International Journal of Control
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713393989

Distributed linear estimation over sensor networks

Giuseppe C. Calafiore [a]; Fabrizio Abrate [a]
[a] Dipartimento di Automatica e Informatica, Politecnico di Torino, Torino, TO, Italy

First Published:May2009

To cite this Article Calafiore, Giuseppe C. and Abrate, Fabrizio(2009)'Distributed linear estimation over sensor networks',International Journal of Control,82:5,868 — 882
To link to this Article: DOI: 10.1080/00207170802350662
URL: http://dx.doi.org/10.1080/00207170802350662

International Journal of Control
Vol. 82, No. 5, May 2009, 868–882

Taylor & Francis
Taylor & Francis Group

Distributed linear estimation over sensor networks

Giuseppe C. Calafiore* and Fabrizio Abrate

Dipartimento di Automatica e Informatica, Politecnico di Torino, Corso Duca degli Abruzzi, Torino, TO, Italy

(*Received 14 November 2007; final version received 17 July 2008*)

We consider a network of sensors in which each node may collect noisy linear measurements of some unknown parameter. In this context, we study a distributed consensus diffusion scheme that relies only on bidirectional communication among neighbour 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. The proposed scheme allows one to consider networks with dynamically changing communication topology, and it is robust to unreliable communication links and failures in measuring nodes. We show that under suitable hypotheses all the local estimates converge to the true parameter value.

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. Although each individual sensor node has limited storage capacity and processing power, the network as a whole has the ability to perform complex tasks. These technological achievements have allowed the growth of various applications of sensor networks, mainly in commercial and industrial endeavours, to manage data that would be difficult or expensive to deal with using wired sensors. Typical applications include environmental monitoring, surveillance, object tracking collaborative information processing, traffic monitoring and mobile agents control; see, for instance Akyildiz, Su, Sankarasubramniam, and Cayirci (2002), Chu, Haussecker, and Zhao (2002) and 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 centralised or distributed. In a centralised scheme, each sensor has to send data (directly or by finding a suitable path in the network) to a data fusion centre. This centre is able to compute the best possible estimate of some unknown parameter (e.g. the maximum likelihood (ML) estimate), but high communication load is imposed on the network. Moreover, continued communication induces high energy consumption at the nodes, and the energy budget is often a critical parameter for smart sensors. In a distributed processing scheme, instead, each sensor exchanges data only with its neighbours, and carries out local computation in order to obtain a good estimate of the unknown parameter of interest. Distributed processing has several advantages with respect to centralised processing: there is no central data fusion centre, each sensor can compute the estimates on its own without having any knowledge of the whole network, and communication takes place only among neighbours. Many sophisticated algorithms for distributed estimation and tracking exist; see, for instance, Tsitsiklis (1993), Alanyali, Venkatesh, Savas, and Aeron (2004), Delouille, Neelamani, and Baraniuk (2004), Luo (2005). 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), and Spanos, Olfati-Saber, and Murray (2005a,b), a distributed version of the Kalman filter (DKF) is analysed for distributed estimation of time-varying parameters.

In this article we start from the setup of Xiao, Boyd, and Lall (2006) and analyse a completely distributed consensus diffusion scheme for linear parameter estimation on networks with unreliable links. Each node in the network may take at each

*Corresponding author. Email: giuseppe.calafiore@polito.it

ISSN 0020–7179 print/ISSN 1366–5820 online

DOI: 10.1080/00207170802350662
http://www.informaworld.com

time t a noisy linear measurement of the unknown parameter. The nodes measurement noise covariances are allowed to be time-varying, thus permitting to model, for instance, sensor failures or measurement precision degradation. The network topology may also change with time. We prove that if the *frequency of connectedness* of the superposition of sequences of communication graphs 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); further details in this respect are discussed in Remark 3.1.

The rest of this article is organised as follows. The proposed distributed scheme for parameter estimation is introduced in § 2. Section 3 contains our main convergence results. Numerical examples are presented in § 4, and conclusions are drawn in § 5. The appendix contains some important preliminary material and notation on graphs.

1.1 *Notation*

For a matrix X, X_{ij} denotes the element of X in row i and column j, and $X^\top$ denotes the transpose of X. $X > 0$ (resp. $X \geq 0$) denotes a positive (resp. non-negative) matrix, that is, a matrix with all positive (resp. non-negative) entries. $\|X\|$ denotes the spectral (maximum singular value) norm of X, or the standard Euclidean norm, in case of vectors. For a square matrix $X \in \mathbb{R}^{n,n}$, we denote with $\sigma(X) = \{\lambda_1(X), \ldots, \lambda_n(X)\}$ the set of eigenvalues, or *spectrum*, of X, and with $\rho(X)$ the spectral radius: $\rho(X) \doteq \max_{i=1,\ldots,n} |\lambda_i(X)|$, where $\lambda_i(X)$, $i = 1, \ldots, n$, are the eigenvalues of X. 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. The consensus-based estimation scheme

