UNIVERSAL DECENTRALIZED ESTIMATION IN A BANDWIDTH CONSTRAINED SENSOR NETWORK

Zhi-Quan Luo and Jin-Jun Xiao

Dept. of Electrical and Computer Engineering, University of Minnesota
200 Union Street SE., Minneapolis, MN 55455, USA
Emails: {luozq,xiao}@ece.umn.edu

ABSTRACT

We consider universal decentralized estimation of a noise-corrupted signal by a bandwidth constrained sensor network with a fusion center (FC). We show that in a homogeneous sensing environment and under a bandwidth constraint of 1-bit per sample per node, there exist universal decentralized estimation schemes (DES) with a mean squared error (MSE) decreasing at the rate $1/K$, where K is the total number of sensors. We extend such 1-bit decentralized estimators to the case of inhomogeneous sensing environment, and propose quantization and transmission power control strategies for local sensors in order to minimize the total consumed sensor energy while ensuring a given MSE performance. We also design a DES for the joint estimation of a vector source based on its noisy and linearly distorted observations, and show that to achieve a MSE within a factor of 2 away from the best linear unbiased estimator (BLUE), the local message length has a nice form of being the channel capacity of "a virtual AWGN channel" from "nature" to each local sensor.

1. INTRODUCTION

A wireless sensor network (WSN) consists of a large number of geographically distributed nodes, each with finite battery power and limited capability in computation, communication and mobility. When properly programmed and networked, sensor nodes in a WSN can cooperate to yield significant signal processing capability with unprecedented robustness and versatility, thus making WSN an attractive low-cost technology for situation awareness applications such as environmental monitoring, smart factory instrumentation, military surveillance, precision agriculture, intelligent transportation and space exploration, to name a few.

Apart from power and bandwidth limitations, distributed signal processing with a WSN differs from the traditional signal processing framework in several important aspects. First, observation data in a WSN is located at different nodes across the network. Secondly, the parametric data model used and the knowledge of sensor noise distributions are not easy to characterize, especially for applications in a dynamic sensing environment. Third, sensor nodes in a WSN can be either static or mobile, and the inter-sensor communication can be either peer to peer (as in ad hoc WSN), or restricted to unidirectional with a common destination called the fusion center (FC). Lastly, WSN size and topology may change dynamically.

This research is supported in part by the Natural Sciences and Engineering Research Council of Canada, Grant No. OPG0090391, by the Canada Research Chair Program and by the National Science Foundation, Grant No. DMS-0312416.

The problem of decentralized estimation has been studied first in the context of distributed control [1] and tracking [9], later in data fusion [2, 7], and most recently in wireless sensor networks [8]. Among these studies, the prevailing assumption has been that the joint distribution of the sensor observations is known, with some also making the additional assumption that the communication links can transmit real values and are distortionless. Our work differs from these studies in that it requires neither the knowledge of noise distributions nor the use of a training sequence except either sensor observation range, or first/second noise moments. In other words, the DESs derived in this paper is universal. Moreover, these schemes have small bandwidth requirement of either 1 bit or a small number of bits per sensor per node. A main objective of this paper is to deal with the signal processing aspect of sensor network research in which the main design objectives are performance, bandwidth/power efficiency, scalability and robustness to changes in the network or environment.

2. DES IN HOMOGENEOUS ENVIRONMENT

To understand how the signal processing capability of a WSN scales with its size, we consider a generic decentralized parameter estimation problem under bandwidth constraints [4]. Specifically, suppose a set of K distributed sensors and a fusion center (FC) wish to cooperate to estimate an unknown parameter θ. Let the sensor observations be described by

$$x_k = \theta + n_k, \quad k = 1, 2, ..., K, \tag{1}$$

where the sensor noise variables $\{n_k : k = 1, 2, ..., K\}$ are assumed to be additive, zero mean, spatially uncorrelated, but otherwise unknown. For simplicity, we further assume that n_k and θ are scalar, real and bounded to a known interval $[-U, U]$. It follows that the observations $x_k \in [-2U, 2U]$, for all k.

