Parameter Estimation Over Noisy Communication Channels in Distributed Sensor Networks

Thakshila Wimalajeewa[1], Sudharman K. Jayaweera[1] and Carlos Mosquera[2]
[1]Dept. of Electrical and Computer Engineering, University of New Mexico,
[2]Dept. de Teoría de la Señal y Comunicaciones, Universidad de Vigo,
[1]USA
[2]Spain

1. Introduction

With the recent development of low-cost, low-power, multi-functional sensor nodes, sensor networks have become an attractive emerging technology in a wide variety of applications including, but not limited to, military surveillance, civilian, industrial and environmental monitoring [1]–[5]. In most of these new applications sensor nodes are capable not only of sensing but also of data processing, wireless communications and networking. It can be argued that it is their ability of ad hoc wireless networking that has attracted much interest to wireless sensor networks in recent years.

A typical sensor network may consist of a large number of spatially distributed nodes to make a decision on a Parameter of Interest (PoI). This can be detection, estimation or tracking of a target or multiple targets. Once the network is deployed, the network resources, such as node power and communication bandwidth, are limited in many situations. This is due to the fact that reinstalling and recharging the batteries might be difficult, or even impossible, once the network is deployed. A common question arising in such networks is how to effectively combine the information from all the nodes in the network to arrive at a final decision while consuming the resources in an optimum way. In a distributed sensor network, the distributed nodes make observations of PoI and process them locally to make a summary of their observations. The final decision is usually made by combining these locally processed data.

Once local decisions are made at each individual sensor node, the natural questions are how to combine the local decisions and where the final decision is made. When there is a possibility that the sensor network can have a central node (generally called as the fusion center) with relatively high processing power compared to distributed nodes, the summary of the local observations can be sent to the fusion center. The fusion center combines all local decisions in an optimum way to arrive at a final decision in what is known as the centralized architecture. The disadvantage of such a system is that if there is a failure in the fusion center, the whole network fails. On the other hand, in some applications, it might be of interest that nodes communicate with each other to reach at a final decision without

depending on a central node. In this set-up the node that makes the final decision may change over time due to the dynamic nature of the sensor network and the PoI. This may lead to a more robust architecture compared to that with centralized architecture.

Irrespective of the data fusion architecture, the local information from sensor nodes needs to be shared over a communication channel that, in general, can undergo both path loss attenuation and multipath fading. As a result, the received signal at a destination node, be it another distributed node or a central fusion center, is corrupted by both multiplicative and additive noise. The performance of the final decision will depend not only on measurement noise at the distributed nodes but also on channel quality of communication links.

The performance of resource-constrained wireless sensor networks with communication and measurement noise has been addressed by many authors in different contexts. For example, performance of the sensor networks under power and bandwidth constraints are analyzed in [6]–[22] and [10], [11], [23]–[36], respectively. Collaborative signal processing, including sequential communication, is addressed in [15], [37]–[42].

In this chapter we address the problem of muti-sensor data fusion over noisy communication channels. The objective of the sensor network is to estimate a deterministic parameter. Distributed nodes make noisy observations of the PoI. Each node generates either an amplified version of its own observation or a quantized message based on its own observation, and shares it with other nodes over a wireless channel. The final decision can be made either at a central node (fusion center) or fully distributively. In the case of centralized architecture, the locally processed messages can be sent to the fusion center over a set of orthogonal channels or a multiple-access channel in which nodes share a common communication channel. In the fully distributive architecture, there is no explicit central fusion center and nodes communicate with each other to arrive at a final decision. There are several variations of this architecture: in one setting, nodes may communicate sequentially with neighbors to sequentially update an estimator (or a sufficient statistic for the parameter). The final decision can be declared by any node during this sequential updating process. On the other hand, it might sometimes be of interest for all nodes in the network to arrive at a common final decision. This leads to a distributed consensus estimation problem. Note that, here all nodes communicate with each other in contrast to the sequential communication architecture above until they reach an agreement.

The rest of this chapter is organized as follows: Section 2 formulates the problem of parameter estimation over noisy communication channels in a distributed sensor network. Section 2-A presents the assumed observation model. The ideal centralized estimation is reviewed in Section 2-B. Ideal estimation is, of course, not possible in a wireless sensor network since communication is over a noisy channel and the network is constrained by available communication resources. By sharing only a summary of the observations with each other, the scarce communication resources can be efficiently utilized. Local processing schemes to achieve this goal are discussed in Section 2-C.

Section 3 focuses on centralized estimation architecture with noisy communication channels between distributed nodes and the fusion center. Estimation performance with orthogonal channels is discussed in Section 3-A and that with non-orthogonal communication channels is discussed in Section 3-B.

Sections 4 and 5 discuss the distributed estimation performance in a sensor network with collaborative information processing. Section 4 considers the distributed sensor network architecture with sequential communication where inter-sensor communication links are

assumed to be noisy. In Section 5 collaborative estimation with distributed consensus is addressed. Here, the nodes are allowed to communicate with a set of other nodes that are considered as their neighbors. Sections 5-A and 5-B address static parameters whereas section 5-C considers, time-varying parameters. Finally chapter summary is given in section 6.

2. Data fusion problem in a wireless sensor network

Throughout, we consider a spatially distributed sensor network that is deployed to estimate a PoI. It is natural to expect that a final decision be obtained by combining the information from different nodes. In a distributed sensor network, nodes share summaries of their observations over noisy communication channels. Since network resources, in particular the node power and the communication bandwidth, are scarce it is important that the observations at each node are locally processed to reduce the observation to a concise summary. The final decision can then be made based on these local outputs that nodes share with each others and/or with a fusion center.

A. Multi-sensor observation model

We consider a situation in which multiple sensors observe a PoI. When these nodes form a sensor network, the final decision can be made in either a centralized or distributed way. In the centralized architecture, each node sends a summary of its observations to a central node called a fusion center. There is no inter-node communication. The fusion center combines all received information in an effective way to arrive at a final decision. In the distributed decision-making architecture, on the other hand, the nodes collaborate with each other to arrive at a final decision distributively, without the aid of a central fusion node.

Irrespective of the architecture, communication between sensors and the fusion center, or among sensors, is over a noisy channel. Thus, the information sent sees distortion due to both additive as well as multiplicative noise. The multiplicative noise is due to path loss attenuation and multipath fading encountered, for example, in a wireless channel. In this section, we first consider the centralized architecture as shown in Fig. 1. The distributed architecture is covered in Sections 4 and 5.

Consider a spatially distributed network of n sensors. Let us assume that the network is to estimate, in general, a vector parameter Θ where Θ is a p-vector. The observation at each node is related to the parameter Θ that we wish to estimate via the following observation model;

$$z_k(t) = f_k(\Theta) + v_k(t),$$

where $z_k(t)$ is the observation at the k-th node at time t, $f_k : \mathbb{R}^p \rightarrow \mathbb{R}$ is a function of the parameter vector Θ (in general, non-linear) and $v_k(t)$ is the additive observation noise at node k. In the special case of linear observation model, the joint observation vector at n nodes at time t can be written as

$$\mathbf{z}(t) = \mathbf{B}\Theta + \mathbf{v}(t), \tag{1}$$

where $\mathbf{B}$ is an $n \times p$ (known) matrix and $\mathbf{v}$ is the observation noise vector having a zero mean and a covariance matrix of $\Sigma_\mathbf{v}$. In this chapter we focus mainly on scalar parameter

estimation (where we assume $p = 1$) although the techniques developed and the results can easily be extended to vector parameter estimation. For a scalar parameter, the observation vector (1) formed by observations at all n nodes reduces to,

$$\mathbf{z} = \theta\mathbf{e} + \mathbf{v}, \tag{2}$$

where we have suppressed the timing index t and $\mathbf{e}$ is the n-vector of all ones.
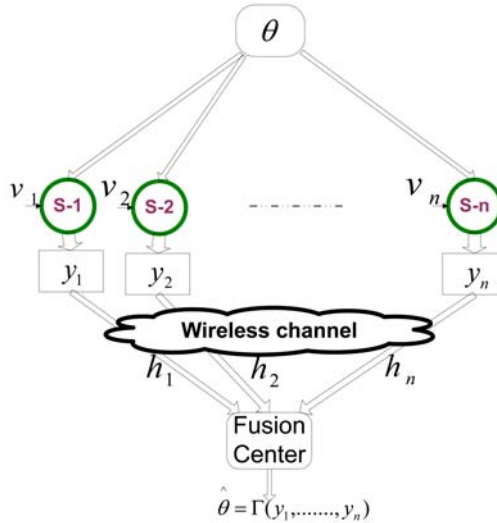

Fig. 1. Distributed estimation with a central fusion center

B. Ideal centralized data fusion

When local observation vector $\mathbf{z}$ is directly available at the fusion center, the problem is termed the ideal centralized data fusion. The optimal final estimator and its mean-squared error performance are summarized in the following lemma:

Lemma 1: [43] *When the observation vector (2) is available at the fusion center, the best linear unbiased estimator (BLUE) for the scalar parameter θ is given by*

$$\hat{\theta}(\mathbf{z}) = \frac{\mathbf{e}^T\boldsymbol{\Sigma_v}^{-1}\mathbf{z}}{\mathbf{e}^T\boldsymbol{\Sigma_v}^{-1}\mathbf{e}}, \tag{3}$$

where x^T denotes the transpose of x. The corresponding mean squared error (MSE) achieved by (3) is

$$MSE = \mathbb{E}\{|\hat{\theta} - \theta|^2\} = \left(\mathbf{e}^T\boldsymbol{\Sigma_v}^{-1}\mathbf{e}\right)^{-1}, \tag{4}$$

where $\mathbb{E}\{.\}$ denotes the mathematical expectation. Further, if the local observations are i.i.d., so that $\boldsymbol{\Sigma_v} = \sigma_v^2\mathbf{I}$, *where $\mathbf{I}$ is the $n \times n$ identity matrix, the estimator in (3) simplifies to the sample mean of the observations,*

$$\hat{\theta}(\mathbf{z}) = \frac{1}{n}\sum_{k=1}^{n} z_k,$$

with the corresponding MSE in (4) simplified to

$$MSE = \frac{\sigma_v^2}{n}.$$

Since communication from distributed nodes to the fusion center is over noisy channels, in practice signals transmitted by the distributed nodes undergo distortion. Hence, direct access from a distant fusion center to the exact observations at distributed nodes may not be possible. However, the lemma 1 will serve as a benchmark for other schemes that we will discuss in this chapter.

C. Local processing at sensor nodes

To facilitate efficient utilization of node and network resources, each node in a sensor network locally processes its observation to generate a useful summary. The transmitted signal at the k-th node is then given by $y_k = g_k(z_k)$. In the following we consider two local processing schemes:

1) Amplify-and-Forward (AF) local processing: In many practical situations where sensor observations are corrupted by additive noise, the amplify and forward strategy has been shown to perform well. In this method, each node directly amplifies its observation and sends it to the fusion center. The transmitted signal from node k is

$$y_k = g_k z_k,$$

where g_k is the amplifying gain at the k-th node. In order to save the node power, it is important to select the amplification gain g_k for $k = 1,..., n$ appropriately depending on the other network parameters such as channel quality and observation quality, etc.. If nodes are operated at the same power level, sometimes it may lead to an unnecessary usage of the network power especially when observation qualities of nodes and channel qualities are not the same for all nodes. Therefore, choosing g_k's in a meaningful way is an important issue to be addressed in designing resource-constrained sensor networks. This problem is discussed in section 3.

With AF local processing, the received signal vector at the fusion center with noiseless communication is given by

$$\mathbf{r} = \mathbf{G}\mathbf{z}, \tag{5}$$

where $\mathbf{G} = \text{diag}(g_1,..., g_n)$ is the channel gain matrix. Then the Best Linear Unbiased Estimator and its corresponding mean squared error is given by the following lemma:

Lemma 2: [34], [43] *If the received signal at the fusion center is as given in (5), then the BLUE estimator based on the received signal vector is given by*

$$\hat{\theta}(\mathbf{r}) = \frac{\mathbf{e}^{\mathbf{T}}\mathbf{\Sigma_v}^{-1}\mathbf{G}^{-1}\mathbf{r}}{\mathbf{e}^{\mathbf{T}}\mathbf{\Sigma_v}^{-1}\mathbf{e}},$$

and the corresponding MSE is

$$MSE = \mathbb{E}\{|\hat{\theta} - \theta|^2\} = \left(\mathbf{e}^{\mathbf{T}}\mathbf{\Sigma_v}^{-1}\mathbf{e}\right)^{-1}.$$

Further, if the local observations are i.i.d., so that $\mathbf{\Sigma_v} = \sigma_v^2\mathbf{I}$, *the MSE simplifies to*

$$MSE = \frac{\sigma_v^2}{n}.$$

Fig. 2. Probabilistic quantization

2) Quantized local processing: To save node energy and communication bandwidth, sensors can compress their observations before transmitting to the fusion center. In this set up, local nodes quantize their observations to generate finite-range messages $m_k(z_k)$ where each m_k is represented by L_k number of bits [9]. Based on the quantized messages received from nodes, the fusion center computes the final estimator

$$\hat{\theta} = \Gamma(\hat{m}_1, \hat{m}_2, \cdots, \hat{m}_n)$$

where $\hat{m}_k$'s are the corrupted versions of quantized messages m_k's received at the fusion center and $\Gamma(.)$ is the final estimator mapping.

