u^{(0)}) \approx p(y|u^{(1)})$ for all y, which means that the resulting channel approaches zero capacity and the condition in Conjecture 1 is satisfied. Then the optimum threshold value that maximizes the expected Fisher Information approaches zero. We now turn our attention to medium to high channel SNR regimes. Consider a fixed sensor SNR, i.e., fixed σ. Since, for nonsymmetric channels, each different channel SNR results in different distributions of $p(y|u^{(0)})$ and $p(y|u^{(1)})$, it is clear from (11) that, (12) should result in different threshold values for different channel SNRs given a fixed sensor SNR. This introduces the concept of channel optimized sensor thresholds for nonsymmetric channels. Considering the other scenario in which the channel SNR is fixed, it is again clear from (11) that (12) should result in different threshold values for different sensor SNRs if the channel is not symmetric. The next section will provide numerical examples of the scenarios given in Conjectures 1 and 2[1].

5. NUMERICAL EXAMPLES

In this section, we will investigate two scenarios where each channel between the sensors and the fusion center is modeled as a unit power Rayleigh fading channel. The first scenario utilizes coherent reception at the fusion center employing BPSK as the modulation scheme. The second scenario utilizes a special case of noncoherent BFSK modulation scheme, which is ON/OFF keying (OOK). Note that the second (noncoherent) scenario does not require phase information at the fusion center which helps avoid training samples to estimate phase and/or complex receivers at the fusion center, and it allows for sensor censoring [8] which can increase network lifetime by saving sensor energy. For both scenarios, the received signal before demodulation can be denoted as $\tilde{y} = he^{j\phi}u + v$, where ϕ is the channel phase and v is a zero-mean complex Gaussian channel noise with independent real and imaginary parts having identical variance σ_v^2.

[1] We are currently working on the proofs of the Conjectures 1 and 2.

Consider the first (coherent) scenario where $u \in \{-1, 1\}$. The observation model after demodulation at the fusion center is given as

$$y = Re\{\tilde{y}e^{-j\phi}\} = hu + Re\{ve^{-j\phi}\}, \tag{13}$$

where $Re\{ve^{-j\phi}\} \sim \mathcal{N}(0, \sigma_v^2)$. Using the unit power Rayleigh fading pdf $p(h) = 2he^{-h^2}$ $(h \geq 0)$, $p(y|u)$ is given as [9]

$$p\,(y|u) = \frac{2\sigma_v}{\sqrt{2\pi}(1+2\sigma_v^2)} e^{\frac{-y^2}{2\sigma_v^2}} \times \left[1 + u\sqrt{2\pi}\alpha y e^{\frac{(\alpha y)^2}{2}} Q(-\alpha u y)\right], \tag{14}$$

where $Q(\cdot)$ is the complementary distribution function of the Gaussian distribution defined as $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$, and $\alpha = 1/(\sigma_v\sqrt{1+2\sigma_v^2})$. Note that (14) results in a symmetric channel model. Now consider the second (noncoherent) scenario where $u \in \{0, 1\}$. The observation model after demodulation (energy detection) at the fusion center is given as

$$y = |uhe^{j\phi} + v|^2 \tag{15}$$

In this case, $p(y|u)$ results in

$$\begin{aligned} p\,(y|u=0) &= \frac{1}{2\sigma_v^2} e^{-\frac{y}{2\sigma_v^2}} \\ p\,(y|u=1) &= \frac{1}{1+2\sigma_v^2} e^{-\frac{y}{1+2\sigma_v^2}}. \end{aligned} \tag{16}$$

Unlike the first scenario, the second scenario where OOK is employed results in a nonsymmetric channel model (16). Note that the resulting observation models (14), (16) can now be substituted in the log-likelihood function $\Lambda(\theta)$ to perform the ML estimation in (6). We should note that the closed forms for the CRLBs (10), (11) can not be obtained for these scenarios. However, the calculations of the CRLBs can be carried out with high accuracy using a numerical method such as composite Simpson's rule [10].

Fig. 2 shows the performance of the MLE estimators for both scenarios with respect to different number of sensors. Mean square error (MSE) of the estimator is the performance criterion and each MSE value is computed based on 10000 Monte Carlo trials. Simulation parameters are as follows. Sensor noise is assumed to be zero mean Gaussian with standard deviation $\sigma = 1$. Channel SNR is 10dB for both scenarios, $\theta = 1$ (sensor SNR = 0 dB) and $\tau = 1.2$. It is clear from the figure that the ML estimators achieve their performance bounds (CRLBs) for relatively small number of sensors and low sensor SNRs, and they do not require the knowledge of instantaneous CSIs. The result in Proposition 1 is exploited to find the global solution to the MLE problem in $P_n(\tau - \theta)$ by using Newton's algorithm and then the invariance property is invoked to find $\hat{\theta}$. The performance of the clairvoyant ML estimator [3] which uses binary data and assumes error-free communications between the sensors and the fusion center is also shown in Fig. 2. Notice that the performances of the ML estimators developed in Section 3 are very close to the performance of the clairvoyant estimator and they only require the information about channel statistics. Especially for the coherent case, the performance is almost the same as the clairvoyant estimator.