If the FC is given all the data samples $\{x_k : k = 1, 2, ..., K\}$, then it can estimate θ using the standard sample mean estimator (also known as BLUE):

$$\bar{\theta}_K = (x_1 + x_2 + \cdots + x_K)/K.$$

This estimator is universal since it is independent of the noise pdf. Its MSE is $\mathrm{E}(|\bar{\theta}_K - \theta|^2) = \sigma^2/K$, where σ^2 is the noise variance. This MSE coincides with the Cramer-Rao lower bound (CRLB) when n_k is Gaussian. In other words, if infinite bandwidth is available so that the sensors can communicate their real-valued observations to the FC error-free, then the WSN has *a signal processing capability that scales linearly with network size K*. Surprisingly, the same remains true even if we constrain each sensor message to be one binary bit.

2.1. A 1-bit Universal DES

Specifically, suppose each sensor compresses its observation x_k to *one binary bit* $m_k(x_k)$ and sends the resulting 0-1 valued message to the FC. (The ensuing results can be easily generalized to the case of any constant number of bits.) Upon receiving these binary messages, the FC combines them to produce a final estimate of $\bar{\theta}_K$:

$$\bar{\theta}_K = f(m_1(x_1), m_2(x_2), ..., m_K(x_K)), \quad (2)$$

where f is a real-valued fusion function. We will refer to $\{f, m_k : k = 1, 2, ..., K\}$ as a *decentralized estimation scheme* or *DES* for short. The problem of decentralized estimation is then to design the local message functions $\{m_k : k = 1, 2, ..., K\}$ and the fusion function f so that $\bar{\theta}_K$ is closest to θ in the MSE sense.

Theorem 2.1 *[4] There exists a universal DES that estimates θ with an MSE not exceeding $4U^2/K$, even if each sensor sends only one-bit information to the FC. The DES that achieves this bound lets $1/2$ of the sensors quantize their observations to the first most significant bit (MSB), $1/4$ of the sensors quantize their observations to the second MSB, and so on.*

Since n_k is bounded in $[-U, U]$, it follows that $\text{Var}(n_k) \leq U^2$, $\forall\ k$. Hence the (worst case) CRLB for any universal unbiased estimator is U^2/K even in the absence of the binary message constraint. Thus, the worst case performance of the universal DES described in Theorem 2.1 is *within a constant factor of* 4 *to being optimal*. Notice also that this DES assigns more sensors to estimate the first MSB of θ than any other bit. This is intuitively satisfying since getting the first MSB of θ right has the highest impact on minimizing the final MSE.

2.2. Extension to an ad hoc WSN

One limitation of the aforementioned universal DES is that it requires the use of a FC and the knowledge of network size in order to specify which sensor should quantize its observation to which bit. Moreover, this DES is not isotropic since sensors quantize their observations to possibly different MSB. This makes it difficult to implement this DES in an ad hoc sensor network where there is little or no coordination among sensors. To overcome this difficulty, we have developed the following *probabilistic DES* [5]:

- With each new sample x_k, sensor k flips a coin and, with probability $1/2$, quantizes x_k to the first MSB, with probability $1/4$ quantizes x_k to the second MSB, and so on. The quantization outcome (one bit) is sent to all its neighbors.
- Sensor messages are forwarded in the network via an underlying WSN protocol. Each sensor recursively computes the average of all received binary messages which are distinct (determined by, say, the sender's ID), and uses it as an estimator of θ.

Intuitively, with the aforementioned coin flipping at each sensor node, there will be roughly $1/2$ of the sensors in the network quantizing their observations to the first MSB, about $1/4$ of the sensors in the network quantizing their observations to the second MSB, and so on. Thus, this probabilistic DES should closely approximate the MSE performance of the DES in Theorem 2.1. This is confirmed in the next result.

Theorem 2.2 *[5] Suppose the aforementioned probabilistic DES is implemented in a connected ad hoc WSN. Assume that each message has a header containing the sender's ID and eventually arrives at its destination without error. Then each node in the WSN produces an unbiased estimate of θ with an MSE of at most $4U^2/(n+1)$, where n denotes the number of distinct messages received by this sensor. Compared to the CRLB, this is also within a constant factor of* 4 *to being optimal.*

The probabilistic DES of Theorem 2.2 is *robust*, *isotropic*, and *universal* in that all sensors operate identically and independently of possible changes in network topology or sensor noise pdf.

