ON OPTIMAL PARAMETRIC FIELD ESTIMATION IN SENSOR NETWORKS

Ke Liu, Hesham El Gamal

Ohio State University
Electrical and Computer Engineering
2015 Neil Ave, Columbus, OH 43210
{liuk,helgamal}@ece.osu.edu

Akbar M. Sayeed

University of Wisconsin-Madison
Electrical and Computer Engineering
1415 Engineering Dr, Madison, WI 53706
akbar@engr.wisc.edu

ABSTRACT

We develop a framework for field estimation using wireless sensor networks, subject to network power and communication channel constraints. Each field snapshot is described by a real-valued parameter vector and the sensor measurements are assumed independent and identically distributed, conditioned on the parameter values. The nodes communicate appropriate local statistics to a fusion center over a wireless multiple access channel (MAC). If the node statistics satisfy a critical *mean condition*, a simple uncoded communication strategy yields the optimal (centralized) $1/k$ squared-error parameter distortion scaling with the number of nodes (k), even with finite total network power. If an additional *additive property* is satisfied, and the network power grows unbounded (sub-linearly) with the number of nodes, then the uncoded strategy achieves the Cramer-Rao lower bound on distortion. Motivated by these general results, we propose a universal parameter estimation framework based on local type/histogram statistics that satisfies both optimality conditions for arbitrary finite alphabet measurements. It is shown that phase coherent transmission of type statistics achieves the optimal power-distortion scaling even over a fading MAC. When reliable phase synchronization is not possible, a simple coded strategy is proposed that achieves logarithmic distortion reduction with total network power.

1. INTRODUCTION

An essential function of wireless sensor networks is reliable extraction of relevant information about the signal field sensed by the spatially distributed sensor nodes. In this paper, we formulate the information extraction problem as a parameter estimation problem in which the parameters carry the relevant signal field information. The parameters may represent different hypotheses in a detection or classification problem, or more generally arbitrary functions of the signal field. The nodes process and encode their local data measurements to facilitate parameter estimation at a fusion center. The communication link from the sensor nodes to the fusion center naturally forms a *multiple access channel (MAC)* due to the shared nature of the wireless medium. The presence of a noisy and possibly fading MAC places a severe constraint on the reliability of distributed estimation. The performance of centralized estimation strategies, in which the sensor data is perfectly available at the fusion center, serves as a performance benchmark: the optimal square-error parameter estimate distortion generally decays as $1/k$ with the number of sensor nodes k. The network power consumed for communication and the nature of sensor data encoding has a direct impact on estimate distortion in a noisy communication channel. Thus, a key question in distributed parameter estimation is: *what is the fundamental power-distortion trade-off in distributed parameter estimation subject to network resource constraints (power, number of nodes, channel characteristics)*?

An overarching goal of this paper is to study the feasibility and optimality of uncoded communication in general parameter estimation problems. This is motivated by [1] in which estimation of a single random variable (parameter) from noisy measurements is considered; it is shown that uncoded communication achieves the optimal $1/k$ distortion scaling as opposed to the $1/\log(k)$ distortion scaling achieved by conventional coded methods. We derive two general optimality conditions on local node statistics. The first is a crucial *mean condition* which requires that the mean of local node statistics is a sufficiently faithful representation of the parameter space. Under this condition, uncoded strategy is naturally matched to the additive structure of the MAC and achieves the optimal $1/k$ distortion scaling, even with finite total network power. The second is an *additive property* which requires that the global sufficient statistics is the sum of local sufficient statistics. If both conditions are satisfied, and the total network power grows unbounded (at a arbitrarily slow sub-linear rate), then uncoded communication asymptotically achieves the Cramer-Rao lower bound for distortion (achieves the best pre-constant in $1/k$ distortion scaling).

Motivated by these general results, we propose a universal parameter estimation framework for uncoded communication in which the nodes communicate the type or histogram statistic of their local data [2, 3, 4, 5]. It is shown that the type statistics satisfy both optimality conditions for arbitrary finite alphabet node measurements. The type-based framework is particularly attractive for estimation problems since the sensors act as dumb counters: the parametric field representation is only used at the fusion center to compute the final estimate.