There are several quantization schemes proposed in the literature each having its own advantages and disadvantages [9], [22], [21]. For simplicity, throughout this chapter we concentrate on the universal decentralized quantization scheme given in [9]. According to this scheme, each node locally quantizes its own observation z_k into a discrete message $m_k(z_k, L_k)$ of L_k bits. Due to the lack of knowledge of probability density function (pdf) of noise, the quantizer $Q_k : z_k \rightarrow m_k(z_k, L_k)$ at local nodes is designed to be a uniform randomized quantizer [9]. To that end, suppose the observation range of each sensor is [-W,W] where W is a known parameter determined by the physical properties of the sensor nodes. At each node the range [-W, W] is divided into 2^{L_k}-1 intervals of length $\Delta_k = 2W/(2^{L_k} - 1)$ each as shown in Fig. 2. The quantizer Q_k rounds-off z_k to the nearest endpoint of one of these intervals in a probabilistic manner. For example, suppose, $i\Delta_k \leq z_k < (i + 1)\Delta_k$ where $-2^{L_k-1} \leq i \leq 2^{L_k-1}$. Then Q_k will quantize z_k into $m_k(z_k, L_k)$ so that

$$P\{m_k(z_k, L_k) = i\Delta_k\} = 1 - r,$$

and

$$P\{m_k(z_k, L_k) = (i + 1)\Delta_k\} = r,$$

where $r \equiv (z_k - i\Delta_k)/\Delta_k \in [0, 1]$. Note that the quantizer noise $q_k(z_k, L_k) = m_k(z_k, L_k) - z_k$ is then a Bernoulli random variable taking values of $q_k(z_k, L_k) = -r\Delta_k$ and $q_k(z_k, L_k) = (1 - r)\Delta_k$ with probabilities

$$P\{q_k(z_k, L_k) = -r\Delta_k\} = 1 - r,$$

and

$$P\{q_k(z_k, L_k) = (1 - r)\Delta_k\} = r.$$

With this local processing scheme the quantized message at node k can be expressed as

$$m_k(z_k, L_k) = z_k + q_k = \theta + v_k + q_k; \text{ for } k = 1, \cdots, n, \tag{6}$$

where we have made use of (2). Note that the quantization noise q_k and the observation noise v_k in (6) will be assumed to be independent. Moreover, q_k is independent across sensors since quantization is performed locally at each sensor.

It can be easily shown that $m_k(z_k, L_k)$ is an unbiased estimator of θ so that $\mathbb{E}\{m_k\} = \theta$ with the MSE (which is, in fact, the variance of the estimator) upper bounded as

$$V_k(m_k) \leq \frac{W^2}{(2^{L_k} - 1)^2} + \sigma_v^2 = \delta_k^2 + \sigma_v^2 \quad \text{for} \quad k = 1, \cdots, n, \tag{7}$$

where $\delta_k^2 = \frac{W^2}{(2^{L_k}-1)^2}$. Hereafter we use the short-hand notation m_k to denote $m_k(z_k, L_k)$, so that the transmitted signal y_k at node k is $y_k = m_k$ for $k = 1, ..., n$.

With quantized processing, the received signal vector at the fusion center, with noiseless communication is

$$\mathbf{r} = \mathbf{m} = \theta \mathbf{e} + \mathbf{v} + \mathbf{q} \tag{8}$$

where $\mathbf{q} = [q_1, ..., q_n]^T$ is the quantization noise vector and $\mathbf{m} = [m_1, ..., m_n]^T$. The BLUE estimator at the fusion center and its performance are characterized in lemma 3 below.

Lemma 3: [9] *The BLUE estimator based on the received signal in (8) is given by*

$$\hat{\theta}(\mathbf{m}) = \frac{\mathbf{e}^T (\Sigma_\mathbf{v} + \Sigma_\mathbf{q})^{-1} \mathbf{m}}{\mathbf{e}^T (\Sigma_\mathbf{v} + \Sigma_\mathbf{q})^{-1} \mathbf{e}},$$

where $\Sigma_\mathbf{q} = diag(\delta_1^2, ..., \delta_n^2)$. *An upper bound for the MSE of above estimator can be found to be (using(7))*

$$MSE = \mathbb{E}\{|\hat{\theta} - \theta|^2\} \leq \left(\mathbf{e}^T (\Sigma_\mathbf{v} + \Sigma_\mathbf{q})^{-1} \mathbf{e}\right)^{-1}. \tag{9}$$

When local observations are i.i.d. the MSE upper bound (9) can be further simplified as

$$MSE \leq \left(\sum_{k=1}^{n} \frac{1}{\sigma_v^2 + \delta_k^2}\right)^{-1}. \tag{10}$$

Of course, in practice the above ideal estimators cannot be realized due to imperfect communications between distributed nodes and the fusion center. These imperfections can be due to multiplicative noise (caused by channel fading and path loss attenuation) and additive noise at the receiver. When the sensor system has to conform with resource constraints on node power and communication bandwidth, it is important to consider the minimum achievable error performance taking into account these channel imperfections. Parameter estimation under imperfect communications in a distributed sensor network is discussed in the next section.

3. Optimal decision fusion over noisy communication channels

The performance of a final estimator when locally processed data are transmitted to the destination over a noiseless channel was discussed in the latter part of Section 2. In this section we discuss the final estimator performance at a fusion center in the presence of noisy communication channels from distributed nodes to the fusion center. In the following we

consider two communication schemes where sensors transmit data over orthogonal or non-orthogonal channels.

A. Communication over orthogonal channels

When locally processed sensor data are transmitted through orthogonal channels (either TDMA, FDMA or CDMA), the received signal vector at the fusion center can be written as

$$\mathbf{r} = \mathbf{H_c u} + \mathbf{w}, \tag{11}$$

where $\mathbf{H_c} = \text{diag}(h_1, ..., h_n)$ are the fading coefficients of each channel and $\mathbf{w}$ is the receiver noise vector with mean zero and the covariance matrix $\mathbf{\Sigma_w}$. Note that in (11) we have assumed flat fading channels between sensors and the fusion center which can be a reasonable assumption in certain WSN's but not all. When the channels are selective one can modify (11) by using a tapped-delay line model. The statistics of h_k is determined by the type of fading distributions. Throughout this chapter we will assume that h_k's are Rayleigh distributed.

1) AF local processing: With AF local processing and orthogonal communication channels, the received signal vector at the fusion center is given by

$$\begin{aligned}\mathbf{r} &= \mathbf{H_c G z} + \mathbf{w} \\ &= \mathbf{H_c G e}\theta + \mathbf{n}, \end{aligned} \tag{12}$$

where $\mathbf{n} = \mathbf{H_c G v} + \mathbf{w}$ is the effective noise vector at the fusion center with mean zero and covariance matrix $\Sigma_n = \mathbf{H_c G \Sigma_v G H_c} + \mathbf{\Sigma_w}$, assuming that the receiver noise and the node observation noise are independent. In the following lemma we summarize the optimal estimator at the fusion center based on the received signal (12) and its performance:

Lemma 4: [34] *If the fusion center has the knowledge of channel fading coefficients, the BLUE estimator and the MSE based on the received signal (12) is given by*

$$\hat{\theta}(\mathbf{r}) = \frac{\mathbf{e^T G H_c \Sigma_n^{-1} r}}{\mathbf{e^T G H_c \Sigma_n^{-1} H_c G e}}, \tag{13}$$

and

$$MSE = \left(\mathbf{e^T G H_c \Sigma_n^{-1} H_c G e}\right)^{-1}. \tag{14}$$

In the special case when local observations and the receiver noise are both i.i.d. such that $\mathbf{\Sigma_v} = \sigma_v^2 \mathbf{I}$ *and* $\mathbf{\Sigma_w} = \sigma_w^2 \mathbf{I}$, *the BLUE estimator (13) and the MSE (14) further simplify to*

$$\hat{\theta}(\mathbf{r}) = \frac{\sum_{k=1}^n \frac{g_k h_k r_k}{\sigma_v^2 h_k^2 g_k^2 + \sigma_w^2}}{\sum_{k=1}^n \frac{h_k^2 g_k^2}{\sigma_v^2 h_k^2 g_k^2 + \sigma_w^2}}, \tag{15}$$

and

$$MSE = \left(\sum_{k=1}^n \frac{h_k^2 g_k^2}{\sigma_v^2 h_k^2 g_k^2 + \sigma_w^2}\right)^{-1}, \tag{16}$$

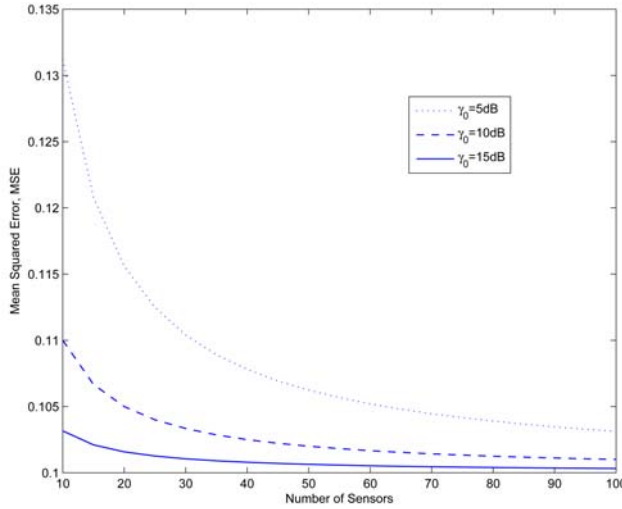*where* σ_w^2 *is the receiver noise power.*

Fig. 3. Mean squared error performance vs. number of nodes: The total network power is constant.

The performance of the BLUE estimator (15) is shown in Figs. 3 and 4 given that the total power in the network is constant. Note that in the both Figs. 3 and 4 the node power is the same at each sensor, so that $g_k = g$ for $k = 1,..., n$ and each channel gain is unity (i.e. $h_k = 1$ for all k). Hence, if total network power is P_T then the individual node power is given by $g^2 = P_T/n$. In this case, the MSE in (16) is further simplified to $MSE = \frac{\sigma_v^2}{n} + \frac{\sigma_w^2}{P_T}$. The local SNR, γ_0 is defined as P_s/σ_v^2 where P_s is the average power at local nodes. In the simulations we have let $P_s = 1$. It can be seen that when either the number of sensors or the total network power is increased, the performance of the BLUE estimator is floored: i.e. $\lim_{n \to \infty} MSE \approx \frac{\sigma_w^2}{P_T}$ and $\lim_{P_T \to \infty} MSE \approx \frac{\sigma_v^2}{n}$. The first of these limits is illustrated in Fig. 3 for a constant total network power, as parameterized by the local observation SNR γ_0. It is seen from Figs. 3 and 4 that when local SNR is high the system shows better performance which intuitively makes sense. From Fig. 4, it can be seen that in the region of low local SNR, the performance of the system can be improved by increasing the number of nodes. But in high local SNR region, increasing the number of nodes may not affect the final performance much since ultimately the performance is limited by the channel quality between nodes and the fusion center.

Allocating equal power for all nodes may not result in the best performance since all nodes may not have the same quality observations or communication channels. This is particularly true when one considers channel fading. Let us consider the power allocation among nodes such that the network consumes the minimum possible energy to achieve a desired performance. The optimization problem can be formulated as

$$\min_{g_k \geq 0, k=1,\cdots,n} \sum_{k=1}^{n} g_k^2 \text{ such that } MSE \leq \epsilon_1,$$

where ϵ_1 is the required MSE threshold at the fusion center. If we assume that the local observations are independent, the optimization problem can be rewritten as

$$\min_{g_k \geq 0, k=1,\cdots,n} \sum_{k=1}^{n} g_k^2 \text{ such that } \epsilon_1^{'} - \sum_{k=1}^{n} \frac{h_k^2 g_k^2}{\sigma_v^2 h_k^2 g_k^2 + \sigma_w^2} \leq 0, \qquad (17)$$

where we have defined $\epsilon_1^{'} = \frac{1}{\epsilon_1}$. The optimal power allocation strategy is stated in the following lemma assuming that the channel state information (CSI) is available at the distributed nodes.
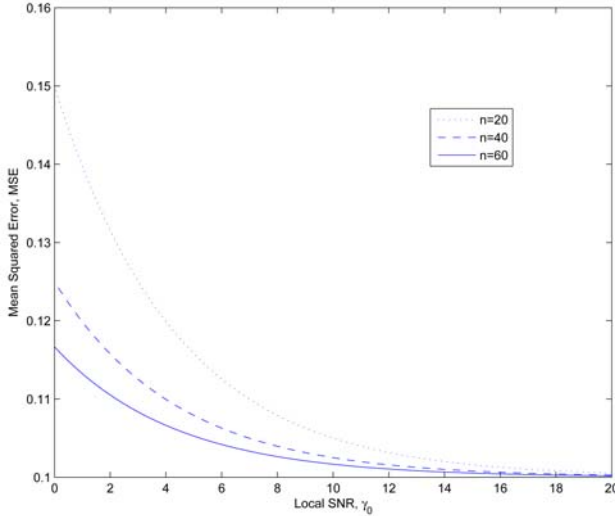

Fig. 4. Performance of mean squared error vs. local SNR, γ_0; The total network power is constant.