3. DES IN INHOMOGENEOUS ENVIRONMENTS

Theorems 2.1 and 2.2 both assume that the sensor noise samples are identically distributed. This assumption may not hold when sensors have variable quality and the sensing environment is inhomogeneous. For example, the sensor closer to the target may have a higher local SNR than those farther away. Let the noise variance of the k-th sensor be σ_k^2. The classical centralized BLUE estimator combines the (real-valued) sensor observations linearly to achieve an MSE of $\left(\sum_{k=1}^K 1/\sigma_k^2\right)^{-1}$.

3.1. A Totally Distributed Estimation Scheme

The classical BLUE requires the knowledge of $\{x_k, \sigma_k^2\}$ from each sensor. Instead, we propose the following universal DES in *inhomogeneous* sensing environment [6]:

- At sensor k, choose

$$L_k = \left\lceil \log \frac{2U}{\sigma_k} \right\rceil, \quad (3)$$

  and take m_k to be the first L_k bits of the binary expansion of $\frac{2U + x_k}{4U} = \sum_{i=1}^{\infty} b_i 2^{-i}$, where $b_i = \{0, 1\}$, plus an extra random bit to make m_k unbiased.
- The final estimator at the FC is

$$\overline{\theta}_K(m_k) = \left(\sum_{k=1}^K 2^{2L_k}\right)^{-1} \sum_{k=1}^K 2^{2L_k} 2U(2m_k - 1). \quad (4)$$

Theorem 3.1 *[6] $\overline{\theta}_K$ in (4) is an unbiased estimator of θ, i.e. $E(\overline{\theta}_K) = \theta$, and $\overline{\theta}_K$ has an MSE*

$$E\left(\left|\overline{\theta}_K - \theta\right|^2\right) < \frac{25}{8}\left(\sum_{k=1}^K \frac{1}{\sigma_k^2}\right)^{-1},$$

which is optimal (up to a factor of 3.125) when compared to the centralized BLUE estimator.

In the above DES, each sensor only needs to know its own noise variance. The final fusion (4) is completely determined by the received messages $\{m_k : k = 1, 2, ..., K\}$. As expected, higher quality sensors (i.e., with lower noise) send more bits and their messages carry more weight at the fusion process.

3.2. Optimal Power Scheduling

In previous sections, we have derived efficient DESs under the assumption that the communication links between the sensors and the FC are distortionless. Such an assumption can be unrealistic in practical situations, especially when the power of the sensors are limited. In this section, we briefly describe how to extend the universal DESs to the case where the communication links are corrupted by Additive White Gaussian Noise (AWGN). Our ultimate goal is to minimize total power consumption per round of estimation, while ensuring a prescribed MSE performance D.

We assume that the channel between sensor k and the FC is experiencing a path-loss $a_k = d_k^\alpha$, where d_k is the transmission distance and α is the pass-loss constant. Sensor k quantizes its observation to L_k bits, where L_k are to be optimized depending on target MSE, sensor noise levels and channel gains from sensors to the FC. We consider a practical model assuming that sensors follow a Time Division Multiple Access (TDMA) scheme, and to minimize the delay of estimation, sensor k adopts adaptive MQAM of constellation size L_k so that all quantized bits from one sample can be transmitted by one single channel use. Suppose P_k is the transmitting power of sensor k (see [3] for details). To minimize the L^2-norm of $P = (P_1, P_2, \ldots, P_K)$, we obtain then the optimal value of L_k [10]:

$$L_k^{opt} = \log\left(1 + \frac{2U}{\sigma_k}\sqrt{\left(\frac{\eta_0}{a_k} - 1\right)^+}\right), \qquad (5)$$

where η_0 is a universal constant decided jointly by target MSE, sensor noise levels and channel gains. The optimal transmission power for sensor k is given as $P_k \sim \frac{a_k}{\sigma_k}\sqrt{\left(\frac{\eta_0}{a_k} - 1\right)^+}$.

When $\frac{\eta_0}{a_k} \leq 1$, or $a_k \geq \eta_0$, we have $L_k = 0$, and therefore $P_k = 0$. Simulation shows that in some cases, a large number of sensors with bad channel qualities or poor observations shut off. Numerical examples also show that an energy savings of up to 70% can be saved when compared to uniform quantization strategy in which each sensor generates the same number of bits (see [10]). The message length formula in (5) is intuitively appealing as it indicates that the message length should be proportional to the logarithm of local SNR scaled by channel path gain. This is in the same spirit as the message length formula in (3) when the channel is ideal such that a real number can be transmitted without distortion.