Finally we address the impact of channel fading. The mean condition for uncoded communication is violated by fading in general, resulting in an error floor. However, if partial channel phase information is available at the nodes, phase-coherent uncoded communication (beamforming) can achieve $1/k$ distortion scaling even over a fading MAC and with finite total network power. The uncoded strategy breaks down in the absence of reliable phase information. In this case, a simple coded strategy is proposed that achieves a $1/\log P_{tot}(k)$ distortion scaling under a variety of fading conditions.

2. SYSTEM MODEL

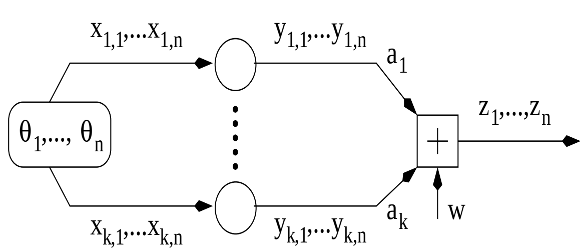

Fig. 1. A schematic illustrating distributed estimation over multiple access channel.

Fig. 1 illustrates a mathematical formulation of distributed estimation over wireless multiple access channels. The signal field information is characterized by a parameter process $\{\boldsymbol{\theta}_j\}$ for $j = 1, \ldots, n$. The parameter $\boldsymbol{\theta}_j$ in the sequence is an m-dimensional real-valued vector from a parameter space $\Theta \subset \mathbb{R}^m$. Sensor readings are induced by the parameter process via an underlining parametric field model. The l-dimensional data measurements are assumed independent and identically distributed (i.i.d.), conditioned on the parameter values, that is, for $i_1 \neq i$ or $j_1 \neq j$,

$$P(\mathbf{x}_{i_1,j_1}, \mathbf{x}_{i,j}) = P_{\boldsymbol{\theta}_{j_1}}(\mathbf{x}_{i_1,j_1}) P_{\boldsymbol{\theta}_j}(\mathbf{x}_{i,j}). \quad (1)$$

We assume identical processing at sensor nodes

$$\mathbf{y}_{i,j} = \pi(\mathbf{x}_{i,j}) \quad (2)$$

where $\mathbf{y}_{i,j}$ is the corresponding t-dimensional sensor output. The design goal for field estimation is to minimize the *mean-squares error (MSE)* of estimation

$$\mathcal{E} = \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 \quad (3)$$

where $\|\mathbf{a}\| = \sqrt{\mathbb{E}[\mathbf{a}^T\mathbf{a}]}$ denotes the L2 norm of random vector $\mathbf{a}$.

Denote by ρ the amplification parameter. The total network power consumption can be written as

$$P_{tot} = k\rho^2. \quad (4)$$

We focus on two extreme cases in terms of network power scaling: 1) *individual power constraint* (IPC) where each sensor node has a constant power budget; and 2) *total power constraint* (TPC) where the total network power is fixed regardless of the network size. The channel signal equation for MAC is given by

$$\mathbf{z} = \rho \sum_{i=1}^{k} a_i \mathbf{y}_i + \mathbf{w} \quad (5)$$

where $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is the t-dimensional white (complex) Gaussian channel noise, independent of source data. In this work we consider both static and fading channels.

In the context of distributed estimation, the MAC channel induces crucial structure constraint on the problem. One approach is the so-called uncoded strategy, in which certain statistics $\mathbf{y}$ of sensor local measurement are sent directly over the channel without additional channel coding and the transmission is analog in the nature. As a result, the receiver gets a superposition of all local statistics plus channel noise. To contrast, the coded strategy calls for the use of digital channel coding to guard again adverse channel effects. Sensor data measurements are digitized and encoded in a rate that yields a reliable communication link to the fusion center. For brevity, we consider a TDMA scheme where each node takes turn to transmit, each with a fraction $1/k$ of the total channel use. Note that the total channel use is kept the same as the uncoded scheme.

A key bottleneck in distributed estimation is the limited channel capacity, which deprives estimation center of a direct (error-free) access to the remote sensor measurements. If data measurements are perfectly available, the problem then essentially reduces to the conventional centralized estimation framework. Although highly unrealistic in wireless sensor networks, this scenario provides an important performance benchmark for studying distributed estimation.