2.1 *Preliminaries*

Consider n-distributed sensors (nodes), each of which may take 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:

$$\begin{aligned}
&\mathrm{E}\{v_i(t)\} = 0,\\
&\mathrm{var}\{v_i(t)\} \doteq \mathrm{E}\{(v_i(t)\} - \mathrm{E}\{v_i(t)\})(v_i(t) - \mathrm{E}\{v_i(t)\})^\top \doteq \Sigma_i(t),\\
&\mathrm{E}\{(v_i(t) - \mathrm{E}\{v_i(t)\})(v_j(\tau) - \mathrm{E}\{v_j(\tau)\})^\top\} = 0\\
&\quad \text{whenever } i \neq j \text{ or } t \neq \tau.
\end{aligned}$$

Allowing the covariance matrices to be time-varying helps when modelling 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 technical characteristics of the i-th sensor. Instead, if 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 centralised information were available, the optimal ML estimate $\hat{\theta}_{\mathrm{ml}}$ of the parameter θ could be obtained. Defining the quantities

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

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:

$$Q_{\mathrm{ml}}(t) \doteq {P^{-1}}_{\mathrm{ml}}(t). \tag{2}$$

However, we assume it is not possible (due to communication constraints, etc.) to construct the optimal centralised 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 § 2.3 that under suitable hypotheses, all local estimates converge asymptotically to the true parameter θ, in mean square sense.

We describe the communication structure among nodes using graph formalism (see the appendix for notation and preliminary results on graphs). 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 $(i, j) \in \mathcal{E}(t)$. 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 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:

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

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

$$\begin{aligned}\tilde{W}_{ij}(t) &= \min\left(1/d_i^{ST}(t), 1/d_j^{ST}(t)\right) \\ &= \frac{1}{\max\left(d_i^{ST}(t), d_j^{ST}(t)\right)} \\ &= \frac{1}{1+\max\left(|\mathcal{N}_i(t)|, |\mathcal{N}_j(t)|\right)} \cdot \frac{1}{t+1}, \\ &\quad \text{for } (i,j)\in\mathcal{E}(t),\ i\neq j. \end{aligned} \tag{3}$$

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

2.2 *Distributed space–time diffusion scheme*

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

$$P_i(t_+) = \frac{t}{t+1} P_i(t) + \frac{1}{t+1} A_i^\top(t)\Sigma_i^{-1}(t)A_i(t) \tag{4}$$

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

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. Notice also that the initial values $P_i(0)$, $q_i(0)$ are irrelevant.

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 neighbours. At this point, a spatial update phase is performed. Considering (3) and defining

$$W_{ij}(t) \doteq \begin{cases} (t+1)\tilde{W}_{ij}(t) & \text{if } (i,j)\in\mathcal{E}(t) \\ 1 - \sum_{j\in\mathcal{N}_i(t)} W_{ij}(t) & \text{if } i=j \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

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

$$P_i(t+1) = P_i(t_+) + \sum_{j\in\mathcal{N}_i(t)} W_{ij}(t)\left(P_j(t_+) - P_i(t_+)\right) \tag{7}$$

$$q_i(t+1) = q_i(t_+) + \sum_{j\in\mathcal{N}_i(t)} W_{ij}(t)\left(q_j(t_+) - q_i(t_+)\right). \tag{8}$$

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

Proposition 2.1: *For $t = 1, 2, \ldots$ 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

$$\Phi(t-1;k) \doteq W(t-1)\cdots W(k). \tag{9}$$

A proof for Proposition 2.1 is given in § A.5.

Remark 2.1: Notice that the recursions in (7) and (8) 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 neighbours. 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 2.1, which are useful for *a posteriori* analysis, do *not* describe the actual computations performed by the nodes, which instead use the recursions (7) and (8).

Remark 2.2 (Measurement and consensus time scales)**:** In some practical cases, it may happen that communication occurs at more frequent intervals than observations, or vice versa, and this would lead to a distinction between the measurement time scale and the distributed averaging one. For notational simplicity, in this work we use a single time

scale for both the averaging and the observation processes. However, this is done without loss of generality, due to the flexibility introduced by the time-varying nature of the process parameters. More precisely, we assume that $t=0,1,\ldots,$ represent the time indices at which either a measurement occurs at some node, and/or a consensus averaging step should be performed through the network. If some sensor i does not have a valid measurement at time t, we just set $\Sigma_i^{-1}(t)=0$, thus covering the situation when communication occurs more frequently than observations. Vice versa, if some sensor takes a measurement at t but no actual consensus iteration should be performed at that time, we simply set the consensus weight matrix equal to the identity, i.e. $W(t)=I$ (this means that nodes are only connected with themselves at these specific time instants), thus covering the situation when measurements occur more frequently than consensus steps. Note further that in any time interval $[T, T+\tau]$ in which measurements persistently occur without averaging, the algorithm evolves according to (4) and (5), with (7) and (8) simply reduced to $P_i(t+1)=P_i(t_+)$, $q_i(t+1)=q_i(t_+)$. It can be readily checked that in this case the algorithm yields optimal ML estimates at each observation step.