4. EXTENSION TO VECTOR SOURCE ESTIMATION

Previous DESs have assumed that sensor observations are in a finite range $[-2U, 2U]$. In this section, we first design a DES estimating a scalar source when sensor observations are unbounded, then extend this strategy to estimate vector sources from their noisy and linearly distorted observations.

4.1. Scalar DES Revisited

We assume that FC has the knowledge of sensor noise variances $\{\sigma_k^2 : k = 1, 2, \ldots, K\}$. This assumption is realistic when the source and whole network are stable, so that the quality of observations at one single sensor does not change fast with time. Once the sensor noise variances are acquired, they can be used for a large number of estimation rounds.

The message function m_k is then taken as the (randomly rounded) integer part of $x'_k = x_k/(2\sigma_k)$. The FC weights these message functions optimally by the knowledge of noise variances. The whole scheme is outlined as follows.

- Suppose $x'_k = \text{sign}(x'_k)(i_k + r_k)$, where i_k and r_k are the integer and decimal parts of $|x'_k|$ respectively. Then $0 \leq r_k < 1$. Construct the message function as

$$m_k(x_k, \sigma_k) = \text{sign}(x'_k)(i_k + d_k), \qquad (6)$$

  where d_k is a binary random variable with $P(d_k = 1) = r_k$ and $P(d_k = 0) = 1 - r_k$.
- The final estimator of θ at the FC is

$$\overline{\theta}_K(m_k) = \left(\sum_{k=1}^K \frac{1}{\sigma_k^2}\right)^{-1} \sum_{k=1}^K \frac{2m_k}{\sigma_k}. \qquad (7)$$

Theorem 4.1 *For the DES in (6)–(7), $\overline{\theta}_K$ is an unbiased estimator of θ, and the MSE*

$$E\left(\overline{\theta}_K - \theta\right)^2 \leq 2\left(\sum_{k=1}^K \frac{1}{\sigma_k^2}\right)^{-1}. \qquad (8)$$

Moreover, for all $1 \leq k \leq K$, the average message length of m_k is not more than $\frac{1}{2}\log(1+\gamma_k)$ binary bits, where $\gamma_k = P_\theta/\sigma_k^2$ with $P_\theta = E(\theta^2)$.

Proof. It is easy to see that $\text{E}(m_k) = \text{E}(x'_k) = \theta/(2\sigma_k)$ which reveals that $\overline{\theta}_K$ is an unbiased estimator of θ from (7). Also $\text{E}(m_k - x'_k)^2 = \text{E}(d_k - r_k)^2 = \text{Var}(d_k) \leq 1/4$ since d_k is a random variable taking values from $\{0, 1\}$. So

$$\text{E}(2\sigma_k m_k - \theta)^2 = \text{E}(2\sigma_k m_k - x_k)^2 + \text{E}(x_k - \theta)^2 \leq 2\sigma_k^2.$$

Thus (8) follows easily from (7). Next we give an upper bound of the average length of the integer message function m_k. We know that $m_k = \lfloor x'_k \rfloor$ or $\lceil x'_k \rceil$, therefore $L_k \approx \lceil \log|m_k| \rceil \leq 1 + \log|m_k| \approx 1 + \log|x'_k| = 1 + \log\frac{|x_k|}{2\sigma_k} = \log\frac{|x_k|}{\sigma_k}$. The average length of m_k

$$\begin{aligned}\text{E}(L_k) &\approx \text{E}\left(\log\frac{|x_k|}{\sigma_k}\right) \leq \log\left(\frac{\text{E}(|x_k|)}{\sigma_k}\right) \\ &= \log\left(\frac{\sqrt{\text{E}(\theta^2) + \sigma_k^2}}{\sigma_k}\right) = \frac{1}{2}\log(1+\gamma_k),\end{aligned}$$

where we applied the fact that $\log x$ is a concave function, and $(\text{E}(|x_k|))^2 \leq \text{E}(x_k^2) = \text{E}(\theta^2) + \sigma_k^2$. □

If we model the sensing process of a parameter from its source to sensor observation x_k as a discrete time AWGN channel, the channel capacity is $C_k = \frac{1}{2}\log(1+\gamma_k)$. The average message length of m_k upper bounded by C_k reveals that the optimal bit assignment is decided by the number of "useful" bits contained in x_k, which is the average information flow from the source measured by channel capacity.