The performance of centralized estimation is governed by the celebrated *Cramer-Rao bound* (CRB) [6, 7], which states that the error covariance matrix of any *unbiased* estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is bounded as follows

$$\mathbf{R} = \mathbb{E}[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}][\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}]^T \geq \frac{1}{k}\mathbf{J}^{-1} \quad (6)$$

where notation $\mathbf{A} \geq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is nonnegative definite. The $m \times m$ matrix $\mathbf{J}$ in the Cramer-Rao bound is the (normalized) *Fisher information matrix* (FIM)

$$\begin{aligned} \mathbf{J} &= -\frac{1}{k}\partial^2_{\boldsymbol{\theta}\boldsymbol{\theta}} \ln P_{\boldsymbol{\theta}}(\mathbf{x}_1, \cdots, \mathbf{x}_k) \\ &= -\,\mathbb{E}\,\partial^2_{\boldsymbol{\theta}\boldsymbol{\theta}} \ln P_{\boldsymbol{\theta}}(\mathbf{x}). \end{aligned} \tag{7}$$

Therefore, one can see that the MSE of centralized estimation decays as $\frac{1}{k}$ with the network size, that is, $\mathcal{E} = \operatorname{tr}\mathbf{R} \geq \frac{1}{k}\operatorname{tr}(\mathbf{J}^{-1})$. For finite k, the Cramer-Rao bound may not be achievable, but there exist many classes of estimators, including *maximal likelihood* estimator, that can attain the bound asymptotically as k increases.

3. OPTIMALITY CONDITIONS FOR UNCODED STRATEGY

3.1. Static Channel, Mean Condition and Additive Property

To best exhibit the optimality conditions, we first consider static MAC channel whose channel coefficients $\{a_i\}$ are deterministic constants. Later in Section 4, we apply the theoretical framework developed in this section to address channel fading. For static channels, one can always re-adjust branch amplification/phase to reduce the problem into the case where $a_i = 1$ for all branches. The corresponding system equation can be written as

$$\mathbf{z} = \rho \sum_{i=1}^{k} \mathbf{y}_i + \mathbf{w}. \tag{8}$$

The mean and variance of $\mathbf{y}$ are given by

$$\mathbb{E}_{\boldsymbol{\theta}}\,\mathbf{y}_i = \mathbf{d}_{\boldsymbol{\theta}}, \quad \operatorname{Var}_{\boldsymbol{\theta}}\mathbf{y}_i = \boldsymbol{\Sigma}_{\boldsymbol{\theta}}, \tag{9}$$

where we have explicitly emphasized the dependence of these quantities upon the hidden parameter $\boldsymbol{\theta}$.

A condition that has profound impact on uncoded scheme is the so-called *mean condition*:

$$\partial_{\boldsymbol{\theta}} \mathbf{d}_{\boldsymbol{\theta}} \neq \mathbf{0}. \tag{10}$$

In other words, the mean of the local statistics $\mathbf{y}_i$ is a faithful representation of the parameter space.

Furthermore, suppose $\mathbf{y}_i$ is a sufficient statistic of local data and its sum over all local $\mathbf{y}_i$'s,

$$\mathbf{y} = \sum_{i=1}^{k} \mathbf{y}_i, \tag{11}$$

is also sufficient for the global data $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_k)$. Then, the statistics $\mathbf{y}_i$'s are said to satisfy the *additive property*.

3.2. Statistical Structure of Uncoded Estimation

We normalize the received signal $\mathbf{z}$ to get a more convenient form

$$\begin{aligned} \tilde{\mathbf{v}} &= \frac{1}{\rho\sqrt{k}}\mathbf{z} = \frac{1}{\sqrt{k}}\sum_i \mathbf{y}_i + \frac{1}{\rho\sqrt{k}}\mathbf{w} \\ &= \mathbf{v} + \tilde{\mathbf{w}} \end{aligned} \tag{12}$$

where $\mathbf{v} = \frac{1}{\sqrt{k}}\sum_i \mathbf{y}_i$, whose variance is $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$, irrespective of the network size. By the network power constraint (4), the equivalent noise $\tilde{\mathbf{w}}$ can be written as

$$\tilde{\mathbf{w}} = \frac{1}{\rho\sqrt{k}}\mathbf{w} = \frac{1}{\sqrt{P_{tot}}}\mathbf{w}. \tag{13}$$

Denote by ϕ the distribution density of the original channel noise $\mathbf{w}$. The density function of the effective noise $\tilde{\mathbf{w}}$ is related to ϕ as

$$P(\tilde{\mathbf{w}}) = (\sqrt{P_{tot}})^{2t}\phi(\sqrt{P_{tot}} \cdot \tilde{\mathbf{w}}) \tag{14}$$

where t is the dimensionality of the (complex) noise signal.