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

2.3 *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 2.2: *The local estimate $\hat{\theta}_i(t)$ is an unbiased estimator of θ, that is*

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

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

$$\begin{aligned} Q_i(t) \doteq \mathrm{var}\{\hat{\theta}_i(t)\} &= \mathrm{E}\left\{(\hat{\theta}_i(t)-\theta)(\hat{\theta}_i(t)-\theta)^\top\right\} \\ &= \frac{1}{t^2}P_i^{-1}(t)\left(\sum_{k=0}^{t-1}\sum_{j=1}^{n}\Phi_{ij}^2(t-1;k)\right. \\ &\quad \left.\times A_j^\top(k)\Sigma_j^{-1}(k)A_j(k)\right)P_i^{-1}(t). \end{aligned} \tag{10}$$

Proof: A proof of the above statement is obtained by direct computation. First, notice that

$$\begin{aligned} \hat{\theta}_i(t) &= P_i^{-1}(t)q_i(t) \\ &= P_i^{-1}(t)\frac{1}{t}\sum_{k=0}^{t-1}\sum_{j=1}^{n}\Phi_{ij}(t-1;k)A_j^\top(k)\Sigma_j^{-1}(k)y_j(k) \\ &= P_i^{-1}(t)\frac{1}{t}\sum_{k=0}^{t-1}\sum_{j=1}^{n}\Phi_{ij}(t-1;k)A_j^\top(k) \\ &\quad \times \Sigma_j^{-1}(k)(A_j(k)\theta+v_j(k)) \\ &= P_i^{-1}(t)\frac{1}{t}\sum_{k=0}^{t-1}\sum_{j=1}^{n}\Phi_{ij}(t-1;k)A_j^\top(k)\Sigma_j^{-1}(k)A_j(k)\theta \\ &\quad + P_i^{-1}(t)\frac{1}{t}\sum_{k=0}^{t-1}\sum_{j=1}^{n}\Phi_{ij}(t-1;k)A_j^\top(k)\Sigma_j^{-1}(k)v_j(k) \\ &= \theta + P_i^{-1}(t)\frac{1}{t}\sum_{k=0}^{t-1}\sum_{j=1}^{n}\Phi_{ij}(t-1;k)A_j^\top(k)\Sigma_j^{-1}(k)v_j(k). \end{aligned} \tag{11}$$

Since $\mathrm{E}\{v_j(k)\}=0$, we have that $\mathrm{E}\{\hat{\theta}_i(t)\}=\theta$, that is, all local estimates are unbiased. Next, notice that whenever ξ_i are independent random vectors and $z=\sum_i K_i\xi_i$, then

$$\mathrm{var}\{z\}=\sum_i K_i\mathrm{var}\{\xi_i\}K_i^\top.$$

Applying this rule to (11) we obtain the covariance of the local estimate:

$$\begin{aligned} Q_i(t) \doteq \mathrm{var}\{\hat{\theta}_i(t)\} &= \mathrm{E}\left\{(\hat{\theta}_i(t)-\theta)(\hat{\theta}_i(t)-\theta)^\top\right\} \\ &= \frac{1}{t^2}P_i^{-1}(t)\left(\sum_{k=0}^{t-1}\sum_{j=1}^{n}\Phi_{ij}^2(t-1;k)A_j^\top(k)\right. \\ &\quad \left.\times \Sigma_j^{-1}(k)\mathrm{E}\{v_j(k)v_j^\top(k)\}\Sigma_j^{-1}(k)A_j(k)\right)P_i^{-1}(t) \\ &= \frac{1}{t^2}P_i^{-1}(t)\left(\sum_{k=0}^{t-1}\sum_{j=1}^{n}\Phi_{ij}^2(t-1;k)\right. \\ &\quad \left.\times A_j^\top(k)\Sigma_j^{-1}(k)A_j(k)\right)P_i^{-1}(t). \end{aligned}$$

□

Note from (10) that if we actually want to compute numerically the local covariance $Q_i(t)$, we need to know the occurred sequence of graphs, since this is needed for constructing the entries of $\Phi(t-1;k)$. Notice also that we can upper bound the local

covariance as detailed in Lemma 2.1 below. To this end, define

$$\begin{aligned} \bar{P}_i(t) &\doteq tP_i(t) = \sum_{k=0}^{t-1}\sum_{j=1}^{n} \Phi_{ij}(t-1;k)H_j(k) \\ H_j(k) &\doteq A_j^\top(k)\Sigma_j^{-1}(k)A_j(k). \end{aligned} \tag{12}$$

The following result holds:

