Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Distributed Maximum Likelihood Estimation with Time-Varying Network Topology

Giuseppe Calafiore * **Fabrizio Abrate** **

* *Dipartimento di Automatica e Informatica, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy, e-mail: giuseppe.calafiore@polito.it*
** *Dipartimento di Automatica e Informatica, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy, e-mail: fabrizio.abrate@polito.it*

Abstract: We consider a sensor network in which each sensor may take at every time iteration a noisy linear measurement of some unknown parameter. In this context, we study a distributed consensus diffusion scheme that relies only on bidirectional communication among neighbor nodes (nodes that can communicate and exchange data), and allows every node to compute an estimate of the unknown parameter that asymptotically converges to the true parameter. At each time iteration, a measurement update and a spatial diffusion phase are performed across the network, and a local least-squares estimate is computed at each node. We show that the local estimates converge to the true parameter value, under suitable hypotheses. The proposed scheme works in networks with dynamically changing communication topology, and it is robust to unreliable communication links and widespread failures in measuring nodes.

Keywords: Distributed estimation, consensus, sensor networks, sensor fusion.

1. INTRODUCTION

Recent technological improvements have allowed the deployment of small inexpensive and low-power devices that can perform local data processing and communicate with other sensors being part of a network. Each sensor node has limited storage capacity and processing power. However, when it is used together with a large number of other nodes, the network as a whole has the ability to perform complex tasks. These technological achievements have allowed the growth of many and varied applications of sensor networks, mainly in commercial and industrial applications, to manage data that would be difficult or expensive to deal with using wired sensors. Typical applications include monitoring, tracking, and controlling. Some of the specific applications are habitat monitoring, surveillance, object tracking, health-care applications, collaborative information processing, traffic monitoring and mobile agents control, see for instance Akyildiz et al. [2002], Chu et al. [2002], Martinez and Bullo [2006].

In each of these application fields, estimation and fusion of data coming from sensors is one of the most challenging tasks. Various schemes for sensor data fusion exist, both centralized or distributed. In a centralized scheme, a sensor has to send data (directly or by finding a suitable path in the network) to a data fusion center. This center is able to compute the best possible estimate of some unknown parameter (e.g. the Maximum Likelihood (ML) estimate). Depending on how data are sent to the data fusion center, communication problems may arise, especially if the network topology changes with time. In a distributed processing scheme, instead, each sensor exchanges data only with its neighbors, and carries out local computation in order to obtain a good estimate of the unknown parameter of interest. Distributed processing has several advantages w.r.t centralized processing: there is no central data fusion center, and each sensor can compute the estimates on its own without having any knowledge of the whole network. One of the most intuitive way to do distributed sensor fusion is flooding. This technique assumes that each sensor diffuses all its data to all the other nodes. In this way each sensor becomes a fusion center, but high communication, storage and computing capabilities need be allocated to the nodes. Many sophisticated algorithms for distributed estimation and tracking exist, see for instance Alanyali et al. [2004], Delouille et al. [2004], Luo [2005], Tsitsiklis [1993], R. Rahman and Saligrama [2007]. In Delouille et al. [2004], an iterative distributed algorithm for linear minimum mean-squared-error (LMMSE) estimation in sensor networks is proposed, while in Alanyali et al. [2004] consensus among distributed noisy sensors observing an event is addressed. In Olfati-Saber [2004], Spanos et al. [2005a,b], a distributed version of the Kalman filter (DKF) is analyzed for distributed estimation of time-varying parameters, while in R. Rahman and Saligrama [2007] distributed tracking of a nonlinear dynamical system via networked sensors is described.

In this paper we start from the setup of Xiao et al. [2006], and analyze a distributed consensus diffusion scheme for linear parameter estimation on networks with unreliable links. Each node in the network is allowed to take at each time t a noisy linear measurement of the unknown parameter. The nodes measurement noise covariances are time-varying, and also the network topology may change with time, being connected at some instants and not connected at other times. We prove that if the *frequency of connectedness* of the communication graph is lower-bounded by a quantity proportional to the logarithm of time, then as $t \to \infty$ the estimates at each local node converge to the true parameter value in the mean square sense. This result may be considered as an extension of the results obtained in Xiao et al. [2006], where the convergence of estimation is rigorously proved only in the case when sensors take a finite number of measurements while performing an infinite number of "spatial" updates.

The rest of this paper is organized as follows. In §2 we describe the distributed scheme for parameter estimation. In §3 we prove our main convergence result for the distributed estimation scheme. In §4 we demonstrate our approach with numerical examples. Finally in §5 we draw the conclusions.