For a function ϕ on $\mathbb{C}^t$ (treated as $\mathbb{R}^{2t}$), scaling by $s > 0$ is defined as

$$\phi_s(\mathbf{x}) = s^{-2t}\phi(s^{-1}\mathbf{x}). \tag{15}$$

Under this terminology, the noise $\tilde{\mathbf{w}}$ represents a scaling of ϕ

$$\tilde{\mathbf{w}} \sim \phi_{s_k}, \quad s_k = \frac{1}{\sqrt{P_{tot}(k)}} \tag{16}$$

where the subscript k is denoted to emphasize the dependence on the network size.

Let $f_{\boldsymbol{\theta},k}$ denote the distribution density of $\mathbf{v}$, induced by parameter $\boldsymbol{\theta}$. The density function of $\tilde{\mathbf{v}}$ is given by

$$\tilde{f}_{\boldsymbol{\theta},k} = f_{\boldsymbol{\theta},k} * \phi_{s_k} \tag{17}$$

where the convolution is a result of independence between noise and data. By central limit theorem, one has[1]

$$f_{\boldsymbol{\theta},k}(\mathbf{u}) = h_{\boldsymbol{\theta},k}(\mathbf{u} - \sqrt{k}\mathbf{d}_{\boldsymbol{\theta}}) \tag{18}$$

with

$$h_{\boldsymbol{\theta},k} \to h_{\boldsymbol{\theta}} = \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}), \quad \text{as } k \to \infty. \tag{19}$$

3.3. Cramer-Rao Bound Analysis

Theorem 1 (Limiting FIM). *For uncoded strategy under static MAC channel, the limiting FIM is given by*

$$\tilde{\mathbf{J}}_\infty = \int \frac{(\mathbf{b}_\infty^T * \phi_{s_\infty}) \cdot (\mathbf{b}_\infty * \phi_{s_\infty})}{h_{\boldsymbol{\theta}} * \phi_{s_\infty}} d\mathbf{u} \tag{20}$$

[1] $\mathbf{u}$ is dummy variable.

where

$$\mathbf{b}_\infty = \partial_{\mathbf{u}} h_{\boldsymbol\theta} \cdot \partial_{\boldsymbol\theta} \mathbf{d}_{\boldsymbol\theta} \tag{21}$$

and

$$s_\infty = \frac{1}{\sqrt{P_{tot}(\infty)}}. \tag{22}$$

Proof. The proof presented here assumes some familiarity with differential computation. Readers are referred to [9] or the like for an introduction.

By convention, the parameter $\boldsymbol\theta$ appears in the subscript, but a notation like $f_{\boldsymbol\theta}(\mathbf{u})$ should really mean $f(\boldsymbol\theta, \mathbf{u})$ when differentiation with respect to $\boldsymbol\theta$ is concerned. We now start a chain of calculations for the FIM associated with $\tilde{\mathbf{v}}$.

$$\begin{aligned}\tilde{\mathbf{J}}_k &= -\frac{1}{k}\,\mathbb{E}\,\partial^2_{\boldsymbol\theta\boldsymbol\theta} \ln \tilde f_{\boldsymbol\theta,k}(\tilde{\mathbf{v}}) \\ &= -\frac{1}{k}\int \partial_{\boldsymbol\theta}\Big(\frac{\partial_{\boldsymbol\theta}\tilde f_{\boldsymbol\theta,k}}{\tilde f_{\boldsymbol\theta,k}}\Big)\cdot \tilde f_{\boldsymbol\theta,k} d\mathbf{u} \\ &= -\frac{1}{k}\int \frac{\tilde f_{\boldsymbol\theta,k}\cdot \partial^2_{\boldsymbol\theta\boldsymbol\theta}\tilde f_{\boldsymbol\theta,k} - \partial_{\boldsymbol\theta}\tilde f^T_{\boldsymbol\theta,k}\cdot\partial_{\boldsymbol\theta}\tilde f_{\boldsymbol\theta,k}}{\tilde f^2_{\boldsymbol\theta,k}}\cdot \tilde f_{\boldsymbol\theta,k} d\mathbf{u} \\ &= \frac{1}{k}\int \frac{\partial_{\boldsymbol\theta}\tilde f^T_{\boldsymbol\theta,k}\cdot\partial_{\boldsymbol\theta}\tilde f_{\boldsymbol\theta,k}}{\tilde f_{\boldsymbol\theta,k}} d\mathbf{u}\end{aligned} \tag{23}$$