Lemma 2.1: *Whenever $\bar{P}_i^{-1}(t)$ is invertible, the covariance matrix of the i-th local estimate satisfies*:

$$Q_i(t) \preceq \bar{P}_i^{-1}(t). \tag{13}$$

Proof: Consider (10) and notice that since $\Phi_{ij}(t-1;k) \in [0,1]$, then $\Phi_{ij}^2(t-1;k) \leq \Phi_{ij}(t-1;k)$. Hence it follows that

$$\begin{aligned} Q_i(t) &= \bar{P}_i^{-1}(t)\left(\sum_{k=0}^{t-1}\sum_{j=1}^{n} \Phi_{ij}^2(t-1;k)H_j(k)\right)\bar{P}_i^{-1}(t) \\ &\preceq \bar{P}_i^{-1}(t)\left(\sum_{k=0}^{t-1}\sum_{j=1}^{n} \Phi_{ij}(t-1;k)H_j(k)\right)\bar{P}_i^{-1}(t) \\ &= \bar{P}_i^{-1}(t), \end{aligned}$$

which proves the statement. □

3. Mean square convergence results

We show in this section that, under suitable hypotheses, 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 $i = 1,\ldots,n$. Notice that this is not obvious since we allow for the existence in the network of nodes that do not collect an infinite number of measurement, or that even do not collect measurements at all, hence the law of large numbers cannot be trivially applied. The convergence 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 centralised ML estimate mean square error goes to zero as $t \to \infty$. This condition is actually necessary, since one cannot hope to make the local estimates converge when even the centralised optimal estimate (which ideally has all the available information) does not converge. The second condition is a technical condition needed for proving convergence of the distributed scheme, and it is detailed in the next section. Loosely speaking, this condition requires that the time-varying communication graphs form, at least 'rarely' in time, subsequences whose union graph is connected, see Appendix A.1 for a definition of graph union. Subsequences such that the union of the graphs in the subsequence forms a connected graph are here called *jointly connected*.

We first state some technical preliminaries. The main theorem is stated in § 3.1. The next paragraphs also require some introductory results and notation that are reported in the appendix.

We start by looking more closely at the structure of the $W(t)$ matrices in (6) and of the transition matrices $\Phi(t-1;k)$ in (9). First, notice that $W(t)$ is non-negative, symmetric and $W_{ij}(t) > 0$ for $(i,j) \in \mathcal{E}(t)$. Moreover, the diagonal entries of $W(t)$ are strictly positive,[1] and the sum over each row or column of $W(t)$ is equal to one ($W(t)\mathbf{1} = \mathbf{1}$, $\mathbf{1}^\top W(t) = \mathbf{1}^\top$). This means that $W(t)$ is a symmetric and doubly stochastic matrix belonging to the set $\mathcal{M}_{\rm ss}$ defined in (25) in the appendix, and that $W(t)$ is *compatible* with graph $\mathcal{G}(t)$, in the sense of Definition A.1. Further note that the set of all possible $W(t)$ generated by the time-varying graphs is *finite*, since the set of Metropolis weights one can obtain from a fixed number of nodes is of finite cardinality.

Using the notation (29)–(31) in the appendix, we write $W(t)$ in the form

$$W(t) = \frac{1}{n}\mathbf{1}\mathbf{1}^\top + Z(t), \quad Z(t) = V(t)D(t)V^\top(t), \tag{14}$$

where $V(t) \in \mathbb{R}^{n,n-1}$ is such that $V^\top(t)V(t) = I_{n-1}$, $V(t)V^\top(t) = I_n - (1/n)\mathbf{1}\mathbf{1}^\top$, $\mathbf{1}^\top V(t) = 0$, 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 non-increasing modulus. Define

$$\Phi(t-1;k) \doteq W(t-1)W(t-2)\cdots W(k), \tag{15}$$

$$\Upsilon(t-1;k) \doteq Z(t-1)Z(t-2)\cdots Z(k), \tag{16}$$

where it holds that

$$\Phi(t-1;k) = \frac{1}{n}\mathbf{1}\mathbf{1}^\top + \Upsilon(t-1;k). \tag{17}$$

3.1 *Main result*

In order to prove our main result on convergence of $\|Q_i(t)\|$ we need to impose an assumption on the connectivity properties of the graph sequence. In its essence, this assumption just requires that any sequence of consecutive graphs of length $k > \bar{k}$ (for some $\bar{k} > 0$) contains a suitable number $N(k)$ of jointly connected subsequences, that is subsequences that contain graphs whose union is connected. Formally, we state the following property:

Definition 3.1 (RJC property)**:** An ordered set of graphs $S = \{\mathcal{G}_k(\mathcal{V},\mathcal{E}_k),\ k = 0,1,\ldots\}$ is said to be repeatedly jointly connected (RJC) if there exist finite

integers $\bar{k} \geq 0$, $M \geq 0$ such that any ordered sequence from S of length $k > \bar{k}$ contains $N(k) > 0$ subsequences of length no larger than M that are jointly connected, where $N(k) : \mathbb{N} \to \mathbb{N}$ is a non-decreasing function such that $N(k) \to \infty$ for $k \to \infty$.