4.2. Joint Estimation of a Vector Source

Suppose the parameter to be estimated is a vector $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_s]^T$. Sensor observations $\mathbf{x}_k$ are linear version of $\boldsymbol{\theta}$ corrupted by additive noises and are described by

$$\mathbf{x}_k = \mathbf{H}_k\boldsymbol{\theta} + \mathbf{n}_k, \quad k = 1, 2, ..., K,$$

where $\mathbf{H}_k$ is a matrix with dimension (r_k, s). We assume that noise $\mathbf{n}_k$ has zero mean and covariance matrix $\mathbf{C}_k$, but otherwise unknown. Noises are spatially uncorrelated among sensors. For simplicity, we assume $\mathbf{P}_{\boldsymbol{\theta}} = \mathrm{E}(\boldsymbol{\theta}\boldsymbol{\theta}^T) = \mathbf{I}_s$. Otherwise we can replace $\mathbf{H}_k$ by $\mathbf{H}'_k = \mathbf{H}_k\mathbf{P}_{\boldsymbol{\theta}}^{1/2}$, and let $\boldsymbol{\theta}' = \mathbf{P}_{\boldsymbol{\theta}}^{-1/2}\boldsymbol{\theta}$ which implies $\mathbf{P}_{\boldsymbol{\theta}'} = \mathbf{I}_s$. Again, we suppose FC knows $\{(\mathbf{H}_k, \mathbf{C}_k) : k = 1, 2, \ldots, K\}$.

If sensors can perfectly send the observations $\{\mathbf{x}_k : k = 1, 2..., K\}$ to the FC, the FC can perform the BLUE estimator minimizing the MSE as follow

$$\overline{\boldsymbol{\theta}}_K(\mathbf{x}) = \left(\sum_{k=1}^{K} \mathbf{H}_k^T \mathbf{C}_k^{-1} \mathbf{H}_K\right)^{-1} \sum_{k=1}^{K} \mathbf{H}_k^T \mathbf{C}_k^{-1} \mathbf{x}_k. \qquad (9)$$

Let $\boldsymbol{\epsilon}(\mathbf{x}) = \overline{\boldsymbol{\theta}}_K(\mathbf{x}) - \boldsymbol{\theta}$. A simple calculation shows that this estimator has an MSE covariance matrix

$$\mathrm{E}\left(\boldsymbol{\epsilon}(\mathbf{x})\boldsymbol{\epsilon}(\mathbf{x})^T\right) = \left(\sum_{k=1}^{K} \mathbf{H}_k^T \mathbf{C}_k^{-1} \mathbf{H}_K\right)^{-1}.$$

Theorem 4.2 *To perform the estimator (9) at the FC, from each sensor k, the number of real messages required to be transmitted is $d_k = rank(\mathbf{H}_k^T\mathbf{C}_k^{-1}\mathbf{H}_k) \le r_k$, where $r_k = \dim(\mathbf{H}_k^T\mathbf{C}_k^{-1}\mathbf{H}_k)$.*

Proof. Consider eigen-decomposition $(\mathbf{H}_k^T\mathbf{C}_k^{-1}\mathbf{H}_K)^{\frac{1}{2}} = \mathbf{U}_k\boldsymbol{\Lambda}_k\mathbf{U}_k^T$ where $\boldsymbol{\Lambda}_k = \mathrm{diag}(\lambda_{k,i} : i = 1, 2, \ldots, r_k)$. Define

$$\mathbf{x}'_k = \mathbf{U}_k\boldsymbol{\Lambda}_k^{-1}\mathbf{H}_k^T\mathbf{C}_k^{-1}\mathbf{x}_k = \mathbf{s}_k + \mathbf{w}_k$$

where $\mathbf{s}_k = \boldsymbol{\Lambda}_k\mathbf{U}_k^T\boldsymbol{\theta}$ and $\mathbf{w}_k = \mathbf{U}_k\boldsymbol{\Lambda}_k^{-1}\mathbf{H}_k^T\mathbf{C}_k^{-1}\mathbf{n}_k$. Suppose $\bar{\mathbf{x}}'_k$ is the vector containing the first d_k components of $\mathbf{x}'_k$, and the remaining $r_k - d_k$ components are filled up by 0, then the estimator