where the fact that $\int \partial^2_{\boldsymbol\theta\boldsymbol\theta}\tilde f_{\boldsymbol\theta,k} = \partial^2_{\boldsymbol\theta\boldsymbol\theta}\int \tilde f_{\boldsymbol\theta,k} d\mathbf{u} = 0$ is used to establish the last equality. Plugging in the density relation (17) and using the rule $\partial_{\boldsymbol\theta}(f_{\boldsymbol\theta,k} * \phi_{s_k}) = (\partial_{\boldsymbol\theta} f_{\boldsymbol\theta,k}) * \phi_{s_k}$, one has

$$\tilde{\mathbf{J}}_k = \frac{1}{k}\int \frac{(\partial_{\boldsymbol\theta} f^T_{\boldsymbol\theta,k} * \phi_{s_k})\cdot(\partial_{\boldsymbol\theta} f_{\boldsymbol\theta,k} * \phi_{s_k})}{f_{\boldsymbol\theta,k} * \phi_{s_k}} d\mathbf{u}. \tag{24}$$

In view of (18), function $f_{\boldsymbol\theta,k}$ can be seen as a composition of $h_{\boldsymbol\theta,k}$ with translation $\mathbf{u} \to (\mathbf{u} - \sqrt{k}\mathbf{d}_{\boldsymbol\theta})$. Hence,

$$\tilde{\mathbf{J}}_k = \frac{1}{k}\int \frac{(\partial \mathbf{h}^T * \phi_{s_k})\cdot(\partial \mathbf{h} * \phi_{s_k})}{h_{\boldsymbol\theta,k}(\mathbf{u} - \sqrt{k}\mathbf{d}_{\boldsymbol\theta}) * \phi_{s_k}} \tag{25}$$

where $\partial\mathbf{h}$ is a shorthand for $\partial_{\boldsymbol\theta} h_{\boldsymbol\theta,k}(\mathbf{u} - \sqrt{k}\mathbf{d}_{\boldsymbol\theta})$. Note that the $\mathbf{u} - \sqrt{k}\mathbf{d}_{\boldsymbol\theta}$ is also a function of $\boldsymbol\theta$. So, the chain rule implies that

$$\partial\mathbf{h} = \partial_{\boldsymbol\theta} h_{\boldsymbol\theta,k}(\mathbf{u} - \sqrt{k}\mathbf{d}_{\boldsymbol\theta}) - \partial_{\mathbf{u}} h_{\boldsymbol\theta}(\mathbf{u} - \sqrt{k}\mathbf{d}_{\boldsymbol\theta}) \cdot \sqrt{k}\partial_{\boldsymbol\theta}\mathbf{d}_{\boldsymbol\theta}. \tag{26}$$

Now, substituting (26) back into (25) and removing translation, one has

$$\tilde{\mathbf{J}}_k = \int \frac{(\mathbf{b}^T_k * \phi_{s_k})\cdot(\mathbf{b}_k * \phi_{s_k})}{h_{\boldsymbol\theta,k} * \phi_{s_k}} d\mathbf{u} \tag{27}$$

where

$$\mathbf{b}_k = -\frac{\partial_{\boldsymbol\theta} h_{\boldsymbol\theta,k} - \sqrt{k}\partial_{\mathbf{u}} h_{\boldsymbol\theta,k}\cdot\partial_{\boldsymbol\theta}\mathbf{d}_{\boldsymbol\theta}}{\sqrt{k}}. \tag{28}$$