1.1 Notations

$X^\top$ denotes the transpose of a square matrix X. $X \succ 0$ (resp. $X \succeq 0$) means X is symmetric, and positive-definite (resp. semidefinite). $\|X\|$ denotes the spectral (maximum singular value) norm of X, or the standard Euclidean norm, in case of vectors. I_n denotes the $n \times n$ identity matrix, and $\mathbf{1}_n$ denotes a n-vector of ones; subscripts with dimensions are omitted whenever they can be inferred from context.

2. CONSENSUS-BASED ESTIMATION SCHEME

2.1 Preliminaries

Consider n distributed sensors (nodes), each of which takes at time t a measurement of an unknown parameter $\theta \in \mathbb{R}^m$ according to the linear measurement equation

$$y_i(t) = A_i(t)\theta + v_i(t), \quad i = 1, \ldots, n; \; t = 0, 1, \ldots$$

where $y_i(t) \in \mathbb{R}^{m_i}$ is the noisy measurement from the i-th sensor at time t, $v_i(t) \in \mathbb{R}^{m_i}$ is measurement noise, and $A_i(t) \in \mathbb{R}^{m_i,m}$ is the time-varying regression matrix.

We assume $v_i(t)$ to be independent zero mean Gaussian random vectors, with possibly time-varying covariances. Allowing the covariance matrices to be time-varying helps modeling realistic circumstances. If a sensor has a correct measurement at time t, we set its covariance matrix to $\Sigma_i(t) = \Sigma_i$, where Σ_i is fixed and determined by the the technical characteristics of the i-th sensor. If instead the sensor does not have a valid measurement at time t (for any reason, including sensor failures), then we set $\Sigma_i(t)^{-1} = 0$, thus neglecting the measurement.
Notice that if full centralized information were available, the optimal Maximum Likelihood (ML) estimation $\hat{\theta}_{\mathrm{ml}}$ of the parameter θ could be obtained. Defining the quantities

$$P_{\mathrm{ml}}(t) \doteq \sum_{k=0}^{t-1}\sum_{j=1}^{n} A_j^\top(k)\Sigma_j^{-1}(k)A_j(k), \tag{1}$$

$$q_{\mathrm{ml}}(t) \doteq \sum_{k=0}^{t-1}\sum_{j=1}^{n} A_j^\top(k)\Sigma_j^{-1}(k)y_j(k), \tag{2}$$

the ML estimate of θ is

$$\hat{\theta}_{\mathrm{ml}}(t) \doteq P^{-1}{}_{\mathrm{ml}}(t)q_{\mathrm{ml}}(t),$$

and the ML error covariance matrix is

$$Q_{\mathrm{ml}}(t) \doteq P^{-1}{}_{\mathrm{ml}}(t). \tag{3}$$

However, we assume it is not possible (due to communication constraints, etc.) to construct the optimal centralized estimate. Instead, our objective is to exploit peer-to-peer information exchange among communicating nodes, in order to build "good" *local* estimates of θ. We shall prove in Section 3 that under suitable hypothesis, all local estimates converge asymptotically to the true parameter θ, in mean square sense.

Let $\mathcal{V} = \{1, 2, \ldots, n\}$ denote the set of nodes of the sensor network, and let $\mathcal{E}(t)$ denote the set of active links at time t; i.e., nodes (i, j) can communicate at time t if and only if $(i, j) \in \mathcal{E}(t)$ (note that $\mathcal{E}(t)$ contains only pairs of *distinct* nodes). The time-varying communication network is represented by the graph $\mathcal{G}(t) = (\mathcal{V}, \mathcal{E}(t))$. We denote with $\mathcal{N}_i(t)$ the set of nodes that are linked to node i at time t (note again that it is assumed that $i \notin \mathcal{N}_i(t)$), and with $|\mathcal{N}_i(t)|$ the cardinality of $\mathcal{N}_i(t)$; $|\mathcal{N}_i(t)|$ is called the *spatial degree* of node i in graph $\mathcal{G}(t)$. Following the notation in Xiao et al. [2006], we define the *time degree* of node i as the number of measurements that node i has collected up to time t, that is $d_i(t) = t + 1$, and the space-time degree as

$$d_i^{ST}(t) = (1 + |\mathcal{N}_i(t)|)(t + 1).$$

With this position, we introduce the weights that shall be employed for information averaging among neighboring nodes. To this end, we use the Metropolis weights (see Xiao et al. [2006]), defined as

$$\tilde{W}_{ij}(t) = \frac{1}{1 + \max\,\left(|\mathcal{N}_i(t)|, |\mathcal{N}_j(t)|\right)} \cdot \frac{1}{t+1}. \tag{4}$$

The distributed space-time diffusion scheme is described in the next section.