$$\overline{\boldsymbol{\theta}}_K(\mathbf{x}) = \left(\sum_{k=1}^{K} \mathbf{H}_k^T \mathbf{C}_k^{-1} \mathbf{H}_K\right)^{-1} \sum_{k=1}^{K} \mathbf{U}_k\boldsymbol{\Lambda}_k\bar{\mathbf{x}}'_k. \qquad (10)$$

is equivalent to (9). □

We can see that $\mathrm{E}(\mathbf{s}_k\mathbf{s}_k^T) = \boldsymbol{\Lambda}_k\mathbf{U}_k^T\mathrm{E}(\boldsymbol{\theta}\boldsymbol{\theta}^T)\mathbf{U}_k\boldsymbol{\Lambda}_k = \boldsymbol{\Lambda}_k^2$ and $\mathrm{E}(\mathbf{w}_k\mathbf{w}_k^T) = \mathbf{U}_k\boldsymbol{\Lambda}_k^{-1}\mathbf{H}_k^T\mathbf{C}_k^{-1}\mathrm{E}(\mathbf{n}_k\mathbf{n}_k^T)\mathbf{C}_k^{-1}\mathbf{H}_k\boldsymbol{\Lambda}_k^{-1}\mathbf{U}_k^T = \mathbf{I}_{r_k}$. Therefore, if we let $\gamma_{k,i}, 1 \le i \le r_k$ be the SNR of the i-th component of $\mathbf{x}'_k$, then $\gamma_{k,i} = \lambda_{k,i}^2$. In addition, all components of $\mathbf{x}'_k$ are mutually independent, and $\gamma_{k,i} = 0$ for all $d_k + 1 \le i \le r_k$.

From Theorem 4.2, we know that FC is only interested in $\bar{\mathbf{x}}'_k$. But $\bar{\mathbf{x}}'_k$ can not be transmitted as real values, so we quantize the first d_k non-zero components of $\bar{\mathbf{x}}'_k$ independently using the strategy in Section 4.1 to get $\bar{\mathbf{m}}_k$, and $\mathbf{m}_k$ is $\bar{\mathbf{m}}_k$ adding $r_k - d_k$ components of 0. Then $\mathrm{E}\left((\mathbf{m}_{k,i} - \bar{\mathbf{x}}'_{k,i})^2\right) \le 1$ for all $1 \le i \le r_k$. Due to the independent property of quantizations, we obtain

$$\mathrm{E}((\bar{\mathbf{x}}'_k - \mathbf{m}_k)(\bar{\mathbf{x}}'_k - \mathbf{m}_k)^T) \preceq \mathbf{I}_{r_k}. \qquad (11)$$

The fusion function is the same as (10) except that $\bar{\mathbf{x}}'_k$ is replaced by $\mathbf{m}_k$, i.e.

$$\overline{\boldsymbol{\theta}}_K(\mathbf{m}) = \left(\sum_{k=1}^{K} \mathbf{H}_k^T \mathbf{C}_k^{-1} \mathbf{H}_K\right)^{-1} \sum_{k=1}^{K} \mathbf{U}_k\boldsymbol{\Lambda}_k\mathbf{m}_k. \qquad (12)$$

Theorem 4.3 *Let $\boldsymbol{\epsilon}(\mathbf{m}) = \overline{\theta}_K(\mathbf{m}) - \boldsymbol{\theta}$. For the estimator (12), its MSE covariance matrix*

$$E\left(\boldsymbol{\epsilon}(\mathbf{m})\boldsymbol{\epsilon}(\mathbf{m})^T\right) \prec 2\left(\sum_{k=1}^{K} \mathbf{H}_k^T \mathbf{C}_k^{-1} \mathbf{H}_K\right)^{-1}. \qquad (13)$$

Moreover, for all $1 \le k \le K$, the average message length of $\mathbf{m}_k$ is approximately $\frac{1}{2}\log\det(\mathbf{I} + \mathbf{H}_k^T\mathbf{C}_k^{-1}\mathbf{H}_k)$ binary bits.