Taking $k \to \infty$, the $\mathbf{b}_k$ in the limit is given by

$$\mathbf{b}_\infty = \partial_{\mathbf{u}} h_{\boldsymbol\theta} \cdot \partial_{\boldsymbol\theta} \mathbf{d}_{\boldsymbol\theta} \tag{29}$$

where $h_{\boldsymbol\theta}$ is the Gaussian density of $\mathcal{CN}(\mathbf{0}, \boldsymbol\Sigma_{\boldsymbol\theta})$, according to (19). Therefore, the limiting value of FIM is settled as

$$\tilde{\mathbf{J}}_\infty = \int \frac{(\mathbf{b}^T_\infty * \phi_{s_\infty})\cdot(\mathbf{b}_\infty * \phi_{s_\infty})}{h_{\boldsymbol\theta} * \phi_{s_\infty}} d\mathbf{u} \tag{30}$$

where $s_\infty = \lim_{k\to\infty} \frac{1}{\sqrt{P_{tot}(k)}}$. If $P_{tot}(k)$ is unbounded, then $s_\infty = 0$ and $\phi_{s_\infty} = \delta$, the *Dirac delta function*. But when $P_{tot}(\infty)$ is finite, the function ϕ_{s_∞} is non-trivial under convolution. □

Any distributed estimation scheme is said to be *first-order* optimal, if the estimation MSE asymptotically decays as $1/k$. In addition, if the scheme actually achieves the centralized benchmark, it is called *second-order* optimal since it has the best decay pre-constant.

Theorem 2 (First Order Optimality). *The uncoded strategy is first-order optimal, regardless of network power scaling, if and only if the mean condition holds*

$$\partial_{\boldsymbol\theta}\mathbf{d}_{\boldsymbol\theta} \neq \mathbf{0}. \tag{31}$$

Otherwise, uncoded schemes exhibit error-floor, that is, the estimation MSE cannot be driven down to zero no matter how large the network size or the total power is.

In a sense, the mean condition is a primary design factor for uncoded strategy. When the local statistics satisfy this condition, the estimation MSE decays as $1/k$ with the number of sensor nodes, even under a finite total network power constraint (TPC).

Theorem 3 (Second Order Optimality). *Assume the mean condition for local statistics. If, in addition, the additive property is also satisfied by the same local statistics, then the uncoded strategy achieves the ideal centralized benchmark as the network size increases, provided the total network power grows* ***unbounded*** *with k.*

3.4. Universal Estimation Based on Types

As seen from the above analysis, the mean condition and additive property are essential prerequisites for optimal uncoded strategy over the MAC channel. For discrete random variables with finite alphabet, there exists a universal statistic, the *type* of data measurements, that automatically satisfies the optimality conditions. Let $\mathcal{A}$ be the finite alphabet. The type $\mathbf{T}$ of a sequence $\mathbf{x} = (x_1, \ldots, x_l)$ is its empirical distribution or histogram:

$$\mathbf{T}(a) = N(a|\mathbf{x})/l, \quad a \in \mathcal{A}, \tag{32}$$

where $N(a|\mathbf{x})$ denotes the number of occurrences for alphabet symbol a in the sequence $\mathbf{x}$. For example, a length-4 sequence $1, 0, 1, 1$ has a type $\mathbf{t} = (3/4, 1/4)$, that is, 3 occurrences for alphabet '1' and 1 for '0'.

Given an i.i.d. sequence $\mathbf{x}$, its (normalized) likelihood can be written as

$$\frac{1}{l} \log P_{\boldsymbol{\theta}}(\mathbf{x}) = [\log \mathbf{P}_{\boldsymbol{\theta}}]^T \mathbf{T} \tag{33}$$

where $\log \mathbf{P}_{\boldsymbol{\theta}} = [\log P_{\boldsymbol{\theta}}(a_1), \ldots, \log P_{\boldsymbol{\theta}}(a_A)]^T$ is a vector of likelihood weights derived from $P_{\boldsymbol{\theta}}$. Since the dependence of data likelihood is via types, the type $\mathbf{T}$ of $\mathbf{x}$ is a sufficient statistic of parameter $\boldsymbol{\theta}$. Moreover, the type statistic is sufficient for arbitrary distribution family $P_{\boldsymbol{\theta}}$ with finite alphabet. It is easy to verify that the mean of type statistics is the corresponding distribution itself

$$\mathbf{d}_{\boldsymbol{\theta}} = \mathbb{E}_{\boldsymbol{\theta}} \, \mathbf{T} = \mathbf{P}_{\boldsymbol{\theta}}, \tag{34}$$

which obviously contains all the information about parameter $\boldsymbol{\theta}$.