2.2 A distributed space-time diffusion scheme

Assume that every node collects a measurement at each time $t = 0, 1, \ldots$ Note that this is done without loss of generality, since one can assume that $\Sigma_i^{-1}(t) = 0$ if sensor i actually does not have an usable measurement at time t. The proposed distributed iterative scheme performs a temporal update phase and a spatial update phase. Using the same notations of Xiao et al. [2006], we assume that each node keeps as local information a *composite information matrix* $P_i(t)$ and a *composite information state* $q_i(t)$.

At time t a measurement is collected at each node, and a temporal (measurement) update phase is performed locally at the nodes. This phase amounts to computing

$$P_i(t_+) = \frac{t}{t+1}P_i(t) + \frac{1}{t+1}A_i^\top(t)\Sigma_i^{-1}(t)A_i(t) \tag{5}$$

$$q_i(t_+) = \frac{t}{t+1}q_i(t) + \frac{1}{t+1}A_i^\top(t)\Sigma_i^{-1}(t)y_i(t), \tag{6}$$

where each node only has to know its local information $P_i(t)$, $q_i(t)$, and the current time degree $d_i(t)$, which is actually constant for all nodes and equal to $d_i(t) = t+1$. Note that the temporal updates are finished instantaneously at each node, thus $t+$ and t are essentially the same integer.

After the temporal update, each node has to broadcast its space degree and its current values of $P_i(t_+)$ and $q_i(t_+)$ to its neighbors. At this point, a spatial update phase is performed. Considering (4) and defining

$$W_{ij}(t) \doteq \begin{cases} (t+1)\tilde{W}_{ij}(t) & \text{if } (i,j) \in \mathcal{E}(t) \\ 1 - \sum_{j \in \mathcal{N}_i(t)} W_{ij}(t) & \text{if } i = j \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

the i-th node updates the composite information matrix and composite information state at time $t + 1$ as follows:

$$P_i(t+1) = P_i(t_+) + \sum_{j \in \mathcal{N}_i(t)} W_{ij}(t)\big(P_j(t_+) - P_i(t_+)\big) \quad (8)$$

$$q_i(t+1) = q_i(t_+) + \sum_{j \in \mathcal{N}_i(t)} W_{ij}(t)\big(q_j(t_+) - q_i(t_+)\big). \quad (9)$$

Merging the temporal update phase and the spatial update phase leads to the following proposition.

Proposition 1. For $t = 1, 2, ...$ the composite information matrix and composite information state at each node $i = 1, \ldots, n$ are given by the expressions

$$P_i(t) = \frac{1}{t} \sum_{k=0}^{t-1} \sum_{j=1}^{n} \Phi_{ij}(t-1; k) A_j^\top(k) \Sigma_j^{-1}(k) A_j(k)$$

$$q_i(t) = \frac{1}{t} \sum_{k=0}^{t-1} \sum_{j=1}^{n} \Phi_{ij}(t-1; k) A_j^\top(k) \Sigma_j^{-1}(k) y_j(k),$$

where

$$\Phi(t-1; k) \doteq W(t-1) \cdots W(k).$$

The proof for Proposition 1 can be found in Calafiore and Abrate [2007].

Remark 1. Notice that the recursions in (8), (9) are well suited for distributed implementation, since at each step each node only needs to know the current time instant, and the space-time degrees and local informations of its neighbors. In particular, the nodes do not need global knowledge of the communication graph, or even of the number of nodes composing the network. Also, no matrix inversion need be performed in this recursion. Notice further that the expressions in Proposition 1, which are useful for a-posteriori analysis, do *not* describe the actual computations performed by the nodes, which use instead the recursions (8), (9).

The properties of the local estimates are discussed in the next section.

3. PROPERTIES OF LOCAL ESTIMATES

At each time when the composite information matrix $P_i^{-1}(t)$ is invertible, each node i in the network is able to compute its local estimate at time t as

$$\hat{\theta}_i(t) \doteq P_i^{-1}(t) q_i(t), \quad i = 1, \ldots, n.$$

The following fact holds.