Notice that the RJC property does not require connectivity at any given time. It only requires that, for sufficiently large k, any ordered sequence of k graphs contains $N(k)$ subsequences such that the *union* of graphs in each of these subsequences is connected. The number $N(k)$ of such subsequences is left unspecified for the time being. However, we shall prove shortly that this number need only increase slowly with the logarithm of k. This means, in turn, that as k grows we shall require the existence of only few graph subsequences whose union is connected, an assumption that appears to be very mild in practice. We preliminarily state the following technical lemma:

Lemma 3.1: *Let $S = \{\mathcal{G}_k(\mathcal{V}, \mathcal{E}_k),\ k = 0, 1, \ldots\}$ be an ordered set of graphs having the RJC property, and let $\mathcal{R} = \{R_k,\ k = 0, 1, \ldots\}$ be a corresponding set of compatible matrices, such that for all k, $R_k \in \mathcal{M}_{ss} \cap \mathcal{R}$, where $\mathcal{R}$ is a set of finite cardinality.*

For $k > \bar{k}$, $t \geq 1$, let $S(t-k, t-1) = \{\mathcal{G}_{t-k}, \ldots, \mathcal{G}_{t-1}\}$ be an ordered sequence of k graphs from S, and let $\{R_{t-k}, \ldots, R_{t-1}\}$ be a corresponding sequence of compatible matrices from $\mathcal{R}$. Define $Z_{t-\tau} = R_{t-\tau} - (1/n)\mathbf{1}\mathbf{1}^\top$. Let further $[s_1, e_1], [s_2, e_2], \ldots, [s_{N(k)}, e_{N(k)}]$, $s_1 \geq t-k$, $e_{N(k)} \leq t-1$, denote the indices delimiting the $N(k)$ subsequences of $S(t-k, t-1)$ that are jointly connected. Then there exist a $\lambda < 1$ such that:

$$\|Z_{e_i} Z_{e_i - 1} \cdots Z_{s_i}\| \leq \lambda, \quad i = 1, \ldots, N(k); \quad \text{for all } t \geq 1,\ k > \bar{k}. \tag{18}$$

See Appendix A.3 for a proof of Lemma 3.1. Notice that the sequence of Metropolis weight matrices $R_t = W(t)$ are compatible with the corresponding communication graphs $\mathcal{G}_t$ and belong to a set of finite cardinality, therefore Lemma 3.1 applies in particular to the sequence $\{Z(t-k), \ldots, Z(t-1)\}$ of matrices of the form (14).

Definition 3.2 (Joint connectivity index)**:** Let S and $\mathcal{R}$ be defined as in Lemma 3.1. The joint connectivity index of S with respect to $\mathcal{R}$ is defined as:

$$\bar{\lambda} \doteq \min \lambda \text{ such that (18) holds.} \tag{19}$$

Let us briefly discuss the meaning of the result in Lemma 3.1 and of Definition 3.2. We know that if the graph set S is RJC then for sufficiently large k every sequence of length k contains $N(k)$ finite subsequences that are jointly connected. We are interested in the products of R_k matrices (which are actually the Metropolis weight matrices $W(k)$ in our specific application) corresponding to the graphs in each of these jointly connected subsequences. More specifically we are interested in the products of the related Z_k matrices, and Lemma 3.1 states that the norm of any such product is upper bounded by a quantity that is strictly less than one. The joint connectivity index in Definition 3.2 is simply the smallest of these upper bounds.

We now define a specific class of lower-bound functions for $N(k)$.

Definition 3.3 (Log-RJC property)**:** Let $S = \{\mathcal{G}_k(\mathcal{V}, \mathcal{E}_k), k = 0, 1, \ldots\}$ be an ordered set of graphs having the RJC property. Let $\mathcal{R} = \{R_k, k = 0, 1, \ldots\}$ be a corresponding set of compatible matrices, such that for all k, $R_k \in \mathcal{M}_{ss} \cap \mathcal{R}$, where $\mathcal{R}$ is a set of finite cardinality, and let $\bar{\lambda} < 1$ be the joint connectivity index of S with respect to $\mathcal{R}$.

Then, S is said to be logarithmical repeatedly jointly connected (log-RJC) with index $\bar{\lambda}$ with respect to $\mathcal{R}$, if there exist a constant $\alpha > 0$ and a finite integer $\bar{k} \geq \alpha$ such that any ordered sequence from S of length $k > \bar{k}$ contains $N(k) \geq N_{lb}(k)$ subsequences that are jointly connected, where

$$N_{lb}(k) \doteq \left\lceil \max\left(c \log\left(\frac{k}{\alpha}\right), 0 \right) \right\rceil, \quad k = 1, 2, \ldots, \qquad c \doteq \frac{2}{\log(1/\bar{\lambda})}, \tag{20}$$

and where $\lceil x \rceil$ denotes the smallest integer larger than or equal to x.