Suppose there are k i.i.d. data $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_k)$. The global type statistic of $\mathbf{X}$ can be computed as

$$\mathbf{T}(a) = \frac{N(a|\mathbf{x})}{kl} = \frac{\sum_{i=1}^k N(a|\mathbf{x}_i)}{kl} = \frac{1}{k} \sum_{i=1}^{k} \mathbf{T}_i(a), \tag{35}$$

or, in other words, the global type statistic is the average of all local type statistics $\mathbf{T} = \frac{1}{k} \sum_i \mathbf{T}_i$. Clearly, type statistics satisfy the additive property.

Therefore, type statistics are a class of universal sufficient statistics that enjoy additive property and carry non-trivial parameter information in the mean. According to our general analysis on uncoded scheme, type-based uncoded estimation achieves the optimal asymptotic performance. But, most importantly, as we can see from (33), the data likelihood value is effectively computed at the fusion center. Therefore, sensor nodes can be oblivion of the detailed knowledge about the parametric field signal model, which leads to an attractive solution based on the idea of *dumb sensor* [3]. Due to space limitation, we refer readers to [2, 3, 4] for more detail about using types in sensor networks.

4. SCALING LAWS FOR FADING CHANNELS

In this section, we investigate the impact of fading MAC channels on distributed estimation. The presence of channel fading poses a great challenge, especially for uncoded scheme. To simplify the discussion, we assume the channel coefficients a_i's are i.i.d. random variables with unit variance. As shown in the following, channel statistics and channel state information (CSI) have profound impact on uncoded strategy.

4.1. No CSI

We first consider the situation when no CSI is available at the sensor nodes. The uncoded scheme sends the raw statistic $\mathbf{y}_i$ over the fading channel. But now, its performance bifurcates depending on the mean of channel coefficients $\mathbb{E}\, a_i$. The system equation can be written as

$$\mathbf{z} = \rho \sum_{i=1}^{k} a_i \mathbf{y}_i + \mathbf{w} = \rho \sum_{i=1}^{k} \mathbf{y}'_i + \mathbf{w} \tag{36}$$

where $\mathbf{y}'_i = a_i \mathbf{y}_i$ is the local statistic under an equivalent static channel. Generally speaking, statistic $a_i \mathbf{y}_i$ will hardly be sufficient for parameter $\boldsymbol{\theta}$ due to the randomness in a_i, and hence additive property fails to hold for the effective statistic $\mathbf{y}'_i$. So, one could not expect uncoded scheme to achieve the ideal centralized benchmark for fading channels. But, the following is true

$$\partial_{\boldsymbol{\theta}} \, \mathbb{E}\, \mathbf{y}'_i = (\mathbb{E}\, a_i) \partial_{\boldsymbol{\theta}} \mathbf{d}_{\boldsymbol{\theta}}. \tag{37}$$

So, if channel coefficients have *non-zero* mean, the estimation MSE can still decays as $1/k$ in uncoded scheme according to our previous analysis. However, if channel coefficients have zero mean, such as in the case of Rayleigh fading channels, then uncoded strategy would completely break down.

As a comparison, coded strategy is very resilient under channel fading. Instead of completely ignoring noise as in uncoded methods, various coded schemes follow a *separate* source-channel coding approach; source are quantized to match the capacity of the MAC channel. To understand the uncoded strategy, we exam a simple scheme using source quantization and TDMA transmission protocol.

Here we code across different blocks in general. Every l-dimensional data block $\mathbf{x}_i$ is digitized into $l\Delta$ bits, which corresponds to a size-Δ quantization alphabet per dimension. Note that quantization will reduce estimation performance, however, the estimation MSE can still maintain a $1/r$ decay, assuming there are r i.i.d. measurements for every source parameter $\boldsymbol{\theta}$.