Lemma 5: [34] *When local observations are i.i.d., the optimal power allocation solution to (17) is given by*

$$g_k^{*2} = \begin{cases} \frac{\sigma_w^2}{h_k^2 \sigma_v^2} \left[\frac{h_k \sum_{j=1}^{K_1} \frac{1}{h_j}}{(K_1 - \epsilon_1^{'} \sigma_v^2)} - 1 \right] & ; if \ k < K_1 \ and \ n > \epsilon_1^{'} \sigma_v^2 \\ 0 & ; if \ k > K_1 \ and \ n > \epsilon_1^{'} \sigma_v^2 \\ infeasible & ; if \ n < \epsilon_1^{'} \sigma_v^2 \end{cases} \qquad (18)$$

where assuming, without loss of generality, $h_1 \geq h_2 \geq ... \geq h_n$, K_1 is found such that $s_1(K_1) < 1$ and $s_1(K_1 + 1) \geq 1$ for $1 \leq K_1 \leq n$ where $s_1(k) = \frac{(k - \epsilon_1^{'} \sigma_v^2)}{h_k \sum_{j=1}^{k} \frac{1}{h_j}}$ for $1 \leq k \leq n$.

Lemma 5 says that the optimal power at each node depends on its observation quality, channel quality and the required MSE threshold at the fusion center. Note that letting $\sqrt{\lambda_0} = \sigma_w \frac{\sum_{k=1}^{K_1} \frac{1}{h_k}}{K_1 - \epsilon_1^{'} \sigma_v^2}$, for $s_1(k) - 1 < 0$ and $n > \epsilon_1^{'} \sigma_v^2$, the optimal g_k^{*2} can be written as $g_k^{*2} = \frac{\sigma_w^2}{h_k^2 \sigma_v^2} \left(\frac{h_k \sqrt{\lambda_0}}{\sigma_w} - 1 \right)$. Hence, when CSI is available at distributed nodes, each node can determine its power using $\sqrt{\lambda_0}$ as a side information that is provided by the fusion center.
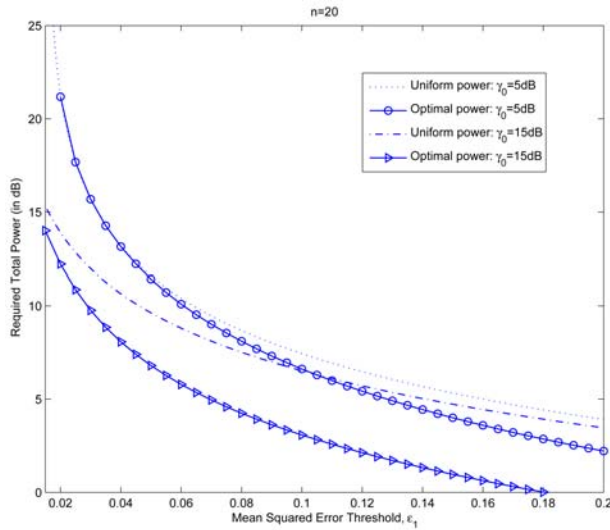
Fig. 5. Optimal power allocation scheme vs. uniform power allocation scheme: The required optimal power to achieve a given MSE of ϵ_1 as given in (18) is shown in the figure parameterized by local SNR γ_0 for n=20 is shown. The comparison of the required uniform power to achieve the same MSE threshold is illustrated.

Figure 5 shows the performance of the optimal power allocation scheme (18) in achieving a desired MSE performance at the fusion center. Figure 5 assumes that fading coefficients are drawn from a Rayleigh distribution with unity mean. It is seen that allocating power optimally as in (18) gives a significant power saving over the uniform power allocation only when either the local observation SNR is high or when the required MSE at the fusion center is not significantly low. This is not surprising since if local observation SNR is high node estimators are good enough on their own and thus perhaps collecting the local estimators from only those nodes with very good fading coefficients can save total power while also meeting the MSE requirement at the fusion center. Moreover, when the MSE required at the fusion center is not very demanding, we may meet it by only collecting local estimators of few nodes (and turning others off), so that the optimal power allocation, may lead to better power savings over the uniform power allocation scheme.

2) Quantized local processing: Recall that with the quantization scheme presented in Section 2-C2, an upper bound for the MSE at the fusion center is given by (9) when the communication between the sensors and the fusion center is noiseless. When discrete messages m_k's are transmitted over noisy communication channels, however, bit errors may occur in a resource constrained network with a finite power.

Let us assume that the discrete messages are transmitted over a noisy channel where bit errors occur due to imperfect communication. Let $\hat{m}_k$ and p_b^k be the decoded message at the fusion center corresponding to the transmitted message m_k from the k-th node and the associate bit error probability, respectively. To compute the resulting MSE of the estimator $\hat{\theta}$ at the fusion center based on the decoded messages $\{\hat{m}_1,..., \hat{m}_n\}$, the bit errors caused by the channel should be taken into account. For the quantization scheme presented in Section

2-C2, a complete analysis of the resulting MSE at the fusion center with noisy channels is given in [9]. According to [9], for i.i.d. local observations an upper bound for the MSE, when the messages are transmitted over a memoryless binary symmetric channel is given by the following lemma:

Lemma 6: [9] *If the bit error rates from node k is* p_b^k, *then the MSE achieved by the fusion center based on the decoded messages* $\{\hat{m}_1,...,\ \hat{m}_n\}$ *is upper bounded by*

$$MSE \leq (1 + p_0)^2 \left(\sum_{k=1}^n \frac{1}{\sigma_v^2 + \delta_k^2} \right)^{-1}, \tag{19}$$

where $\delta_k^2 = \frac{W^2}{(2^{L_k}-1)^2}$ *and* $p_0 > 0$ *satisfies the following condition:*

$$p_0 \geq \frac{8W}{\sigma_v} \sqrt{\frac{np_b^k}{3}} \text{ for all } k = 1, \cdots, n.$$

By comparing (19) with (10) it is observed that the achievable MSE with imperfect communication deviates by that with noiseless communication by a constant factor.

Let the communication channel between node k and the fusion center undergo path loss attenuation a_k proportional to d_k^α where d_k is the transmission distance from node k to fusion center and α is the path loss attenuation index. Assuming that node k sends L_k bits using quadrature amplitude modulation (QAM) with constellation size 2^{L_k}, at a bit error probability of p_b^k the transmission power spent by node k is $P_k = B_s E_k$, where B_s is the transmission symbol rate and E_k is the transmission energy per symbol, given by,

$$E_k = \varsigma_k a_k \left(\ln \frac{2}{p_b^k} \right) (2^{L_k} - 1), \tag{20}$$

with $\varsigma_k = 2N_f N_0 G_d$ where N_f is the receiver noise figure, N_0 is the single sided thermal noise spectral density and G_d is a system constant [9].

It can be easily seen from (20) that $L_k = \log \left(1 + \frac{P_k}{B_s \varsigma_k a_k \left(\ln \frac{2}{p_b^k} \right)} \right)$. Thus, to determine the optimal number of bits L_k to be allocated to node k in order to meet a desired MSE performance at the fusion center while minimizing the total network power, [9] solves the following optimization problem (assuming ς_k, B_s and p_b^k are the same for all nodes):

$$\min \|\mathbf{P}\|_2 \text{ such that } MSE \leq \epsilon_2, \tag{21}$$

where $\|\mathbf{P}\|_2 = \sqrt{\sum_{k=1}^n P_k}$ is the L^2-norm of the power vector $\mathbf{P} = [P_1,..., P_n]^T$, ϵ_2 is the desired MSE threshold at the fusion center and the MSE is as given by (19). The optimal number of bits L_k^* to quantize the observations at node k, that is given by the solution to (21), are characterized in the following lemma.

Lemma 7: [9] *The optimal number of bits used to quantize the observations at the k-th node found by solving (21) is*

$$
L_k^* = \begin{cases} \log\left(1 + \frac{W}{\sigma_v}\sqrt{\frac{\eta_0}{a_k} - 1}\right) & ; if\ k \leq K_2\ and\ n > \frac{\sigma_\gamma^2}{\epsilon_2} \\ 0 & ; if\ k \geq K_2 + 1\ and\ n > \frac{\sigma_\gamma^2}{\epsilon_2} \\ infeasible & ; if\ n < \frac{\sigma_\gamma^2}{\epsilon_2} \end{cases}
$$

where $\eta_0 = \left(\frac{K_2}{\sigma_v^2} - \frac{1}{\epsilon_2'}\right)^{-1} \sum_{k=1}^{K_2} \frac{a_k}{\sigma_v^2}$, $\epsilon_2' = \frac{\epsilon_2}{(1+p_0)^2}$ and assuming, without loss of generality, $a_1 \leq a_2 \leq ... \leq a_n$, K_2 is found such that $s_2(K_2) < 1$ and $s_2(K_2 + 1) \geq 1$ for $1 \leq K_2 \leq n$ where $s_2(k) = a_k\left(\frac{k}{\sigma_v^2} - \frac{1}{\epsilon_2'}\right)\left(\sum_{j=1}^{k} \frac{a_k}{\sigma_v^2}\right)^{-1}$. Then the optimal transmission power at the k-th node is given by

$$
P_k^* = \varsigma_k B_s \left(\ln \frac{2}{p_b}\right) \frac{Wh_k}{\sigma_v} \sqrt{\left(\frac{\eta_0}{a_k} - 1\right)^+},
$$

where $(x)^+$ equals to zero if $x < 0$ and equals to x otherwise.

Note that again the optimal power at node k is determined by the observation quality, channel quality and the required MSE threshold as was the case with AF local processing we saw in lemma 5. Figure 6 shows the number of sensors that are active in the network to achieve a desired MSE threshold at the fusion center. In Fig. 6, the network size $n = 1000$ and $\alpha = 2$. The distance from node k to fusion center, d_k, is drawn from a uniform distribution on [1, 2]. It is observed that when the required MSE threshold increases the number of active sensors decreases greatly. That is, the network discards the observations at nodes with poor observation and channel quality. This is similar to what we observed in Fig. 5 earlier.
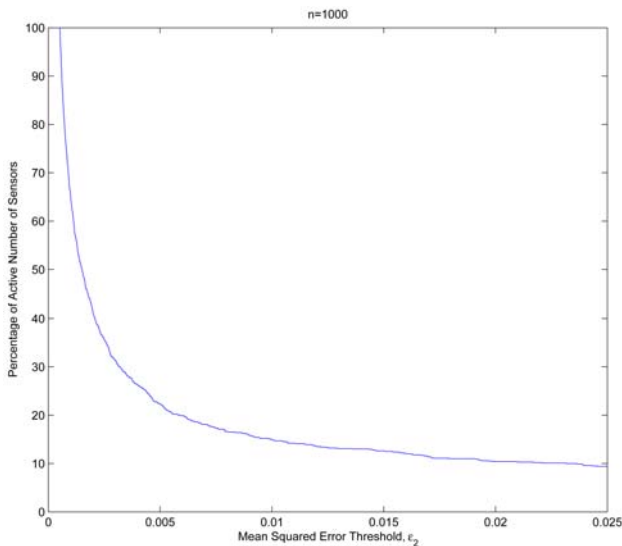

Fig. 6. Number of active sensors in the network according to the optimal power allocation scheme given by lemma 7. The number of total sensors in the network is $n = 1000$ and $\alpha = 2$

Figure 7 shows the energy saving due to the optimal power allocation scheme given in lemma 7 compared to the uniform power allocation scheme. Clearly Fig. 7 shows that significant energy savings are possible by optimally selecting number of bits, especially at moderate levels of desired MSE at the fusion center.
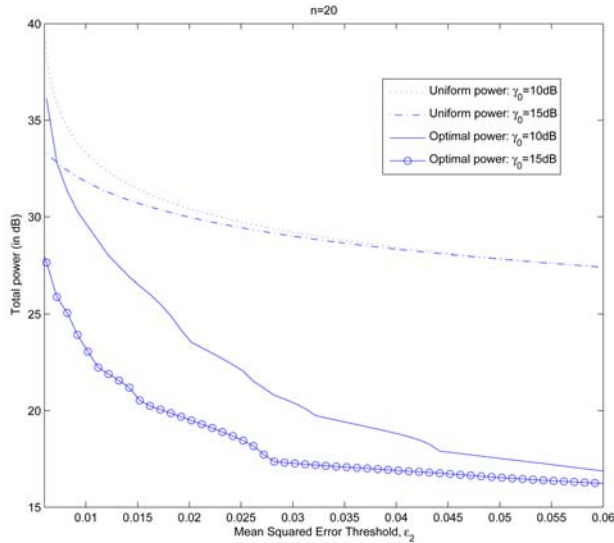

Fig. 7. Performance of optimal power allocation scheme given in lemma 7 vs. uniform power allocation scheme: network size n = 20

B. Communication over multiple access channels

One of the disadvantages of using orthogonal channels to transmit local decisions is the large bandwidth consumption as the number of distributed nodes n increases. An alternative is to allow multiple sensor nodes to share a common channel. Such multiple access communication (MAC) in bandwidth constrained wireless sensor networks has been investigated in, among others, [10], [11], [17], [24], [25], [34], [36], [44]. For example, in [24], [25], [44] the authors proposed a type based multiple-access communication in which sensors transmit according to the type of their observation in a shared channel where the *type* is as defined in [45]. An analysis of both orthogonal and MAC channels for distributed detection in a sensor network was presented in [44]. MAC with correlated observations was considered in [34] and [46]. The use of CDMA signaling in distributed detection of deterministic and Gaussian signals under strict power constraints was presented in [10] and [11], respectively. When all sensor nodes communicate with the fusion center coherently, with amplify-and-forward local processing the estimator performance can be improved compared to that of orthogonal communication due to the coherent beam-forming gain [47], [46]. Performance of MAC communication with asynchronous transmissions was discussed in [48].