Proposition 2. The local estimate $\hat{\theta}_i(t)$ is an unbiased estimator of θ, that is

$$\mathrm{E}\left\{\hat{\theta}_i(t)\right\} = \theta.$$

Moreover, the covariance of the local estimate is given by the expression

$$Q_i(t) = \frac{1}{t^2} P_i^{-1}(t) R(t) P_i^{-1}(t). \quad (10)$$

where

$$R(t) = \left(\sum_{k=0}^{t-1} \sum_{j=1}^{n} \Phi_{ij}^2(t-1; k) A_j^\top(k) \Sigma_j^{-1}(k) A_j(k) \right)$$

$\star$

3.1 Mean square convergence results

We show in this section that, under suitable hypothesis, as the number of measurements goes to infinity, *all* the local estimates $\hat{\theta}_i(t)$ converge to the true parameter value θ, in the mean square sense. That is, $\lim_{t \to \infty} \|Q_i(t)\| = 0$, for $i = 1, \ldots, n$. This result holds for time-varying network topology, and it is derived under two assumptions. The first condition is a very natural one, and requires that the centralized ML estimate mean square error goes to zero at $t \to \infty$. This condition is actually necessary, since one cannot hope to make the local estimates converge when even the centralized optimal estimate (who has all the available information) does not converge. The second condition is a technical sufficient condition needed for proving convergence of the distributed scheme, and is detailed in the sequel. Loosely speaking, this condition requires that the time-varying communication graph is connected at least "rarely" in time.

We first state some technical preliminaries in the next section, whereas the main theorem is stated in Section 3.1.2.

Technical preliminaries Define

$$\bar{P}_i(t) \doteq t P_i(t) = \sum_{k=0}^{t-1} \sum_{j=1}^{n} \Phi_{ij}(t-1; k) H_j(k) \quad (11)$$

$$H_j(k) \doteq A_j^\top(k) \Sigma_j^{-1}(k) A_j(k).$$

The following result holds.

Lemma 1. Whenever $\bar{P}_i^{-1}(t)$ is invertible, the covariance matrix of the i-th local estimate satisfies

$$Q_i(t) \preceq \bar{P}_i^{-1}(t). \quad (12)$$

A proof of This Lemma can be found in Calafiore and Abrate [2007]. We now look more closely at the structure of the $W(t)$ matrices and of the transition matrix $\Phi(t-1; k)$. Notice that $W_{ij}(t) \in [0, 1]$, $\forall i, j$, and $W_{ii}(t) > 0$, $i = 1, \ldots, n$. Also, $W(t)\mathbf{1} = \mathbf{1}$, $\mathbf{1}^\top W(t) = \mathbf{1}^\top$, hence the weight matrices $W(t)$ are symmetric and doubly stochastic. This means that $W(t)$: (a) is unitarily diagonalizable, that is it has a set of orthogonal eigenvectors; and (b) all its eigenvalues are real and have modulus no larger than one. Since $\mathbf{1}/\sqrt{n}$ is always an eigenvector of $W(t)$ associated with the eigenvalue $\lambda_1 = 1$, we may write $W(t)$ in the form

$$W(t) = \frac{1}{n}\mathbf{1}\mathbf{1}^\top + Z(t), \quad Z(t) = V(t) D(t) V^\top(t)$$

where $V(t) \in \mathbb{R}^{n,n-1}$ is such that

$$V^\top(t) V(t) = I_{n-1},$$
$$V(t) V^\top(t) = I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^\top,$$
$$\mathbf{1}^\top V(t) = 0,$$

and $D(t) = \mathrm{diag}(\lambda_2(t), \ldots, \lambda_n(t)) \in \mathbb{R}^{n-1,n-1}$ is a diagonal matrix containing the last $n-1$ eigenvalues of $W(t)$ arranged in order of decreasing modulus. Since $\lambda_1 = 1$ is an eigenvalue of $W(t)$ of maximum modulus, $\lambda_2(t)$ denotes the second-largest-modulus eigenvalue of $W(t)$. Moreover, it is shown in §II.A of Xiao et al. [2005] (see also Xiao et al. [submitted], Xiao and Boyd [2004]) that if the graph is connected at time t, then the spectral radius of $Z(t)$ is strictly less than one. This means that, whenever the graph is connected, $\lambda_1 = 1$ is a simple eigenvalue of $W(t)$, and $|\lambda_2(t)| < 1$, thus the symmetric matrix $Z(t)$ is contractive.

Observe now that from the definition

$$\Phi(t-1;k) \doteq W(t-1)\cdots W(k),$$

follows that all entries of $\Phi(t-1;k)$ lie in the interval $[0,1]$, and that $\Phi(t-1;k)\mathbf{1} = \mathbf{1}$, $\mathbf{1}^\top\Phi(t-1;k) = \mathbf{1}^\top$, hence also $\Phi(t-1;k)$ is doubly stochastic. Further, since $\mathbf{1}\mathbf{1}^\top$ is orthogonal to *all* matrices $Z(t)$, it results that

$$\Phi(t-1;k) = \frac{1}{n}\mathbf{1}\mathbf{1}^\top + \Upsilon(t-1;k), \tag{13}$$

where we defined

$$\Upsilon(t-1;k) \doteq Z(t-1)Z(t-2)\cdots Z(k). \tag{14}$$

Note that for any $t \geq 1$, $Z(t)$ is symmetric, hence $\|Z(t)\| = |\lambda_2(t)|$. Therefore, from (14) and from the submultiplicativity of matrix norm, we have that

$$\|\Upsilon(t-1;k)\| \leq |\lambda_2(t-1)|\cdot|\lambda_2(t-2)|\cdots|\lambda_2(k)|, \tag{15}$$

where $\|\cdot\|$ denotes the spectral (maximum singular value) matrix norm.