Now we evaluate the effects of sensor thresholds. Fig. 3 shows the expected per sensor Fisher Information with respect to different sensor threshold values for different channel conditions. It is clear from the figure that for symmetric channels, the optimum sensor

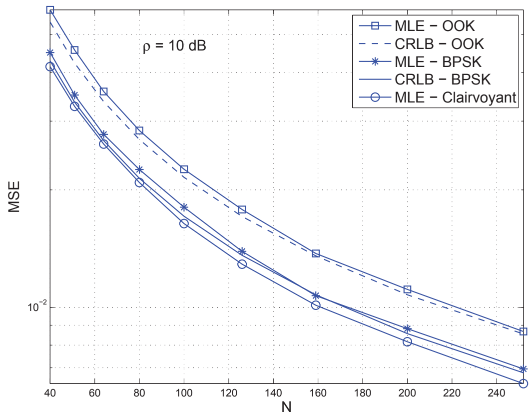

Fig. 2. Performance comparison of ML estimators for different communication scenarios

threshold is always zero. However, for nonsymmetric channels, optimum sensor threshold is no longer zero, and it depends on both the the channel SNR and the sensor SNR. One can argue that the optimum sensor threshold obtained through the maximization of the expected Fisher Information does not necessarily minimize the CRLB due to the nonlinear relationship between them. However, our experimental results in Fig. 4 shows that such T^{opt} leads to the smallest MSE for this particular problem. For each MSE value in Fig. 4, the unknown parameter θ is randomly drawn from $[-1, 1]$ using 10000 Monte Carlo trials and N is set to 100. It is clear from the figure that numerical results support our argument.

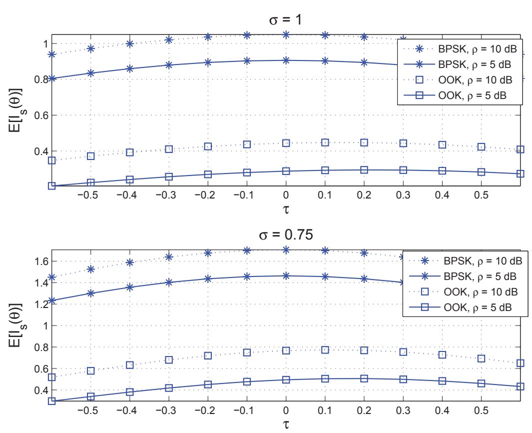

Fig. 3. Expected per sensor Fisher Information with respect to different sensor thresholds

6. CONCLUSIONS

This paper has studied the distributed parameter estimation problem where the channels are modeled as fading and noisy, and only the channel statistics are available. We have shown that the resulting ML formulation which includes channel statistics is concave provided that at least one of the channels has nonzero capacity and therefore can be easily utilized to find an estimate which is unbiased and efficient for asymptotic regimes. We have also shown that, for symmetric sensor noise distributions, the sensor threshold design problem is coupled with the channel layer and the sensor noise level for nonsymmetric channels. Our future work includes the extension of this analysis to the case of multi-bit data. Furthermore, analytical expressions for the optimal threshold values need to be derived for nonsymmetric channels.

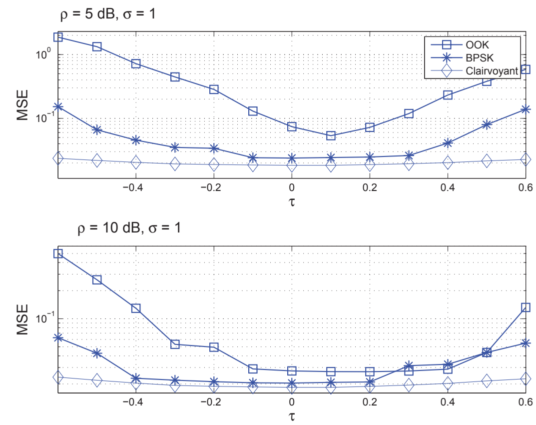

Fig. 4. Performance of ML estimators with respect to different sensor thresholds

7. REFERENCES

[1] C. Tepedelenlioglu, M. K. Banavar, and A. Spanias, "Asymptotic analysis of distributed estimation over fading multiple access channels," in *Proc. 41st Asilomar Conf. Signals, Systems and Computers*, Oct. 2007, pp. 2140–2144.

[2] T. Wimalajeewa and S. K. Jayaweera, "Power efficient distributed estimation in a bandlimited wireless sensor network," in *Proc. 41st Asilomar Conf. Signals, Systems and Computers*, Oct. 2007, pp. 2156–2160.

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

[4] ——, "Bandwidth-constrained distributed estimation for wireless sensor networks-part II: unknown probability density function," *IEEE Trans. Signal Process.*, vol. 54, no. 7, pp. 2784–2796, Jul. 2006.

[5] T. C. Aysal and K. E. Barner, "Constrained decentralized estimation over noisy channels for sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1398–1410, Apr. 2008.

[6] X. Wang and C. Yang, "Decentralized estimation methods for wireless sensor networks with known channel state information," in *Proc. IEEE 64th Vehicular Technology Conference*, Sep. 2006, pp. 1–5.

[7] ——, "Optimal and suboptimal decentralized estimation underlying fading channels in wsns," in *Proc. IEEE Global Telecommunications Conference*, Nov. 2007, pp. 2863–2867.

[8] R. Jiang and B. Chen, "Fusion of censored decisions in wireless sensor networks," *IEEE Trans. Wireless Comm.*, vol. 4, no. 6, pp. 2668–2673, Nov. 2005.

[9] R. Niu, B. Chen, and P. K. Varshney, "Fusion of decisions transmitted over rayleigh fading channels in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1018–1027, Mar. 2006.

[10] J. H. Mathews and K. K. Fink, *Numerical Methods Using Matlab*. Prentice Hall, 2004.