In the following we consider the form and performance of the final estimator at the fusion center when communications from distributed nodes to the fusion center is over noisy multiple-access channels. Assuming perfect synchronization among sensor transmissions, the received signal at the fusion center over a MAC can be written as

$$r = \sum_{k=1}^{n} h_k y_k + w,$$

where w is the receiver noise with zero mean and variance of σ_w^2 and h_k is the channel fading coefficient from node k to the fusion center, as defined earlier. For the AF local processing, substituting $y_k = g_k z_k$, the resulting received signal is given by

$$r = \sum_{k=1}^{n} h_k g_k z_k + w. \tag{22}$$

Fusion center computes the final estimator based on the received coherent signal r. The resulting BLUE estimator and its performance is given by the following lemma.

Lemma 8: [34] *The BLUE estimator and the resulting MSE based on the received signal (22) can be shown to be*

$$\hat{\theta}(r) = \frac{r}{\sum_{k=1}^{n} h_k g_k},$$

and

$$MSE = \frac{\mathbf{e^T G H_c \Sigma_v H_c G e} + \sigma_w^2}{(\mathbf{e^T H_c G e})^2}.$$

With i.i.d. local observations the MSE simplifies to

$$MSE = \frac{\sigma_v^2 \sum_{k=1}^{n} h_k^2 g_k^2 + \sigma_w^2}{(\sum_{k=1}^{n} h_k g_k)^2}.$$

The MSE performance of the BLUE estimator under both orthogonal and multiple-access channels, with i.i.d. observations, is depicted in Fig. 8 as a function of total network power. Figure 8 assumes equal node powers and unity channel gains. Moreover, MAC communication is assumed to be perfectly synchronized among nodes. As seen from Fig. 8, when total network power is small, the MAC communication leads to a better MSE performance compared to that with orthogonal communication. But as total network power increases both schemes converge to the same performance level. This is because when the network can afford a large transmission power, irrespective of the communication scheme the overall estimator performance is only limited by the local observation quality and the effects of additive/multiplicative channel noise is mitigated by the large gain in the transmission. However, when a practical sensor network is power-constrained the MAC communication may be able to provide a much better performance over that of the orthogonal transmissions when nodes are perfectly synchronized.

Figure 8 assumes equal transmission powers at all nodes. However, when the fusion center needs to achieve only a target estimator quality, say an MSE of ϵ_3, one can consider non-uniform power allocations such that,

$$\min_{g_k \geq 0, k=1, \cdots, n} \sum_{k=1}^{n} g_k^2 \text{ such that } MSE \leq \epsilon_3, \tag{23}$$

where MSE is as given in lemma 8. When the observations are i.i.d., a tractable analytical solution for the above optimization problem was given in [34] that is stated in the following lemma.
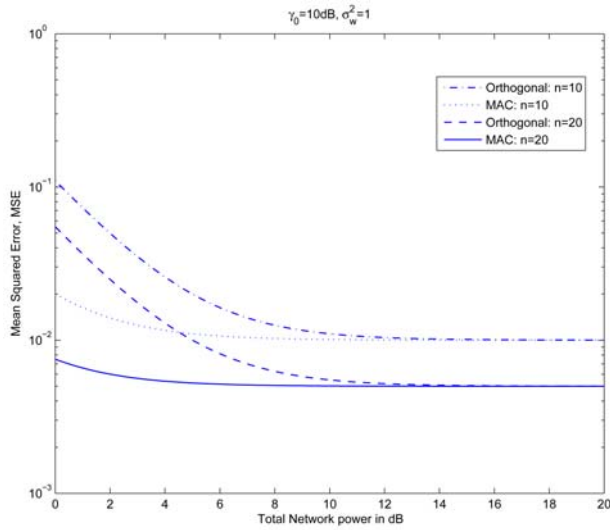
Fig. 8. Mean squared error performance vs. total network power for different network sizes
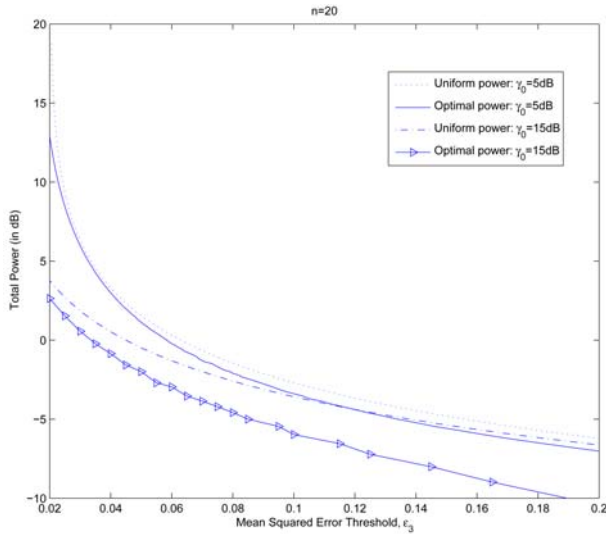

Fig. 9. Performance of optimal power allocation scheme vs. uniform power allocation scheme

Lemma 9: With i.i.d. local observations, the optimal power at node k, g_k^{*2} that solves the optimization problem in (23) with the MSE as given in lemma 8 is $g_k^{*2} = \dfrac{\mu^2}{4} \dfrac{h_k^2}{(1+\eta_0 h_k^2)^2}$ for k = 1, 2,..., n, where η_0 and μ can be found numerically by solving the equations $\sum_{k=1}^{n} \dfrac{\eta_0 h_k^2}{(1+\eta_0 h_k^2)} = \dfrac{1}{\epsilon_3'}$ and $\mu = 2\dfrac{\sigma_w}{\sigma_v} \left(\dfrac{1}{\eta_0^2 \epsilon_3'} - \sum_{k=1}^{n} \dfrac{h_k^4}{(1+\eta_0 h_k^2)^2} \right)^{-\frac{1}{2}}$ where $\epsilon_3' = \dfrac{\epsilon_3}{\sigma_v^2}$.

It is observed that η_0 has a feasible solution only when $n > \frac{\sigma_v^2}{\epsilon_3}$ [34]. The total power spent by the network with the above optimal power allocation scheme is given by $P_{total} = \sum_{k=1}^{n} g_k^{*2} = \frac{\sigma_u^2}{\sigma_v^2} \eta_0$.

Figure 9 shows the performance of the optimal power allocation scheme compared to that of uniform power allocation scheme for a network size of n = 20. Again, the optimal power scheduling scheme has a significant performance gain over the uniform power allocation scheme especially when local SNR is high or the required MSE threshold at the fusion center is moderate, similar to what was observed in Section 3-A in the case of orthogonal communication.

C. Effects of synchronization errors on MAC

To achieve coherent gain with MAC transmissions, it is important that the sensor transmissions are synchronized. For this discussion on node synchronization, we assume, i.i.d observations and AF local processing. Analysis would remain essentially the same for other network models as well.

To achieve synchronization in node transmissions, one may assume that there is a master-node (that can be taken as the fusion center itself, for simplicity) that broadcasts the carrier and timing signals to the distributed nodes [47]. Suppose that the k-th node is located at a distance of $d_k + \delta_k$ from the fusion center, for k = 1, 2,..., n, where d_k and δ_k are the nominal distance and the sensor placement error of the k-th node, respectively. The fusion center broadcasts a carrier signal $\cos(2\pi f_0 t)$ where f_0 is the carrier frequency. The received carrier signal at the k-th node is a noisy version of $\cos(2\pi f_0 t + \psi_k + \psi_{ek})$ where $\psi_k = \frac{2\pi f_0 d_k}{c}$ and $\psi_{ek} = \frac{2\pi f_0 \delta_k}{c}$. Each node employs a Phase Locked Loop (PLL) to lock onto the carrier. If each node precompensates for the difference in their nominal distances d_k, by transmitting its locally processed and modulated observation with a proper delay and phase shift ψ_k, then the received signal at the fusion center is corrupted only by the phase shift due to the sensor placement error δ_k. Considering only the phase shift due to this sensor placement error, the received signal at the fusion center is given by $r = \sum_{k=1}^{n} h_k g_k z_k \cos(\psi_{ek}) + w$, assuming AF local processing at sensor nodes. In the following lemma we assume that the placement error δ_k is Gaussian with zero mean and variance σ_δ^2.

Lemma 10: [34] *Assuming that σ_δ^2 << ι_0 where $\iota_0 = \frac{c}{f_0}$, so that phase error $\psi_{ek} \sim \mathcal{N}(0, \sigma_\psi^2)$ where σ_ψ^2 is small, the BLUE estimator at the fusion center when local observations are i.i.d. is $\hat{\theta}(r) = r \left(e^{-\frac{\sigma_\psi^2}{2}} \sum_{k=1}^{n} h_k g_k \right)^{-1}$. The resulting MSE with the phase synchronization errors is*

$$MSE = \frac{\sigma_v^2 \sum_{k=1}^{n} h_k^2 g_k^2 + e^{\sigma_\psi^2} \sigma_w^2}{\left(\sum_{k=1}^{n} h_k g_k \right)^2}. \tag{24}$$

Figure 10 shows the MSE performance (24) of a sensor system in the presence of phase synchronization errors. It can be seen that the performance is robust against synchronization errors as long as the variance of the phase error σ_ψ^2 is sufficiently small.

4. Sequential communication

In Section 3 it was assumed that the final decision on the PoI is made at a central fusion center, and all nodes were to send their locally processed data to this fusion center. In a distributed network, however, it might be desirable in some applications that the final decision be made fully distributively without depending on a central node. To achieve this goal, nodes may communicate with each other to reach a final decision albeit at the cost of inter-sensor communications. One such architecture of distributed estimation is to communicate with nodes sequentially until the desired performance is reached.
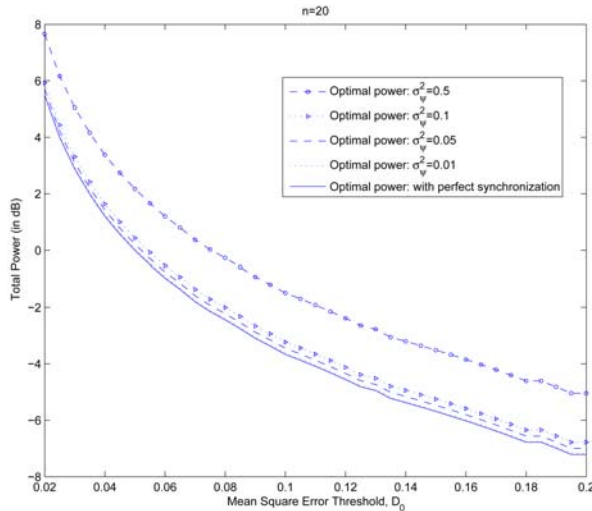

Fig. 10. Performance of optimal power allocation scheme with synchronization errors: $n = 20$, $\eta_0 = 10 dB$

The basic problem of sequential detection for statistical hypotheses was first formulated by Wald in [49] who derived the sequential probability ratio test (SPRT). Analysis of SPRT and its comparison with fixed sample-size test for centralized detection of a constant signal was given in [50]. The decentralized version of the binary sequential detection problem was addressed by [41] and a more general formulation of the distributed sequential detection problem was presented in [40]. Decentralized sequential detection problem with multiple hypotheses was considered in [51] and [52].