Notice that $N_{lb}(k) = 0$ for $k \leq \alpha$, whereas $N_{lb}(k)$ grows at sub-linear (specifically logarithmic) rate for $k > \alpha$. Figure 1 shows an example of plot of $N_{lb}(k)$.

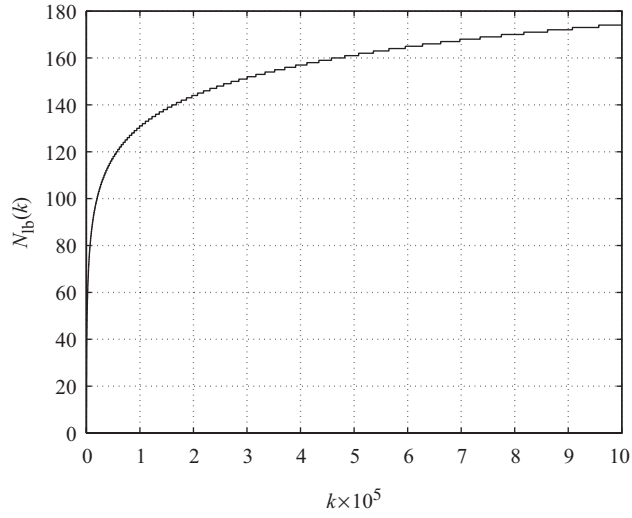

Figure 1. Plot of $N_{lb}(k)$, for $\bar{\lambda} = 0.9$, $\alpha = 100$.

The following key technical result holds, see Appendix A.4 for a proof.

Lemma 3.2: *Let* $S=\{\mathcal{G}_k(\mathcal{V},\mathcal{E}_k), k=0,1,\ldots\}$ *be an ordered set of graphs having the log-RJC property with respect to a compatible sequence of matrices* $\{R_k\in\mathcal{M}_{ss}\cap\mathcal{R}, k=0,1,\ldots\}$, *where* $\mathcal{R}$ *is a set of finite cardinality. Define* $Z_k=R_k-(1/n)\mathbf{1}\mathbf{1}^\top$, $k=0,1,\ldots$. *Then*,

$$\lim_{t\to\infty}\sum_{k=1}^{t}\|Z_{t-1}Z_{t-2}\cdots Z_{t-k}\|=\text{constant}<\infty.$$

We can now state the main result of this section in the following theorem:

Theorem 3.1: *Let the occurring communication graph sequence* $S=\{\mathcal{G}_k(\mathcal{V},\mathcal{E}_k),\ k=0,1,\ldots\}$ *have the log-RJC property with respect to the compatible sequence of Metropolis weight matrices* $\{W(k), k=0,1,\ldots\}$, *and let* $\|H_j(k)\|\le C$, *for all* $j=1,\ldots,n$, $k=0,1,\ldots$

If $\lim_{t\to\infty}\|P_{\mathrm{ml}}(t)\|=\infty$ (*or*, *equivalently*, *if the centralised ML 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 (12), and substitute (17) to obtain

$$\begin{aligned}\bar{P}_i(t)&=\sum_{k=0}^{t-1}\sum_{j=1}^{n}\frac{1}{n}H_j(k)+\sum_{k=0}^{t-1}\sum_{j=1}^{n}\Upsilon_{ij}(t-1;k)H_j(k)\\&=\frac{1}{n}P_{\mathrm{ml}}(t)+\sum_{k=0}^{t-1}\sum_{j=1}^{n}\Upsilon_{ij}(t-1;k)H_j(k).\end{aligned}\tag{21}$$

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 $\|A+B\|\ge\|A\|-\|B\|$. Applying this inequality to (21), and taking the spectral norm, we have

$$\|\bar{P}_i(t)\|\ge\frac{1}{n}\|P_{\mathrm{ml}}(t)\|-\left\|\sum_{k=0}^{t-1}\sum_{j=1}^{n}\Upsilon_{ij}(t-1;k)H_j(k)\right\|.\tag{22}$$

Further notice that

$$\begin{aligned}&\left\|\sum_{k=0}^{t-1}\sum_{j=1}^{n}\Upsilon_{ij}(t-1;k)H_j(k)\right\|\\&\quad\le\sum_{k=0}^{t-1}\sum_{j=1}^{n}|\Upsilon_{ij}(t-1;k)|\cdot\|H_j(k)\|\\&\quad\le C\sum_{k=0}^{t-1}\sum_{j=1}^{n}|\Upsilon_{ij}(t-1;k)|\\&\quad\text{[from }\|\Upsilon\|_\infty\le\sqrt{n}\|\Upsilon\|\text{, see § 5.6 of}\\&\qquad\text{Horn and Johnson (1985)]}\\&\quad\le\sqrt{n}C\sum_{k=0}^{t-1}\|\Upsilon(t-1;k)\|\\&\quad\text{[reversing the summation]}\\&\quad=\sqrt{n}C\sum_{k=1}^{t}\|\Upsilon(t-1;t-k)\|.\end{aligned}$$