Main result We are now in position to state in the next theorem our main result for convergence of distributed least squares estimation. Our convergence proof hinges upon the following assumption.

Assumption 1. (Frequency of connectedness). Let $\mathcal{G}(t)$ denote the connection graph at time t, and let $W(t)$ be the corresponding weight matrix, defined in (7). Let $\bar\lambda < 1$ be a positive constant such that each time $\mathcal{G}(t)$ is connected, it holds that $|\lambda_2(t)| \leq \bar\lambda$.

We assume that for any interval of time $[\tau, \tau+k)$ of length k, the graph is connected at least $N(k)$ times, with

$$N(k) = \frac{2}{\log(1/\bar\lambda)}\log(k/\alpha), \quad \text{for some } \alpha > 0. \tag{16}$$

⋆

Notice that this assumption requires the time-varying graph to be connected only quite "rarely", since $N(k)$ grows slowly with the logarithm of the time length. The following key result holds.

Theorem 1. Let Assumption 1 hold, and let $\|H_j(k)\| \leq C$, for all $j = 1,\ldots,n$, $k = 0,1,\ldots$

If $\lim_{t\to\infty}\|P_{\rm ml}(t)\| = \infty$ (or, equivalently, if the centralized maximum likelihood error covariance goes to zero), then

$$\lim_{t\to\infty}\|Q_i(t)\| = 0, \quad i = 1,\ldots,n.$$

⋆

Proof. Consider the expression of $\bar P_i(t)$ in (11), and substitute (13) to obtain

$$\begin{aligned}\bar P_i(t) &= \sum_{k=0}^{t-1}\sum_{j=1}^{n}\frac{1}{n}H_j(k) + \sum_{k=0}^{t-1}\sum_{j=1}^{n}\Upsilon_{ij}(t-1;k)H_j(k)\\ &= \frac{1}{n}P_{\rm ml}(t) + \sum_{k=0}^{t-1}\sum_{j=1}^{n}\Upsilon_{ij}(t-1;k)H_j(k). \end{aligned}\tag{17}$$

Recall now that for any two matrices A, B and any norm, applying the triangle inequality to the identity $A = (-B)+(B+A)$, it results that $\|A+B\| \geq \|A\|-\|B\|$. Applying this inequality to (17), and taking the spectral norm, we have

$$\|\bar P_i(t)\| \geq \frac{1}{n}\|P_{\rm ml}(t)\| - \|\sum_{k=0}^{t-1}\sum_{j=1}^{n}\Upsilon_{ij}(t-1;k)H_j(k)\|. \tag{18}$$

Notice further that

$$\begin{aligned}\|\sum_{k=0}^{t-1}\sum_{j=1}^{n}\Upsilon_{ij}(t-1;k)H_j(k)\| &\leq \sum_{k=0}^{t-1}\sum_{j=1}^{n}|\Upsilon_{ij}(t-1;k)|\cdot\|H_j(k)\|\\ &\leq C\sum_{k=0}^{t-1}\sum_{j=1}^{n}|\Upsilon_{ij}(t-1;k)|\\ &\leq \sqrt{n}C\sum_{k=0}^{t-1}\|\Upsilon(t-1;k)\|\\ \text{[from (15)]} &\leq \sqrt{n}C\sum_{k=0}^{t-1}|\lambda_2(t-1)|\cdot|\lambda_2(t-2)|\cdots|\lambda_2(k)|\\ &= \sqrt{n}C\sum_{k=1}^{t}\prod_{\tau=1}^{k}|\lambda_2(t-\tau)|.\end{aligned}$$

From Assumption 1, let $N(k)$ denote the lower bound (16) on the number of times in which the graph is connected in any time interval of length k, and recall that $|\lambda_2(t)| \leq \bar\lambda < 1$, if the the graph is connected at t. Therefore, we have that

$$\prod_{\tau=1}^{k}|\lambda_2(t-\tau)| \leq \bar\lambda^{N(k)} = \bar\lambda^{\frac{2\log(k/\alpha)}{\log(1/\bar\lambda)}} = \frac{\alpha^2}{k^2}.$$