Proof. The proof of (13) follows from (11) and the fact that $\mathrm{E}(\mathbf{m}_k) = \mathbf{x}'_k$. Moreover, let $L_{k,i}$ be the length of message $m_{k,i}$, and L_k be the total number of bits in $\mathbf{m}_k$. Using the message length bound for scalar case in Theorem 4.1, we have

$$\begin{aligned}\mathrm{E}(L_k) &= \sum_{i=1}^{d_k} L_{k,i} \approx \sum_{i=1}^{d_k}\left(\frac{1}{2}\log\left(1 + \gamma_{k,i}\right)\right) \\ &= \frac{1}{2}\log\prod_{i=1}^{d_k}(1 + \gamma_{k,i}) = \frac{1}{2}\log\det(\mathbf{I} + \mathbf{H}_k^T\mathbf{C}_k^{-1}\mathbf{H}_k),\end{aligned}$$

where we have used the fact that $\gamma_{k,i} = \lambda_{k,i}^2$, and $\lambda_{k,i}^2$ are the eigenvalues of $\mathbf{H}_k^T\mathbf{C}_k^{-1}\mathbf{H}_k$. □

The quantity $C_k = \frac{1}{2}\log\det(\mathbf{I} + \mathbf{H}_k^T\mathbf{C}_k^{-1}\mathbf{H}_k)$ is the Shannon capacity of a "virtual AWGN channel" from nature to sensor k with channel given by $\mathbf{H}_k$, noise covariance matrix $\mathbf{C}_k$ and input power $\mathbf{P}_{\boldsymbol{\theta}} = \mathbf{I}_s$. The fact that C_k is channel capacity nicely indicates that the message length is decided by the number of "useful" bits contained in $\mathbf{x}_k = \mathbf{H}_k\boldsymbol{\theta} + \mathbf{n}_k$.

5. REFERENCES

[1] D.A. Castanon and D. Teneketzis, "Distributed Estimation Algorithms for Nonlinear Systems," *IEEE Transactions on Automatic Control*, Vol. AC-30, pp. 418–425, 1985.

[2] Z. Chair and P.K. Varshney, "Distributed Bayesian Hypothesis Testing with Distributed Data Fusion," *IEEE Transactions on Systems, Man and Cybernetics,* Vol. 18, pp. 695–699, 1988.

[3] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization for both uncoded and coded systems," submitted to *IEEE Transactions on Wireless Communications*, 2003. Also available at http://wsl.stanford.edu/Publications.html.

[4] Z.-Q. Luo, "Universal Decentralized Estimation in a Bandwidth Constrained Sensor Network," submitted to *IEEE Transactions on Information Theory*. Also available at http://www.ece.umn.edu/users/luozq/.

[5] Z.-Q. Luo, "An Isotropic Universal Decentralized Estimation Scheme for a Bandwidth Constrained Ad Hoc Sensor Network," submitted to *IEEE Journal on Selected Areas in Communications*. Also available at http://www.ece.umn.edu/users/luozq/.

[6] Z.-Q. Luo, J.-J. Xiao, "Decentralized Estimation in an Inhomogeneous Environment," *Proceedings of IEEE International Syposium on Information Theory*, pp. 520, Chicago, June 2004.

[7] Z.-Q. Luo and J.N. Tsitsiklis, "Data Fusion with Minimal Communication," *IEEE Transactions on Information Theory*, Vol. 40, pp. 1551–1563, 1994.

[8] H.C. Papadopoulos, G.W. Wornell, and A.V. Oppenheim, "Sequential Signal Encoding from Noisy Measurements Using Quantizers with Dynamic Bias Control," *IEEE Transactions on Information Theory*, Vol. 47, pp. 978–1002, 2001.

[9] A.S. Willsky, M. Bello, D.A. Castanon, B.C. Levy, and G. Verghese, "Combining and Updating of Local Estimates and Regional Maps Along Sets of One-dimensional Tracks," *IEEE Transactions on Automatic Control,* Vol. AC-27, pp. 799–813, 1982.

[10] J.-J. Xiao, S. Cui, Z.-Q. Luo and A.J. Goldsmith, "Joint Estimation in Sensor Networks under Energy Constraints ," To be presented in *SECON*, Santa Clara, CA, September 2004.