Going back to (22), we hence obtain that

$$\begin{aligned}\|\bar{P}_i(t)\|&\ge\frac{1}{n}\|P_{\mathrm{ml}}(t)\|-\left\|\sum_{k=0}^{t-1}\sum_{j=1}^{n}\Upsilon_{ij}(t-1;k)H_j(k)\right\|\\&\ge\frac{1}{n}\|P_{\mathrm{ml}}(t)\|-\sqrt{n}C\sum_{k=1}^{t}\|\Upsilon(t-1;t-k)\|.\end{aligned}$$

Recalling (16) and applying Lemma 3.2 we now have that there exist $K<\infty$ such that

$$\lim_{t\to\infty}\sum_{k=1}^{t}\|\Upsilon(t-1;t-k)\|=K,$$

therefore

$$\lim_{t\to\infty}\|\bar{P}_i(t)\|\ge\frac{1}{n}\lim_{t\to\infty}\|P_{\mathrm{ml}}(t)\|-\sqrt{n}CK.$$

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

$$\lim_{t\to\infty}\|\bar{P}_i(t)\|=\lim_{t\to\infty}\|P_{\mathrm{ml}}(t)\|=\infty.$$

Finally, from (13) it follows that whenever $\bar{P}_i(t)$ is invertible

$$\|Q_i(t)\|\le\|\bar{P}_i^{-1}(t)\|=\frac{1}{\|\bar{P}_i(t)\|},$$

hence

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

which concludes the proof. □

Remark 3.1: A few remarks are in order with respect to the result in Theorem 3.1. 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. These hypotheses basically require that the graph process forms at least seldom in

time subsequences whose union is connected, and that the total information collectively gathered by all nodes is sufficient to make an hypothetical centralised ML estimate converge to the 'true' parameter value as $t \to \infty$. Our hypothesis of joint graph connectivity is consistent with similar hypotheses that appeared in the literature on consensus, formation and agreement problems, see, e.g. Jadbabaie, Lin, and Morse (2003), Roy and Saberi (2005), Roy, Saberi, and Herlugson (2006).

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 $t \to \infty$; on the 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' (consensus) iterations only. This in turn permits the authors of Xiao et al. (2006) to apply standard tools for convergence of products of stochastic matrices, see Xiao et al. (2006) and the references therein. These results cannot be directly applied to our setup, due to the persistent presence of new measurements, which act as a forcing term in the local iterations (4), (5). In some sense, Xiao et al. (2006) deals with convergence of a particular type of homogeneous (unforced) system, whereas we here deal with convergence of a system persistently 'forced' by external measurements. As a matter of fact, §6 of Xiao et al. (2006) also contains an extension to the case of infinite measurements, which is stated in Theorem 4 of this reference for a special scalar case. On the one hand this result does not require assumptions on graph connectivity, but on the other hand it guarantees convergence only for some of the nodes that indeed collect an infinite number of measurements as $t \to \infty$. Since such nodes, left alone, would be able to build estimates that converge to the true parameter value, the mentioned result shows that the diffusion scheme does not worsen the situation with respect to a case when no communication exchange is made among nodes, but does not actually prove that some benefit is taken from communication. On the contrary, the convergence result stated in Theorem 3.1 of the present manuscript does require an hypothesis on graph connectivity (the log-RJC condition), but then guarantees convergence for all the nodes of the network, both those that collect an infinite as well as a finite number of measurements.

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 with fixed topology, for three different scenarios with increasing sensor measurement rate. The second example shows the convergence of the proposed distributed scheme in a time-varying topology setting with a ring network structure.

4.1 *Example 1*

We considered a network with fixed topology constructed by drawing $n = 50$ nodes at random on the unit square $[0, 1] \times [0, 1]$, and assuming that any two nodes can communicate whenever their distance is less than 0.25. The communication graph of the network (shown in Figure 2) resulted in 184 edges and was connected.

We consider a vector θ of unknown parameters with dimension $m = 5$. 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 $\mathbb{R}^5$. When the sensor takes a measurement its measurement noise v_i is i.i.d. Gaussian with unit variance. We model the possibility of intermittent measurements by assuming that each sensor takes a valid measurement every $p \geq 1$ iterations. As discussed in Remark 2.2, this situation is captured by assuming that the i-th sensor covariance is such that

$$\Sigma_i^{-1}(t) = \begin{cases} I & \text{if } (t \bmod p) = 0 \\ 0 & \text{otherwise.} \end{cases}$$