Continuing the previous chain of inequalities, we thus have

$$\begin{aligned}\sqrt{n}C\sum_{k=1}^{t}\prod_{\tau=1}^{k}|\lambda_2(t-\tau)| &\leq \sqrt{n}C\sum_{k=1}^{t}\frac{\alpha^2}{k^2}\\ &= \alpha^2\sqrt{n}C\sum_{k=1}^{t}\frac{1}{k^2}.\end{aligned}$$

Going back to (18), we hence obtain that

$$\begin{aligned}\|\bar P_i(t)\| &\geq \frac{1}{n}\|P_{\rm ml}(t)\| - \|\sum_{k=0}^{t-1}\sum_{j=1}^{n}\Upsilon_{ij}(t-1;k)H_j(k)\|\\ &\geq \frac{1}{n}\|P_{\rm ml}(t)\| - \alpha^2\sqrt{n}C\sum_{k=1}^{t}\frac{1}{k^2}.\end{aligned}$$

Recalling now the value of the convergent series $\sum_{k=1}^{\infty}\frac{1}{k^2} = \frac{\pi}{6}$, we have in the limit

$$\lim_{t\to\infty}\|\bar P_i(t)\| \geq \frac{1}{n}\lim_{t\to\infty}\|P_{\rm ml}(t)\| - \frac{\pi}{6}\alpha^2\sqrt{n}C.$$

Since by hypothesis $\lim_{t\to\infty}\|P_{\rm ml}(t)\| = \infty$, we obtain that

$$\lim_{t\to\infty}\|\bar P_i(t)\| = \lim_{t\to\infty}\|P_{\rm ml}(t)\| = \infty.$$

Finally, from (12) it follows that whenever $\bar P_i(t)$ is invertible

$$\|Q_i(t)\| \leq \|\bar P_i^{-1}(t)\| = \frac{1}{\|\bar P_i(t)\|}$$

hence

$$\lim_{t\to\infty} \|Q_i(t)\| \leq \lim_{t\to\infty} \frac{1}{\|\bar{P}_i(t)\|} = 0,$$

which concludes the proof. □

Remark 2. A few remarks are in order in respect to the result in Theorem 1. First, we notice that it is not required by the theorem that each individual node collects an infinite number of measurements as $t \to \infty$. Indeed, the local estimate at a node may converge even if this node *never* takes a measurement, as long as the other hypotheses are satisfied. As an extreme situation, even if only one node in the network takes measurements, then local estimates at *all* nodes converge, if the hypotheses are satisfied.

Further, it is worth to underline that the convergence result presented here is quite different from related results given in Xiao et al. [2006]. The main situation considered in Xiao et al. [2006] assumes that the total number of measurements collected by the whole set of sensors remains finite as $t \to \infty$; contrary, we allow this number to grow as time grows, which seems a more natural requirement. Besides technicalities, considering the number of measurements to remain finite essentially amounts to assuming that, from a certain time instant on, the network evolves with "spatial" iterations only. This in turn permits to apply standard tools for convergence of products of stochastic matrices, and to give results under weaker hypotheses of connectedness of the union of the infinitely occurring graphs, see Xiao et al. [2006] and the references therein. It appears instead that these results cannot be directly applied to our setup, due to the persistent presence of new measurements, which acts as a forcing term in the local iterations (5), (6). ♢

4. NUMERICAL EXAMPLES

In this section, we illustrate the distributed estimation algorithm on some numerical examples. We considered two different situations:

- A first example shows the estimation performance in a middle-sized network, in two different scenarios, with increasing sensor measurement rate.
- The second example is built in order to show that the proposed distributed scheme may converge even under weaker hypothesis than the ones assumed in Theorem 1.

4.1 Example 1

We considered a sensor network with $n = 50$ nodes chosen uniformly at random on the unit square $[0, 1] \times [0, 1]$. We assumed that two nodes in the network are connected by an edge if their distance is less than 0.25. In this particular example we hence obtained a fixed-topology network with 184 edges.

The vector of unknown parameters has dimension $m = 5$, and each sensor takes a scalar measurement $y_i = a_i^\top \theta + v_i$, where the vectors a_i have been chosen from an uniform distribution on the unit sphere in $\mathbb{R}^5$. The noise is i.i.d. Gaussian with unit variance. To quantify the estimation performances in the network, we define an average index of the local mean square estimation errors:

$$\overline{\mathbf{MSE}}(t) = \frac{1}{n}\sum_{i=1}^{n} \text{Tr } (Q_i(t)).$$

For the purpose of comparison, we also compute the *Maximum Likelihood Error* (MLE) as

$$\mathbf{MLE}(t) = \text{Tr } (Q_{\text{ml}}(t)).$$

Two experiments with increasing sensor measurement rate have been carried out. The sensor measurement rate is, in this context, the probability p with which a sensor takes a measurement at any time iteration, meaning that at each step, the inverse covariance matrix $\Sigma_i^{-1}(t)$ associated to the i-th sensor is set to the identity with probability p or to zero with probability 1-p. In the first experiment the measurement rate of each node is $p = 0.01$ (Figure 1), in the second one it is $p = 1$ (Figure 2).