The total channel use is fixed to be t per source parameter. Thus, for a block of n source parameters, every sensor node is allocated with nt/k channel uses. We choose a nominal bit-budget B for each sensor:

$$B = \frac{nt}{k} \log P_{tot}. \tag{38}$$

Since each data block contributes $l\Delta$ bits, the corresponding number of source symbols is

$$D = \frac{B}{l\Delta} = \frac{nt}{kl\Delta} \log P_{tot}. \tag{39}$$

Although each sensor can only send the (quantized) data measurement corresponding to D source symbols, they can collaborate with each other to cover all the n source symbols, collectively. One simple scheme is as follows: sensor 1 covers source symbol 1 – D, sensor 2 covers 2 – $(D+1)$, and so on, that is, sensor nodes send a length-D block in a cyclic fashion. It is easy to check that every source symbol is covered by equal amount of times

$$r = \frac{kD}{n} = \frac{t}{l\Delta} \log P_{tot}, \tag{40}$$

or, in other words, for every source parameter $\boldsymbol{\theta}$, there are r number of data measurements sent by sensor nodes collectively.

However, not all transmission would be successful due to channel fading outage. Let γ be the probability of a successful transmission, which occurs when the instantaneous channel capacity is larger than the bit budget:

$$\gamma = P(B \leq \frac{nt}{k} \log |a|^2 P_{tot}) = P(|a|^2 \geq 1) \tag{41}$$

where a is the channel fading coefficient. Thus, on average, a γr number of data measurements are reliably received by the estimation center for each source symbol, which implies an MSE scaling law of $1/\gamma r$. Since r is related to P_{tot} logarithmically, we have the following.

Theorem 4 (Logarithmic Decay). *The coded strategy exhibits* $\log P_{tot}(k)$ *decay in MSE. In particular, the decay is on the order of* $\log k$ *for individual power constraint (IPC), while error-floor occurs for total power constraint (TPC).*

4.2. CSI Feedback

When CSI is available at the sensor nodes, uncoded strategy can try to "actively" compensate the channel fading. The key is to deconstructing composite fading coefficients with nonzero mean.

Besides the channel inversion which inverts the channel completely (when $\mathbb{E}\,\frac{1}{|a|^2}$ is bounded), a very important approach is the *beam-forming*. Let b_i be the feedback CSI for the channel coefficient a_i. Every sensor node then rotates the transmit signal by the corresponding CSI. The resulting composite channel coefficient becomes

$$c_i = b_i^* a_i. \tag{42}$$

As long as $\mathbb{E}\, c_i \neq 0$, uncoded strategy achieves $1/k$ scaling. A perfect CSI feedback ($b_i = a_i$) will certainly do the job. But, noisy CSI feedback is generally sufficient to deconstruct a composite channel with non-zero mean. This suggests a little feedback is actually quite efficient in terms of scaling laws.

5. CONCLUSIONS

We have studied the asymptotic performance of distributed estimation over MAC channels. The optimality conditions given in this paper provide a good principles to design uncoded strategy for sensor networks. Moreover, type-based framework is shown to be a universal solution for discrete random variables. Although we have focused mainly on estimation problem, the parallel results can be obtained for detection problem as well.

6. REFERENCES

[1] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: Lossy source-channel communication revisited," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003.

[2] K. Liu and A. M. Sayeed, "Asymptotically optimal decentralized type-based detection in wireless sensor networks," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'04)*, Montreal, Quebec, Canada, May 2004, vol. 3, pp. 873–876.

[3] K. Liu and A. M. Sayeed, "Optimal detection strategies for wireless sensor networks," in *the 42nd Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, USA, Oct. 2004.

[4] G. Mergen and L. Tong, "Estimation over multiaccess channels," submitted to IEEE Trans. Signal Processing, July 2004.

[5] A. Giridhar and P. R. Kumar, "Computing and communicating functions over sensor networks," submitted to IEEE J. Select. Areas Commun., July 2004.

[6] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, Addison-Wesley, New York, 1990.

[7] Jun Shao, *Mathematical Statistics*, Springer, New York, 2nd edition, 2003.

[8] G. B. Folland, *Real Analysis: Modern Techiques and Their Applications*, Wiley, New York, 2nd edition, 1999.

[9] M. Spivak, *Calculus on Manifolds*, Benjamin, New York, 1965.