Distributed estimation with sequential communication in wireless sensor networks has been addressed by several authors in recent years. For example, in [37]–[39], [53] information driven approaches for distributed sequential estimation of a source location have been investigated. In this architecture only one node communicates with another at a given time and the final decision can be made at any node once a required performance level is reached. However, it is worth mentioning that most of these are concerned with random parameters. To be consistent with our analysis in section 3, we, on the other hand, will consider non-random parameter estimation with sequential estimation, as considered, for example, in [42]. Thus, in the following, we consider the distributed sequential estimation of a non-random parameter where each node makes a local decision based on its own observation and the decision from the previous node.

A. Distributed sequential estimation

Let us assume the same linear observation model (2) as in sections 2 and 3:

$$z_k = \theta + v_k, \ \text{for} \ k = 1, \cdots, n. \tag{25}$$

Let $\hat{\theta}_{k-1}$ be the local estimator at the $(k - 1)$-th node. The $(k - 1)$-th node sends its estimator to the k-th node over a noisy channel. Then the effective observation vector at node k is [42]

$$\mathbf{z}_k = \begin{bmatrix} z_k \\ y_k \end{bmatrix} = \begin{bmatrix} \theta + v_k \\ \hat{\theta}_{k-1} + w_k \end{bmatrix} \ \text{for} \ k = 2, \cdots, n, \tag{26}$$

with $\mathbf{z}_1 = z_1$ where w_k is the channel noise at inter-sensor communication link from node $(k - 1)$ to node k and, as before, v_k is the observation noise at the k-th node. Both $\mathbf{v}$ and $\mathbf{w}$ are assumed to be zero mean with covariance matrices $\mathbf{\Sigma_v}$ and $\mathbf{\Sigma_w}$, respectively. Throughout this section we assume that the channel noise $\{w_k\}_{k=1}^n$ is independent with the covariance matrix $\mathbf{\Sigma_w} = \mathrm{diag}(\sigma_{w_1}^2, \cdots, \sigma_{w_n}^2)$. Moreover the observation noise v_k and the channel noise w_k are assumed to be independent of each other. Given the effective observation vector $\mathbf{z}_k$ the node k computes the BLUE estimator of parameter θ. In the following we consider the cases of independent and correlated observations, separately.

1) Independent observations: When local observations are independent, $\mathbf{\Sigma_v} = \mathrm{diag}(\sigma_1^2, \cdots, \sigma_n^2)$. The following lemma from [42] summarizes the BLUE estimator and its performance at the k-th node for independent observations.

Lemma 11: [42] *When observation noise is independent the BLUE estimator at the k-th node is given by*

$$\hat{\theta}_k(\mathbf{z}_k) = \frac{P_{k-1} + \sigma_{w_k}^2}{G(k)} z_k + \frac{\sigma_k^2}{G(k)} y_k, \ \text{for} \ k = 2, \cdots, n, \tag{27}$$

where $G(k) = P_{k-1} + \sigma_{w_k}^2 + \sigma_k^2$ *and* P_k *is the MSE at k-th node which can be shown as*

$$P_k = P_{k-1} - \frac{(P_{k-1} + \sigma_{w_k}^2)P_{k-1} - \sigma_{w_k}^2\sigma_k^2}{G(k)}, \ \text{for} \ k = 2, \cdots, n. \tag{28}$$

For k=1, we have

$$\hat{\theta}_1(\mathbf{z}_1) = z_1, \tag{29}$$

with corresponding MSE of

$$P_1 = \sigma_1^2. \tag{30}$$

It is seen from lemma 11 that the BLUE estimator at node k can be determined by observation vector at node k along with the variance of the estimator P_{k-1} at the previous node. It is also interesting to see that as the inter-sensor communication link noise vanishes,

$$\lim_{\sigma_{w_k}^2 \to 0} P_k \approx \frac{P_{k-1}}{1 + \frac{P_{k-1}}{\sigma_k^2}}, \ \text{for} \ k = 2, \cdots, n. \tag{31}$$
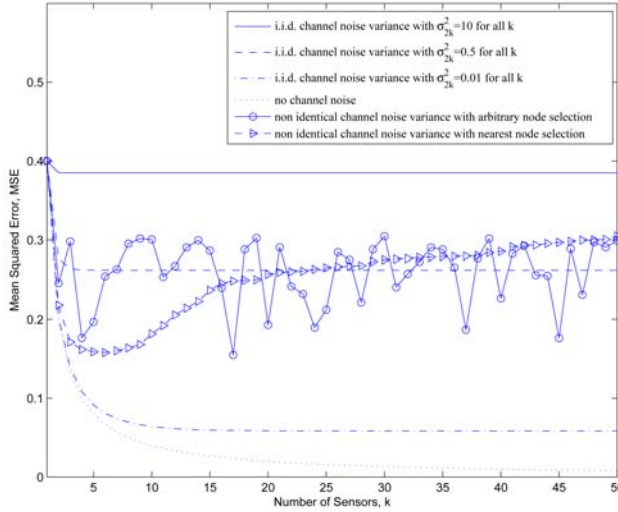
Fig. 11. Mean squared error at the k-th node: The observations are assumed to be i.i.d. with $\sigma_k^2 = \sigma_0^2 = 0.4$ for all k. Three instances of channel noise variance $\sigma_{w_k}^2$ are considered where case 1: $\sigma_{w_k}^2$ is i.i.d. with values of 0.01, 0.5 and 10; case 2: $\sigma_{w_k}^2$'s are drawn from a uniform distribution in [0, 1]; case 3: no channel noise, i.e. $\sigma_{w_k}^2 = 0$ for all k

In other words, when the inter-sensor communication links are very good we have $P_k < P_{k-1}$ for $k = 2,..., n$ irrespective of how the next node is selected. That implies that, sending the $(k -1)$-th node's decision to the k-th node always improves the MSE performance. On the other hand, if the quality of inter-sensor communication links is poor, we get

$$\lim_{\sigma_{w_k}^2 \to \infty} P_k \approx \sigma_k^2, \text{ for } k = 2, \cdots, n. \tag{32}$$

(32) implies that when the quality of inter-sensor communication links is poor, the MSE at the k-th node is not affected by the $(k - 1)$-th node's decision. Therefore, in that case sequential communication will not improve the MSE performance. In fact, we can see that the above sequential estimation process gives improved performance only when the following condition is satisfied for the observation quality:

Lemma 12: [42] *If the current node's MSE is P_k, then $P_k \leq P_{k-1}$ if and only if the observation quality at the k-th node satisfies the following condition:*

$$\sigma_k^2 \leq \left(1 + \frac{P_{k-1}}{\sigma_{w_k}^2}\right) P_{k-1}. \tag{33}$$

For i.i.d. observations and channel noise (such that $\mathbf{\Sigma_v} = \sigma_0^2 \mathbf{I}$ and $\mathbf{\Sigma_w} = \sigma_w^2 \mathbf{I}$) it can be shown that $P_k \leq P_{k-1}$ for all $k \geq 1$ [42]. In this case, the MSE at the k-th node also simplifies to

$$P_k = \frac{\sigma_0^2 (P_{k-1} + \sigma_w^2)}{P_{k-1} + \sigma_w^2 + \sigma_0^2}. \tag{34}$$

As $k \rightarrow \infty$, the asymptotic variance converges to $P_k = P_{k-1} = P_\infty$. Then from (34), it can be shown that [42]

$$P_\infty = \frac{\sigma_w^2}{2} \left(\sqrt{1 + \frac{4\sigma_0^2}{\sigma_w^2}} - 1 \right),$$ (35)

and we have $P_\infty \leq \sigma_0^2$. It is also of interest to see that

$$if \ \sigma_w^2 \gg \sigma_0^2 : \qquad P_\infty \ \approx \sigma_0^2$$ (36)

$$if \ \sigma_w^2 \ll \sigma_0^2 : \qquad P_\infty \ \approx \sigma_w \sigma_0 \left(1 - \frac{\sigma_w}{2\sigma_0} \right).$$ (37)

Figure 11 shows the performance of the distributed sequential estimator for i.i.d. observation noise (i.e. $\sigma_k^2 = \sigma_0^2$ for $k = 1,..., n$).

In Fig. 11 we have let $\sigma_0^2 = 0.4$. If the channel noise variance is also i.i.d. so that $\sigma_{w_k}^2 = \sigma_w^2$, [42] has shown that $P_k \leq P_{k-1}$ for all $k \geq 1$ which is illustrated in Fig. 11 as well. However, it is seen that in this case, when the channel noise variance σ_w^2 increases the MSE is limited by the observation noise variance σ_0^2 as predicted by (32). When channel noise variances are non-identical, Fig. 11 depicts the MSE performance at node k with two different schemes for the next node selection. In one scheme, next node is selected randomly and in the other scheme the next node is chosen to be the node with the minimum distance to the current node (note that here we are assuming that $\sigma_{w_k}^2$'s are in ascending order with k). As can be seen from Fig. 11, with random node selection, whenever the condition (33) in lemma 12 is satisfied, we have $P_k < P_{k-1}$. On the other hand, when the next node is selected to be the one at the minimum distance, it is seen from Fig. 11 that after a certain number of nodes, the MSE starts to monotonically increase. Therefore, with this scheme it is important to identify the node at which the MSE is minimum, and terminate the sequential updating process at the particular node.

Fig. 11 only shows the MSE performance when the observations are i.i.d.. However, if the observations are not identical, it might be of interest to perform an information driven distributed sequential estimation process, in which the sequence of nodes are selected to capture the highest information gain as well as with the lowest communication cost.

2) Correlated observations: When the observations are correlated, the covariance matrix of the effective received signal $\mathbf{z}_k$ (26) at k-th node can be written as,

$$\Sigma_{\mathbf{z}_k} = \begin{bmatrix} \sigma_k^2 & r_{k,k-1} \\ r_{k,k-1} & \sigma_{w_k}^2 \end{bmatrix}, \text{ for } m = 1, \cdots, k,$$ (38)

where $r_{k,k-m} \triangleq \mathbb{E}\{v_k \tilde{\theta}_{k-m}\}$ with $\tilde{\theta}_k = \hat{\theta}(\mathbf{z}_k) - \theta$. With the covariance matrix in (38), the BLUE estimator and MSE are summarized in the following lemma.

Lemma 13: [42] *When local observations are correlated with the covariance matrix in (38), the BLUE estimator at k-th node is given as*

$$\hat{\theta}_k(\mathbf{z}_k) = \frac{P_{k-1} + \sigma_{w_k}^2 - r_{k,k-1}}{G'(k)} z_k + \frac{\sigma_k^2 - r_{k,k-1}}{G'(k)} w_k,$$ (39)

where $G'(k) = P_{k-1} + \sigma_k^2 + \sigma_{w_k}^2 - 2r_{k,k-1}$ and the corresponding MSE P_k, at k-th node is

$$P_k = \frac{\sigma_k^2(P_{k-1} + \sigma_{w_k}^2) - r_{k,k-1}^2}{G'(k)}, \tag{40}$$

with $P_1 = \sigma_1^2$ and $r_{k,k-1}$ can be computed at node k using following recursion formula:

$$r_{k,k-1} = \frac{\sigma_{k-1}^2 - r_{k-1,k-2}}{G'(k-1)}r_{k,k-2} + \frac{P_{k-2} + \sigma_{w_k}^2 - r_{k-1,k-2}}{G'(k-1)}\mathbb{E}\{v_k v_{k-1}\}. \tag{41}$$

In general, computing $r_{k,k-1}$ from (41) needs a recursion that spans over all $r_{j,j-m}$ for $m = 1,..., j$ for each $j = k, k - 1,..., 1$. In the special case where v_k is wide sense stationary with identical variance σ_0^2 such that $\mathbb{E}\{v_k v_{k-m}\} = \rho^m \sigma_0^2$, [42] has shown that the recursion in (41) can be computed based at node k only on that is computed at node $k - 1$. Then the distributed sequential algorithm can be summarized as [42],
1) Initialization

$$\hat{\theta}_1(\mathbf{z}_1) = z_1$$
$$P_1 = \sigma_0^2 \text{ and } r_{2,1} = \rho\sigma_0^2$$

2) For $k = 2, 3,...$

$$G'(k) = P_{k-1} + \sigma_0^2 + \sigma_{w_k}^2 - 2r_{k,k-1}$$
$$\hat{\theta}_k(\mathbf{z}_k) = \frac{P_{k-1} + \sigma_{w_k}^2 - r_{k,k-1}}{G'(k)}z_k + \frac{\sigma_0^2 - r_{k,k-1}}{G'(k)}w_k$$
$$F(k) = \frac{\sigma_{w_k}^2 + P_{k-1} - \sigma_0^2}{G'(k)}$$
$$P_k = \frac{\sigma_0^2(P_{k-1} + \sigma_{w_k}^2) - r_{k,k-1}^2}{G'(k)}$$
$$r_{k,k-1} = \frac{\rho}{2}\left[(1 + F(k))\sigma_0^2 + (1 - F(k))r_{k,k-1}\right]$$

The above algorithm implies that the sequential estimation process for correlated observations at node k can be performed with information received from node $k - 1$, ($\hat{\theta}_{k-1}(\mathbf{z}_{k-1})$, P_{k-1} and $r_{k,k-1}$) and its own information.