In each plot, the $\overline{\mathbf{MSE}}(t)$ and the $\mathbf{MLE}(t)$ are shown. It can be noticed that, after a transient, the algorithm performance appears to improve proportionally with the sensors measurement rate.

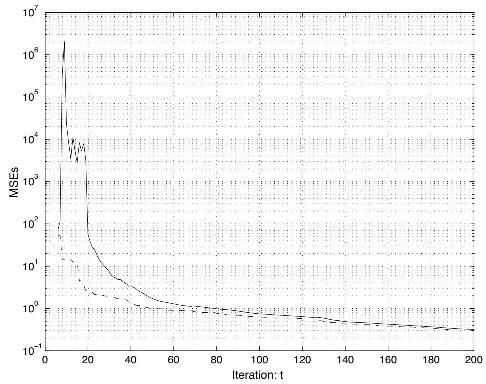

Fig. 1. Example 1: Sensor measurement rate equal to 0.01; Dashed line: $\mathbf{MLE}(t)$; Solid line: $\overline{\mathbf{MSE}}(t)$.

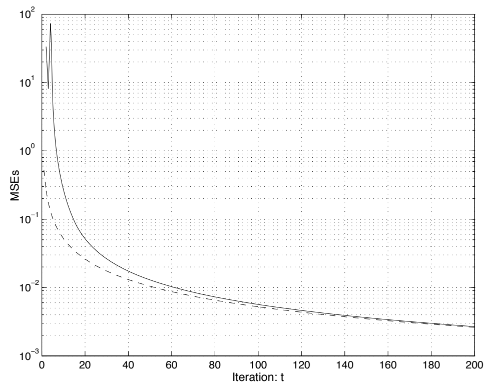

Fig. 2. Example 1: Sensor measurement rate equal to 1; Dashed line: $\mathbf{MLE}(t)$; Solid line: $\overline{\mathbf{MSE}}(t)$.

4.2 Example 2

In a second numerical example, we considered a network with three nodes and a switching connection topology. The connection is as in Figure 3(a) for t odd, and as in Figure 3(b) for t even.

The vector of unknown parameters is of dimension $m = 2$, and only sensors 2 and 3 are able to collect measurements. These sensors take at each iteration a scalar measurement $y_{2,3} = a_{2,3}^\top \theta + v_{2,3}$, where $a_2 = [\,1\ 0\,]$, $a_3 = [\,0\ 1\,]$, and where the measurement covariance matrices are defined as follows:

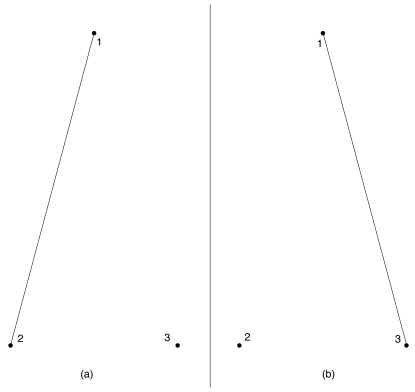

Fig. 3. Example 2: When t is odd the occurring graph is the shown in (a), while when t is even the occurring graph is shown in (b).

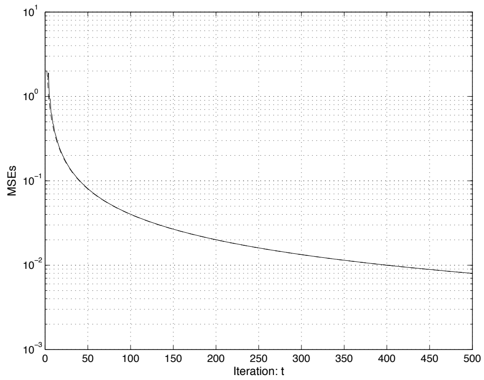

Fig. 4. Example 2: Dashed line: $\mathbf{MLE}(t)$; Solid line: $\overline{\mathbf{MSE}}(t)$.

$$\Sigma_2^{-1}(t) = \begin{cases} 0 & \text{if } t \text{ is odd} \\ 1 & \text{otherwise} \end{cases}, \qquad \Sigma_3^{-1}(t) = \begin{cases} 1 & \text{if } t \text{ is odd} \\ 0 & \text{otherwise,} \end{cases}$$

that is, the two sensors take a valid measurement alternatively in time. The weight matrix defined in (7) is

$$W(t) = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 0.5 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \text{ for } t \text{ odd, } W(t) = \begin{bmatrix} 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \end{bmatrix}, \text{ for } t \text{ even.}$$