Figure 12 shows the MSE performance at node k with the number of nodes assuming i.i.d. channel noise with $\sigma_w^2 = 1$. It is seen that, when the observation noise is highly correlated, the sequential estimation process does not give significant performance which intuitively makes sense.

5. Distributed collaborative data fusion with consensus

The specific topology of the network becomes an issue if all the participant nodes must collaborate to improve on their individual estimates, without a central fusion center that collects and processes all measurements as assumed in Section III. Still it is possible to
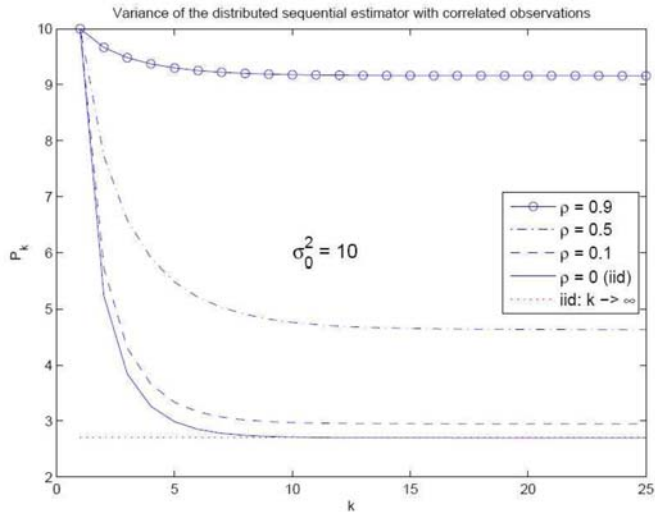
Fig. 12. MSE performance at *k*-th node with correlated observations

obtain distributively a function of the observation vector such as the BLUE estimator, as we will detail later. In this section we assume that information is exchanged only locally with neighbors, and can reach distant nodes through an iterative process.

For a proper characterization of the collaboration in such a network, it is necessary to make use of analytical tools which describe the topology of a sensor net. Thus we model the sensor network as a graph $G = (V,E)$, with nodes (sensors) $\nu_k \in V$ and edges $e_{kj} \in E$ if there is a path from node ν_k to node ν_j. Note that a path exists in the network if transmissions from node ν_k reaches node ν_j. The elements of the adjacency matrix $\mathcal{A}$ are defined as $[\mathcal{A}]_{kj} = 1$ if $e_{jk} \in E$, otherwise $[\mathcal{A}]_{kj} = 0$. If there is a sequence of edges to go from any node k to any other node j the graph is said to be connected. The degree matrix $\mathbf{D}$ of graph G is a diagonal matrix such that $[\mathbf{D}]_{kk}$ is equal to the number of connections entering node k. With that, the Laplacian matrix $\mathbf{L}$ is defined as $\mathbf{L} = \mathbf{D} - \mathcal{A}$. Specifically, the elements $[\mathbf{L}]_{kj}$ of the Laplacian matrix $\mathbf{L}$ are defined as

$$[\mathbf{L}]_{kj} = \begin{cases} -1, & if \ e_{jk} \in E \\ [\mathbf{D}]_{kk}, & if \ k = j. \end{cases} \tag{42}$$

It turns out that the eigenvalues $\{\lambda_k\}_{k=1}^{n}$ of $\mathbf{L}$ contain important information about the topology of the graph G. In fact, if they are ordered as $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$, we always have that $\lambda_1 = 0$ and $\lambda_2 > 0$ for a connected graph (that is, a graph in which there exists a path connecting any two nodes). This second eigenvalue λ_2 is known as the algebraic connectivity of the graph, and its value plays a major role in the speed at which information can be diffused through the network [55]. The corresponding eigenvectors are denoted by $\mathbf{u}_k$, with $\mathbf{u}_1 = \mathbf{e}$ where, as before, $\mathbf{e}$ is the n-vector of all ones. As an illustration, the four-node network in Fig. 13 has the following associated Laplacian matrix:

$$\mathbf{L} = \begin{pmatrix} 1 & 0 & 0 & -1 \\ -1 & 1 & 0 & 0 \\ -1 & -1 & 2 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix}. \tag{43}$$

The connection between nodes k and j is symmetric if whenever node k sends data to node j it can also get data from node j. If all the connections were symmetric, the corresponding Laplacian matrix of the graph in Fig. 13 would look like as follows:

$$\mathbf{L} = \begin{pmatrix} 3 & -1 & -1 & -1 \\ -1 & 2 & -1 & 0 \\ -1 & -1 & 3 & -1 \\ -1 & 0 & -1 & 2 \end{pmatrix}. \tag{44}$$

Note that in this case the Laplacian matrix is also symmetric. In the first case the eigenvalues of $\mathbf{L}$ are given by $\{0, 1.5 \pm 0.866j, 2\}$, whereas in the latter case, they are $\{0, 2, 4, 4\}$. Clearly the network is more strongly connected in the second case, and this fact is reflected in the magnitude of the second eigenvalue or, in other words, in the algebraic connectivity.
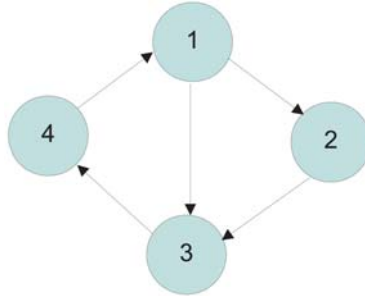

Fig. 13. Four node graph

A. Distributed estimation of static parameters

We consider again the same observation model (2) at distributed nodes. However, now all sensors attempt to share information with everyone else in order to update each of their local estimators. Moreover, all sensors wish to agree on a common estimate for θ, in what is known as distributed consensus estimation. This cannot be achieved in one shot, however, due to the constraints imposed by the network topology, since sensors can only access information from other nodes corresponding to the non-zero entries in $\mathbf{L}$ after one exchange, in $\mathbf{L}^2$ after two exchanges, and so on. Nodes will have to iteratively keep updating and exchanging their local estimators to reach a consensus. Let $\mathbf{z}(i)$ denote the information vector of all nodes after the i-th information exchange with $\mathbf{z}(0) = \mathbf{z}$ where $\mathbf{z}$ is as given in (2). If, for simplicity, we first consider ideal noiseless links, the following lemma shows how to lead all sensors to the same estimate.

Lemma 14: [55] *The recursion*

$$\mathbf{z}(i+1) = (\mathbf{I} - \gamma\mathbf{L})\mathbf{z}(i), \quad \text{for } i = 0, 1, 2, \dots \tag{45}$$

converges to the vector $(\mathbf{u}^T\mathbf{z})\mathbf{e}$, *where* $\mathbf{u} = [u_1, u_2, ..., u_n]^T$ *denotes the left eigenvector of* $\mathbf{L}$ *associated with the eigenvalue 0 normalized as* $\sum_{k=1}^{n} u_k = 1$, *provided that* $0 < \gamma < 1/\max_k[\mathbf{D}]_{kk}$.

According to the recursion (45), at each iteration, the current estimates are exchanged with nodes defined by the active links of the Laplacian matrix $\mathbf{L}$, and updated after weighting the neighbors contributions by γ. Thus, an agreement on the value $\mathbf{u}^T\mathbf{z}$ can be reached after convergence[1]. The convergence speed is a function of both γ and the algebraic connectivity of the graph. Hence sensor local estimates will converge faster in more densely connected networks.

In many situations, however, it might be of interest to ensure that the network converges to an arbitrary final consensus value of $\mathbf{c}^T\mathbf{z}$. We can indeed modify (45) to achieve this, so that the final asymptotic consensus estimator is given by $(\mathbf{c}^T\mathbf{z})\mathbf{e}$.

Lemma 15: The recursion

$$\mathbf{z}(i+1) = (\mathbf{I} - \gamma\mathbf{C}^{-1}\mathbf{U}\mathbf{L})\mathbf{z}(i), \quad i = 0, 1, 2, \dots \tag{46}$$

converges to the vector $(\mathbf{c}^T\mathbf{z})\mathbf{e}$, *where* $\mathbf{C}$ *and* $\mathbf{U}$ *are diagonal matrices such that* $\mathbf{C} = \text{diag}(\mathbf{c})$ *and* $\mathbf{U} = \text{diag}(\mathbf{u})$.

Note that the local exchange of values (peer-to-peer) serves to improve the individual estimates at nodes even not directly connected with each other, as the information is diffused through the network at each iteration.

B. Robust consensus schemes

Clearly, the assumption of ideal noiseless links is not valid in practice: exchanges contain additive noise, and it turns out that the recursion (45) does not converge to a consensus [56], unless some provision is taken[2]. One possible remedy is to use a decreasing sequence of positive steps $\gamma(i)$ as suggested in [57]:

$$\mathbf{z}(i+1) = (\mathbf{I} - \gamma(i)\mathbf{C}^{-1}\mathbf{U}\mathbf{L})\mathbf{z}(i) + \text{Diag}\{(\mathbf{I} - \gamma(i)\mathbf{C}^{-1}\mathbf{U}\mathbf{L})\mathbf{W}(i)\}, \quad i = 0, 1, 2, \dots \tag{47}$$

where $\text{Diag}(A)$ denotes the vector formed by diagonal elements of A and $\text{Diag}\{(\mathbf{I} - \gamma(i)\mathbf{C}^{-1}\mathbf{U}\mathbf{L})\mathbf{W}(i)\}$ in (47) accounts for the channel noise leaking into the sensors. Note that the $n \times n$ matrix $\mathbf{W}(i)$ contains the links noise values $w_{kj}(i)$, as illustrated in Fig. 14 (if two different nodes are not connected, the corresponding value in $\mathbf{W}(i)$ is irrelevant).

Lemma 16: [57] Asymptotic consensus is achieved in (47) as long as the positive sequence $\{\gamma(i)\}$ is such that $\gamma(i) < 1/\max_k[D]_{kk}$ for all i and satisfies the following conditions:

$$\sum_{i=0}^{\infty} \gamma(i) = \infty \text{ and } \sum_{i=0}^{\infty} \gamma^2(i) < \infty. \tag{48}$$

---

[1] If the network is symmetric, that is, $\mathbf{L} = \mathbf{L}^t$, then $\mathbf{u} = \mathbf{e}/n$, and the consensus value is the average of the initial observations.

[2] If the noise is correlated with the exchanged values, as it is the case for quantization noise, the strategies to follow can be different from those presented here, which are more general.

In the noiseless case the asympotic consensus value is that corresponding to the constant step of previous section, which turns out to be independent of the convergence speed. This is no longer the case in the presence of noise. In fact, the stepsize sequence $\{\gamma(i)\}$ should be designed to speed up convergence while minimizing the impact of channel noise, which will be proportional to the sequence-energy $\sum_{i=0}^{\infty} \gamma^2(i)$[58]. For instance, at each step each sensor could compute the BLUE estimator by combining what it receives from its neighbors with its own value, taking into account the exchange noise statistics in the analysis. Invariably, as the exchanges progress the estimates at different nodes become correlated even for initially independent observations. Hence the nodes should be able to estimate the degree of correlation among them for proper combining. On the other hand, it is also possible to use filtering to fight noise (see e.g. [59]) although this causes strict consensus not be achievable.
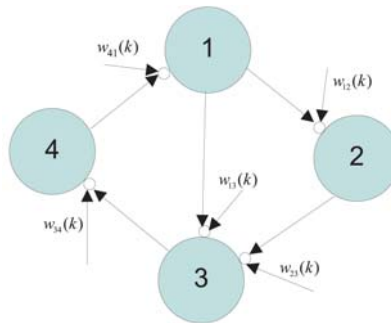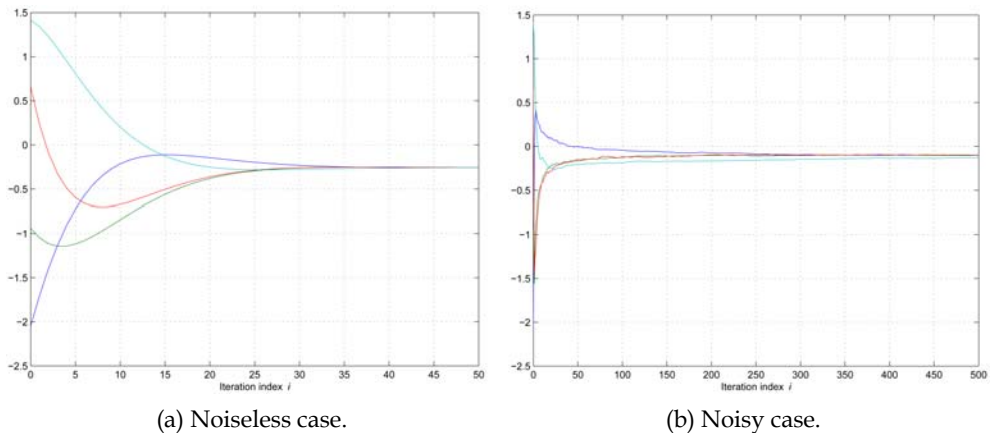

Fig. 14. Four node graph with noisy links.

(a) Noiseless case.   (b) Noisy case.

Fig. 15. Convergence curves towards consensus: (a) noiseless links, $\gamma = 0.1$, (b) noisy links, $\gamma(i) = 0.5 = i$, $i = 1, 2,..., \sigma_w^2 = 0.1$.

Figure 15 shows the evolution of the values in the four node network shown in Figs. 13 and 14, respectively. Clearly, consensus is achieved even in the noisy case, although the asymptotic estimate differs from the ideal case due to the noise. Observe the different horizontal axis scaling used in the two figures, due to the slower convergence in the noisy case.

C. Cooperative tracking: distributed Kalman filters

Dynamic parameter estimation, including target tracking, using wireless sensor networks has been considered by many authors in the literature. Common methods used in dynamic parameter estimation (or target tracking) include Kalman filtering, Bayesian estimation methods, particle filtering and their variants. Kalman filtering for distributed parameter estimation or tracking has been studied by many authors, [60]–[65]. When the state-space model is nonlinear or non-Gaussian, authors in [66]–[70] have proposed the use of particle filters for the dynamic parameter estimation. In this section we discuss how cooperative tracking can be achieved when inter-node communication is possible. In this case, we assume that each node observes an underlying state which evolves, for instance, according to a classical Gauss-Markov model:

$$x(t+1) = F(t)x(t) + u(t). \tag{49}$$

We denote the observation at node k at time instant t by $z_k(t)$, where

$$z_k(t) = H_k(t)x(t) + v_k(t), k = 1, 2, \ldots, n \text{ and } t = 1, 2, \ldots \tag{50}$$

If a centralized observer collects observations from all nodes, it could run a Kalman filter, which is the optimum MSE estimator provided that the noise processes $u(t)$ and $v_k(t)$, $k = 1, \ldots, n$, are jointly Gaussian (otherwise, optimality is still true only among linear estimators). The process noise $u(t)$ and observation noise $\mathbf{v}(t)$ are assumed to be white, independent and have covariances $\mathbf{\Sigma}_u(t)$ and $\mathbf{\Sigma}_v(t)$ respectively, with

$$\mathbf{\Sigma}_v(t) \doteq \begin{pmatrix} \Sigma_{v(1)}(t) & 0 & \cdots & 0 \\ 0 & \Sigma_{v(2)}(t) & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \Sigma_{v(n)}(t) \end{pmatrix}. \tag{51}$$

However, if a central fusion center is not available, nodes can benefit from cooperation with neighbors. The degree of improvement as well as the performance degradation with respect to that of a central fusion center will depend on how close to an agreement the network can go before the next state update in (49). This statement can be made more clear after the formulation of the *information form* centralized Kalman filter for the recursive update of the linear estimator $\hat{x}(t)$ of $x(t)$ based on the observations up to time instant t [71]:

$$\begin{aligned} P^{-1}(t)\hat{x}(t) &= F^{-1}(t)(1 - K(t))P^{-1}(t-1)\hat{x}(t-1) \\ &+ \mathbf{H}^T(t)\mathbf{\Sigma}_v^{-1}(t)\mathbf{z}(t) \end{aligned} \tag{52}$$

with

$$\mathbf{z}(t) \doteq \begin{bmatrix} z_1(t) & z_2(t) & \cdots & z_n(t) \end{bmatrix}^T, \tag{53}$$

$$\mathbf{H}(t) \doteq \begin{bmatrix} H_1(t) & H_2(t) & \cdots & H_n(t) \end{bmatrix}^T, \tag{54}$$

and where $P(t) = \mathbb{E}\{(x(t) - \hat{x}(t))^2\}$ is computed through the following updates:

$$R(t) = \Sigma_u^{-1}(t) + F^{-1}(t)P^{-1}(t-1)F^{-1}(t), \tag{55}$$

$$K(t) = F^{-1}(t)P^{-1}(t-1)F^{-1}(t)R^{-1}(t), \tag{56}$$

$$\begin{aligned} P^{-1}(t) = {} & F^{-1}(t)P^{-1}(t-1)F^{-1}(t) \\ & + \mathbf{H}^T(t)\mathbf{\Sigma}_v^{-1}(t)\mathbf{H}(t) - K(t)R(t)K(t), \end{aligned} \tag{57}$$

for an initial uncertainty $P(-1)$. As observed in [61] the above equations show that each sensor would be able to emulate a central fusion center should it be able to compute the following quantities:

$$\mathbf{H}^T(t)\mathbf{\Sigma}_v^{-1}(t)\mathbf{z}(t) = \sum_{k=1}^{n} H_k(t)z_k(t)/\Sigma_{v(k)}(t) \tag{58}$$

$$\mathbf{H}^T(t)\mathbf{\Sigma}_v^{-1}(t)\mathbf{H}(t) = \sum_{k=1}^{n} H_k^2(t)/\Sigma_{v(k)}(t) \tag{59}$$

As stated above, any linear combination of the initial sensor values can be asymptotically achieved in a distributed form via collaborative exchanges of local information, although in a time-varying setting the number of available exchanges will determine the level of alignment of the quantities (58) and (59). As expected, if the number of exchanges without channel noise goes to infinity, the performance of the network approaches that of a centralized observer [72]. Clearly, this requires the exchange of the observations as well as associated signal-to-noise ratios.

As an illustration, let us consider in detail the setting with one hop exchanges in which each sensor has access to the (noisy) observations from its closest neighbors exclusively. From (58) and (59) we see that nodes must exchange $H_k(t)z_k(t)/\Sigma_{v(k)}(t)$ and $H_k^2(t)/\Sigma_{v(k)}(t)$, so the combined observations can be written as

$$\bar{\mathbf{z}}(t) = \bar{\mathbf{H}}(t)x(t) + \bar{\mathbf{v}}(t) = \mathbf{A}\begin{pmatrix} H_1(t)z_1(t)/\Sigma_{v(1)}(t) \\ \vdots \\ H_n(t)z_n(t)/\Sigma_{v(n)}(t) \end{pmatrix} + \mathrm{Diag}\{\mathbf{A}\mathbf{W}(t)\} \tag{60}$$

where we have defined the extended observation vector as

$$\bar{\mathbf{z}}(t) \doteq \begin{bmatrix} \bar{z}_1(t) & \bar{z}_2(t) & \cdots & \bar{z}_n(t) \end{bmatrix}^T. \tag{61}$$

If the weighting matrix is chosen, for simplicity, as $\mathbf{A} = \mathbf{I} - \gamma\mathbf{L}$, although more sophisticated weighthing coefficients can also be used, then the k-th element of the n-vector $\bar{\mathbf{H}}(t)$ and the covariance matrix of the vector noise $\bar{\mathbf{v}}(t)$ in (60) are given in the following lemma.

Lemma 17: The state-space model at the k-th node when merging observations as in (60) is given by

$$x(t+1) \quad = \quad F(t)x(t) + u(t) \tag{62}$$

$$\bar{z}_k(t) \quad = \quad \bar{H}_k(t)x(t) + \bar{v}_k(t), \tag{63}$$

for k = 1,..., n, with

$$\bar{H}_k(t) = \frac{H_k^2(t)}{\Sigma_{\bar{v}(k)}(t)} - \gamma \sum_{j=1}^{n} \frac{H_j^2(t)}{\Sigma_{\bar{v}(j)}(t)} [\mathcal{A}]_{kj}. \tag{64}$$

In addition, if the noise power σ_w^2 is the same for all communication links, then

$$\mathbf{\Sigma}_{\bar{v}}(t) = \mathbf{A} \begin{pmatrix} H_1^2(t)/\Sigma_{v(1)}(t) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & H_n^2(t)/\Sigma_{v(n)}(t) \end{pmatrix} \mathbf{A}^T + \gamma^2 \sigma_w^2 \mathbf{D}.$$

It is important to note that the covariance matrix of the vector noise $\bar{\mathbf{v}}(t)$ containing the n components $\bar{v}_k(t)$ in (17) is now a function of the original observation noise statistics, the weighting matrix used, and the communication noise. Each node can iterate the corresponding Kalman filter provided that the communication of $H_k^2(t)/\Sigma_{v(k)}(t)$ is noiseless (Note that, in practice, these values may change slowly making robust communication easier). As a consequence, after each measurement the corresponding values are exchanged with neighbors to assist the Kalman updating at each node. The weight γ should be chosen as a function of the channel noise, with a low value for highly unreliable channels [73].

It is also possible to exchange state estimates instead of the described observations merging, or combine both if enough bandwidth is available. In any case, the number of exchanges will determine the performance of the distributed Kalman filter.

6. Chapter summary

In this chapter, we discussed the fusion performance of estimation of a scalar parameter, be it static or dynamic by a distributed sensor network in the presence communication noise. The distributed nodes in the network make observations on a PoI and make local decisions based on the observations. Then by communicating the local summaries over wireless channels (to a central node or to their local neighbors), a final estimator is obtained.

In section 2, the multi-sensor data fusion problem was formulated and two types of local processing schemes were discussed. Then these locally processed data was transmitted over wireless channels to a central node or shared among each other. The communication noise was allowed to be multiplicative (due to path loss attenuation and the multi-path fading) or additive (due to receiver noise). In section 3, the data fusion was analyzed when nodes communicate with a central node that forms the final estimator.

When there is no central fusion center available in the network, nodes may communicate with each other to improve their local estimators by combining them with those of their neighbors. Sections 4 and 5 addressed this type of sensor nets in two different contexts.

Section 4 discussed the distributed sequential estimation in which nodes communicate sequentially to update their estimator and the final decision can be made at any distributed node. In Section 5, the distributed estimation with consensus was considered where all nodes try to reach a final decision that agrees with each other via inter-node communication among neighbors.

7. References

[1] R. S. Smith and F. Y. Hadaegh, "Distributed estimation, communication and control for deep space formations," *IET control Theory and Applications*, vol. 1, no. 2, pp. 445 – 451, March 2007.

[2] A. Speranzon, C. Fischione, and K. H. Johansson, "A distributed estimation algorithm for tracking over wireless sensor networks," in *IEEE Int. Conf. on Communications (ICC)*, June 2007, pp. 3088 – 3093.

[3] R. Nowak, U. Mitra, and R. Willett, "Estimating inhomogeneous fields using wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 999–1006, Aug. 2004.

[4] H. Wang, J. Elson, L. Girod, D. Estrin, and K. Yao, "Target classification and localization in habitat monitoring," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Hong kong, April 2003.

[5] E. Biagioni and K. Bridges, "The application of remote sensor technology to assist the recovery of rare and endangered species," *Int. Journ. of High Performance Computing Applications*, vol. 16, no. 3, pp. 112–121, Fall 2002.

[6] A. Krasnopeev, J.-J. Xiao, and Z.-Q. Luo, "Minimum energy decentralized estimation in a wireless sensor network with correlated sensor noises," *EURASIP Journal on Wireless Commun. and Networking*, vol. 2005, no. 4, pp. 473–482, 2005.

[7] J. F. Chamberland and V. V. Veeravalli, "Asymptotic results for decentarlized detection in power constrained wireless sensor networks," *IEEE J. Select. Areas Commun.*, vol. 22, no. 6, pp. 1007–1015, Aug. 2004.

[8] S. Appadwedula, V. V. Veeravalli, and D. L. Jones, "Energy efficient detection in sensor networks," *IEEE J. Select. Areas Commun.*, vol. 23, no. 4, pp. 693–702, Apr 2005.

[9] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Joint estimation in sensor networks under energy constraints," in *Proc. IEEE first conf. Sensor and Ad Hoc commun. and Networks*, Santa Clara, CA, Oct. 2004.

[10] S. K. Jayaweera, "Large system decentralized detection performance under communication constraints," *IEEE Commun. Lett.*, vol. 9, pp. 769–771, Sep. 2005.

[11] — —, "Bayesian fusion performance and system optimization in distributed stochastic Gaussian signal detection under communication constraints," *IEEE Trans. Signal Processing.*, vol. 55, no. 4, pp. 1238–1250, April. 2007.

[12] J. F. Chamberland and V. V. Veeravalli, "The impact of fading on decentralized detection in power constrained wireless sensor networks," in *Proc. Acoust., Speech, Signal Processing 2004. (ICASSP '04)*, vol. 3, Montreal, Canada, May. 2004, pp. 837–840.

[13] X. Zhang, H. V. Poor, and M. Chiang, "Optimal power allocation for distributed detection in wireless sensor networks," *IEEE Trans. Signal Processing.*, vol. 56, no. 9, pp. 4124–4140, Sept. 2008.