In this example, none of the occurring graphs is connected, hence the assumptions of Theorem 1 are clearly violated. However, analyzing the estimation performances on average in the network using the index $\overline{\mathbf{MSE}}(t)$ adopted in §4.1, we notice that it converges numerically to zero as $t \to \infty$, as can be seen in Figure 4. This fact suggests that the distributed estimation scheme may converge under weaker hypotheses than those assumed in Theorem 1. It also confirms a conjecture along the lines of Xiao et al. [2006] §6, that improvements in the theory could perhaps be found in the direction of relaxing the connectivity hypothesis, by requiring that only the union of the communication graphs be connected.

5. CONCLUSIONS

In this paper we discussed a distributed estimation scheme for sensor networks. The nodes can communicate with their instantaneous neighbors and maintain a common data structure. At each time iteration, a node may collect a new measurement and compute a local estimate of the unknown parameter. We showed in Theorem 1 that *all* local estimates converge asymptotically to the true parameter, even for nodes that collect only a finite number of measurements. Convergence is proved under a necessary condition of convergence of a virtual centralized estimate and under an hypothesis on the frequency of connectivity of the communication graphs. We conjecture that this latter hypothesis can be further relaxed by requiring that only the union of the repeatedly occurring graphs be connected, and current research is being devoted to this purpose.

REFERENCES

I. Akyildiz, W. Su, Y. Sankarasubramniam, and E. Cayirci. A survey on sensor networks. *IEEE Communication Magazine*, 40(8):102–114, 2002.

M. Alanyali, S. Venkatesh, O. Savas, and S. Aeron. Distributed bayesian hypothesis testing in sensor networks. In *Proceedings of the American Control Conference*, pages 5369–5374, 2004.

G.C. Calafiore and F. Abrate. A distributed scheme for least squares estimation over sensor networks. Technical report, 2007.

M. Chu, H. Haussecker, and F. Zhao. Scalable information-deriven sensor querying and routing for ad hoc heterogeneous sensor networks. *International Journal of High-Performance Computing Applications*, 16(3):293–313, 2002.

V. Delouille, R. Neelamani, and R. Baraniuk. Robust distributed estimation in sensor networks using the embedded polygons algorithm. In *Proceedings of The Third International Conference on Information Processing in Sensor Networks*, pages 405–413, 2004.

Z.Q. Luo. An isotropic universal decentralized estimation scheme for a bandwidth constrained ad hoc sensor network. *IEEE Journal on Selected Areas in Communication*, 23(4):735–744, 2005.

S. Martinez and F. Bullo. Optimal sensor placement and motion coordination for target tracking. *Automatica*, 40:661–668, 2006.

R. Olfati-Saber. Consensus problems in networks of agents with switching topology. *IEEE Transactions on Automatic Control*, 49(9), 2004.

M. Alanyali R. Rahman and V. Saligrama. Distributed tracking in multihop sensor networks with communication delays. *IEEE Transactions on Signal Processing*, 55(9):4656–4668, 2007.

D. P. Spanos, R. Olfati-Saber, and R. Murray. Approximate distributed Kalman filtering in sensor networks with quantifiable performance. In *Fourth International Symposium on Information Processing in Sensor Networks (IPSN 2005)*, pages 133–139, 2005a.

D. P. Spanos, R. Olfati-Saber, and R. Murray. Distributed sensor fusion using dynamic consensus. In *Proceedings of IFAC World Congress*, 2005b.

J.N. Tsitsiklis. *Advances in Signal Processing, Vol. 2*, chapter Decentralized Detection, pages 297–344. H. V. Poor and J. B. Thomas, editors, JAI Press, 1993.

L. Xiao and S. Boyd. Fast linear iterations for distributed averaging. *Systems and Control Letters*, 53:65–78, 2004.

L. Xiao, S. Boyd, and S. Lall. A scheme for robust distributed sensor fusion based on average consensus. In *International Conference on Information Processing in Sensor Networks (IPSN 2005)*, pages 63–70, 2005.

L. Xiao, S. Boyd, and S. Lall. A space-time diffusion scheme for peer-to-peer least-squares estimation. In *Proceedings of Fifth International Conference on Information Processing in Sensor Networks (IPSN 2006)*, pages 168–176, 2006.

L. Xiao, S. Boyd, and S. Lall. Distributed average consensus with time-varying Metropolis weights. *Automatica*, submitted.