[14] I. Bahceci and A. K. Khandani, "Energy-efficient estimation of correlated data in wireless sensor networks," in *40th Annual Conf. on Information Sciences and Systems (CISS)*, Princeton, NJ, March 2006, pp. 973–978.

[15] A. Speranzon, C. Fischione, and K. H. Johansson, "Distributed and collaborative estimation over wireless sensor networks," in *Proc. 45th IEEE Int. Conf. on Decision and Control (CDC)*, San Diego, CA, Dec. 2006, pp. 1025–1030.

[16] T. Wimalajeewa and S. K. Jayaweera, "Optimal power scheduling for correlated data fusion in wireless sensor networks via constrained PSO," *IEEE Trans. on Wireless commun.*, vol. 7, no. 9, pp. 3608–3618, Sept. 2008.

[17] S. K. Jayaweera and K. Altarazi, "Large system decision fusion performance in inhomogeneous sensor networks," in *IEEE Int. Conf. on Advanced Video and Signal based Surveillance (AVSS)*, Sydney, Australia, Nov. 2006.

[18] S. Marano, V. Matta, and P. Willett, "Distributed estimation in large wireless sensor networks via a locally optimum approach," *IEEE Trans. on Signal Processing*, vol. 56, no. 2, pp. 748–756, Feb. 2008.

[19] I. D. Schizas, G. B. Giannakis, and Z.-Q. Luo, "Distributed estimation using reduced-dimensionality sensor observations," *IEEE Trans. on Signal Processing*, vol. 55, no. 55, pp. 4284 – 4299, Aug. 2007.

[20] Q. Tian and E. J. Coyle, "Optimal distributed estimation in clustered sensor networks," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 4, Toulouse, France, May 2006.

[21] V. Megalooikonomou and Y. Yesha, "Quantizer design for distributed estimation with communication constraints and unknown observation statistics," *IEEE Trans. Communications*, vol. 48, no. 2, pp. 181–184, Feb. 2000.

[22] K. Zhang and X. R. Li, "Optimal sensor data quantization for best linear unbiased estimation fusion," in *Proc. 43rd IEEE Int. Conf. on Decision and Control (CDC)*, Atlantis, Paradise Island, Bahamas, Dec. 2004, pp. 2656–2661.

[23] W. Li and H. Dai, "Distributed detection in wireless sensor networks using a multiple access channel," *IEEE Trans. Signal Processing.*, vol. 55, no. 3, pp. 822–833, March 2007.

[24] G. Mergen and L. Tong, "Type based estimation over multiaccess channels," *IEEE Trans. Signal Processing.*, vol. 54, no. 2, pp. 613–626, Feb. 2006.

[25] G. Mergen, V. Naware, and L. Tong, "Asymptotic detection performance of type-based multiple access over multiaccess fading channels," *IEEE Trans. Signal Processing.*, vol. 55, no. 3, pp. 1081–1092, March 2007.

[26] S. Marano, V. Matta, L. Tong, and P. Willett, "A likelihood-based multiple access for estimation in sensor networks," *IEEE Trans. on Signal Processing*, vol. 55, no. 11, pp. 5155–5166, Nov. 2007.

[27] T. C. Aysal and K. E. Barner, "Blind decentralized estimation for bandwidth constrained wireless sensor networks," *IEEE Trans. on Wireless commun.*, vol. 7, no. 5, pp. 1466–1471, May 2008.

[28] C. Thejaswi and R. K. Patro, "Distributed detection in bandwidth constrained wireless sensor networks," in *IEEE Radio and Wireless Symposium (RWS)*, San Diego, CA, Jan. 2006, pp. 95–98.

[29] Z.-Q. Luo and J.-J. Xiao, "Universal decentralized estimation in a bandwidth constrained sensor network," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2210–2219, Jun 2005.

[30] Z.-Q. Luo, "An isotropic universal decentralized estimation scheme for a bandwidth constrained ad hoc sensor network," *IEEE Journal on Selected areas in Commun.*, vol. 23, no. 4, pp. 735–744, April 2005.

[31] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks- part II: Unknown probability density function," *IEEE Trans. on Signal Processing*, vol. 54, no. 7, pp. 2784–2796, July 2006.

[32] — —, "Bandwidth-constrained distributed estimation for wireless sensor networks-part I: Gaussian case," *IEEE Trans. on Signal Processing*, vol. 54, no. 3, pp. 1131–1143, Mar. 2006.

[33] S. Cui, J.-J. Xiao, A. J. Goldsmith, Z.-Q. Luo, and H. V. Poor, "Estimation diversity and energy efficiency in distributed sensing," *IEEE Trans. on Signal Processing*, vol. 55, no. 9, pp. 4683 – 4695, Sept. 2007.

[34] T. Wimalajeewa and S. K. Jayaweera, "Power efficient distributed estimation in a bandlimited wireless sensor network," in *41st Asilomar Conference on Signals, Systems and Computers (ACSSC)*, Pacific Grove, CA, Nov. 2007, pp. 2156–2160.

[35] P. Ishwar, R. Puri, K. Ramchandran, and S. S. Pradhan, "On rate-constrained distributed estimation in unreliable sensor networks," *IEEE Jour. on Selected Areas of Communications*, vol. 23, no. 4, pp. 765 – 775, April 2005.

[36] T. M. Duman and M. Salehi, "Optimal quantization for distributed estimation via a multipleaccess channel," in *IEEE Int. Symp. on* Inform. Theory (ISIT), Whistler, Canada, Sept. 1995.

[37] T. Zhao and A. Nehorai, "Distributed sequential Bayesian estimation of a diffusive source in wireless sensor networks," *IEEE Trans. Signal Processing.*, vol. 55, no. 4, pp. 1511–1524, April 2007.

[38] M. Chu, H. Haussecker, and F. Zhao, "Scalable information-driven sensor querying and routing for ad hoc heterogeneous sensor networks," *Int. J. High-Perform. Comput. Applicat.*, vol. 16, no. 3, pp. 90–110, 2002.

[39] T. Zhao and A. Nehorai, "Information-driven distributed maximum likelihood estimation based on Gauss-Newton method in wireless sensor networks," *IEEE Trans. on Signal Processing*, vol. 55, no. 9, pp. 4669–4682, Sept. 2007.

[40] H. R. Hashemi and I. B. Rhodes, "Decentralized sequential detection," *IEEE Trans. Inform. Theory.*, vol. 35, no. 3, pp. 509–520, May 1989.

[41] D. Teneketzis and Y. C. Ho, "The decentralized Wald problem," *Inform. Computation*, vol. 73, no. 1, pp. 23–44, 1987.

[42] S. K. Jayaweera and C. Mosquera, "Distributed sequential estimation with noisy, correlated observations," *IEEE Signal Processing Letters*, July 2008 accepted.

[43] S. M. Kay, Fundementals of Statistical Signal Processing: Estimation Theory. New Jersy, USA: Prentice Hall PTR, 1998.

[44] K. Liu and A. Sayeed, "Optimal distributed detection strategies for wireless sensor networks," in *42nd annaul Allerton conf. commun., control and computing*, Monticello, IL, Oct. 2004.

[45] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Willey and Sons, Inc., NY, 2006.

[46] K. Altarazi, S. K. Jayaweera, and V. Aravinthan, "Performance of decentralized detection in a resource-constrained sensor network," in *39th Anuual Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2005.

[47] G. Barriac, R. Mudumbai, and U. Madhow, "Distributed beamforming for information transfer in sensor networks," in *Third Int. Symp. on Information Processing in Sensor Networks (IPSN 2004)*, vol. 26-27, Berkeley, CA, April 2004, pp. 81–88.

[48] Y. Lin, B. Chen, and L. Tong, "Distributed detection over multiple access channels," in *IEEE International Conference on, Acoustics, Speech and Signal Processing, (ICASSP)*, Honolulu, Hawaii, USA, April 2007, pp. 541–544.

[49] A. Wald, "Sequential tests of statistical hypotheses," *The Annals of Mathematical Statistics*, vol. 9, no. 2, pp. 117–186, June 1945.

[50] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1994.

[51] V. Dragalin, A. G. Tartakovsky, and V. V. Veeravalli., "Multihypothesis sequential probability ratio tests, part I: Asymptotic optimality," *IEEE Trans. on Info. Theory*, vol. 45, no. 7, pp. 2448–2462, Nov. 1999.

[52] ——, "Multihypothesis sequential probability ratio tests, part II: Accurate asymptotic expansions for the expected sample size," *IEEE Trans. on Info. Theory*, vol. 46, no. 4, pp. 1366–1383, July 2000.

[53] F. Zhao, J. Shin, and J. Reich, "Information-driven dynamic sensor collaboration," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 61–72, March 2002.

[54] A. H. Sayed and C. G. Lopes, "Distributed recursive least-squares strategy over adaptive networks," in *Proc. 40th Asilomar Conf. on Signals, Systems and Computers (ACSSC)*, Pacific Grove, CA, 2006, pp. 233–237.

[55] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, Jan. 2007.

[56] L. Xiao, S. Boyd, and S.-J. Kim, "Distributed average consensus with least-mean-square deviation," *Journal of Parallel and Distributed Computing*, vol. 67, no. 1, pp. 33–46, Jan 2007.

[57] Y. Hatano and M. Mesbahi, "Agreement over random networks," *IEEE Trans. on Automatic Control*, vol. 50, no. 11, pp. 1867–1872, Nov. 2005.

[58] C. Mosquera, R. López-Valcarce, and S. Jayaweera, "Distributed Estimation with Noisy Exchanges," in *9th IEEE Int. Workshop on Signal Processing Advances in Wireless Communications, Recife, Brazil*, July 2008.

[59] S. Barbarossa and G. Scutari, "Decentralized maximum-likelihood estimation for sensor networks composed of nonlinearly coupled dynamical systems," *IEEE Trans. on Signal Processing*, vol. 55, no. 7-Part 1, pp. 3456 – 3470, July 2007.

[60] R. Olfati-Saber, "Distributed Kalman filtering for sensor networks," in *46th IEEE Conf. on Decision and Control (CDC)*, New Orleans, LA, USA, Dec. 2007, pp. 5492 – 5498.

[61] ——, "Distributed Kalman Filter with Embedded Consensus Filters," in *Proc. of 44th IEEE Conf. on Decision and Control (CDC)*, Sevilla, Spain, Dec 2005, pp. 8179–8184.

[62] D. P. Spanos, R. O. Saber, and R. M. Murray, "Approximate distributed Kalman filtering in sensor networks with quantifiable performance," in *Proc. 4th Int. Symp. Information Processing in Sensor Networks (IPSN)*, Los Angeles, CA, USA, Apr. 2005, pp. 133–139.

[63] U. A. Khan and J. F. Moura, "Distributed kalman filters in sensor networks: Bipartite fusion graphs," in *14th Workshop on Statistical Signal Processing (SSP)*, Madison, WI, USA, Aug. 2007, pp. 700 – 704.

[64] C.-J. Ran and Z.-L. Deng, "Two average weighted measurement fusion Kalman filtering algorithms in sensor networks," in *7th World Congress on Intelligent Control and Automation (WCICA)*, Chongqing, China, June 2008, pp. 2387–2391.

[65] S. Kirti and A. Scaglione, "Scalable distributed Kalman filtering through consensus," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Las Vegas, NV, USA, March-April 2008, pp. 2725–2728.

[66] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. on Signal Processing*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

[67] M. Coates, "Distributed particle filters for sensor networks," in *3rd Int. Symp. on Information Processing in Sensor Networks (IPSN)*, April 2004, pp. 99 – 107.

[68] D. Gu, J. Sun, Z. Hu, and H. Li, "Consensus based distributed particle filter in sensor networks," in *Int. Conf. on Information and Automation (ICIA)*, Hunan, China, June 2008, pp. 302–307.

[69] P. M. Djuric, M. Vemula, and M. F. Bugallo, "Target tracking by particle filtering in binary sensor networks," *IEEE Trans. on Signal Processing*, vol. 56, no. 6, pp. 2229–2238, June 2008.

[70] A. Ahmad and M. R. Gani, "Distributed estimation for nonlinear stochastic signals over sensor networks," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Las Vegas, NV, USA, March-April 2008, pp. 3425 – 3428.

[71] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Prentice Hall, 2000.

[72] I. Schizas, G. B. Giannakis, A. Ribeiro, and S. I. Roumeliotis, "Consensus in Ad Hoc WSNs with Noisy Links - Part II: Distributed Estimation and Smoothing of Random Signals," *IEEE Trans. on Signal Processing*, vol. 56, no. 4, pp. 1650–1666, April 2008.

[73] C. Mosquera and S. Jayaweera, "Entangled Kalman Filters for Cooperative Estimation," in *5th IEEE Sensor Array and Multichannel Signal Processing Workshop, Darmstadt (Germany)*, 2008.