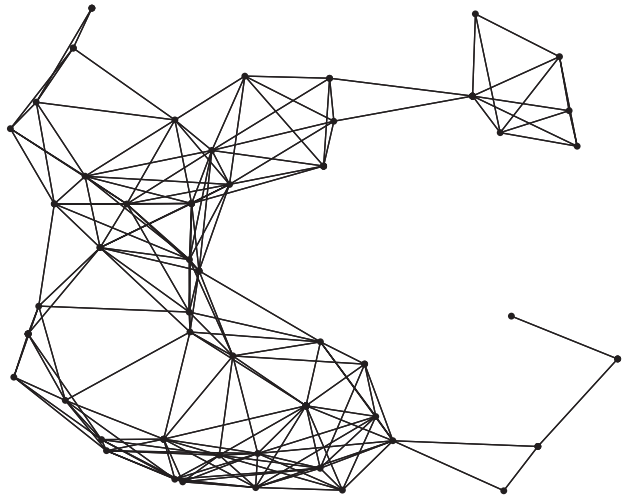

Figure 2. Network used in Example 1.

To quantify the estimation performances, we define an average index of the local mean square estimation errors:

$$\mathbf{MSE}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Tr}(Q_i(t)),$$

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

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

Three experiments have been carried out, with measurement rates $1/p = 0.01$, $1/p = 0.1$, $1/p = 1$, see Figures 3–5. In case $p = 1$, measurements and consensus iterations happen at the same rate, whereas in cases $p = 10$, $p = 100$, consensus iterations are more frequent than measurement iterations, hence in the instants among measurements the algorithm performs 'consensus-only' steps, leading the local estimates to approach the current ML estimate. This effect is more evident as p increases, see Figures 3 and 4. Since more information is globally gathered by the network as p decreases, we observe as expected that the final local estimation error decreases with p.

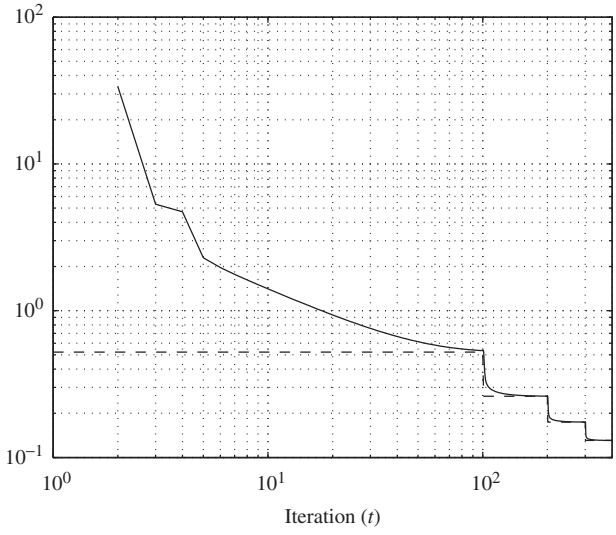

Figure 3. First experiment $p = 100$: **MLE**(t) (dashed), **MSE**(t) (solid).

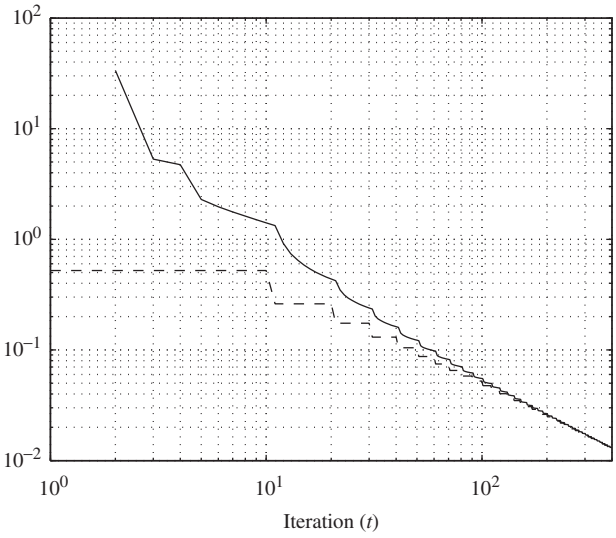

Figure 4. Second experiment $p = 10$: **MLE**(t) (dashed), **MSE**(t) (solid).

4.2 *Example 2*

In a second example, we considered a network with ring structure and time-varying topology, in two cases with $n = 3$ and $n = 6$ nodes. Specifically, we assumed that at each time $t = 0, 1, \ldots,$ only two sensors are able to communicate and collect measurements. We denote the two sensors that are active at t with $s_1(t)$ and $s_2(t)$, respectively, where these indices are defined as:

$$s_1(t) = [t \bmod n] + 1$$
$$s_2(t) = [(t+1) \bmod n] + 1.$$

Figure 6 shows the time varying graph topology for $n = 6$.

In this situation, the Metropolis weight matrix is written as

$$W(t) = I + \frac{1}{2}e_{s_1(t)}e_{s_2(t)}^{\top} + \frac{1}{2}e_{s_2(t)}e_{s_1(t)}^{\top} - \frac{1}{2}e_{s_1(t)}e_{s_1(t)}^{\top} - \frac{1}{2}e_{s_2(t)}e_{s_2(t)}^{\top},$$

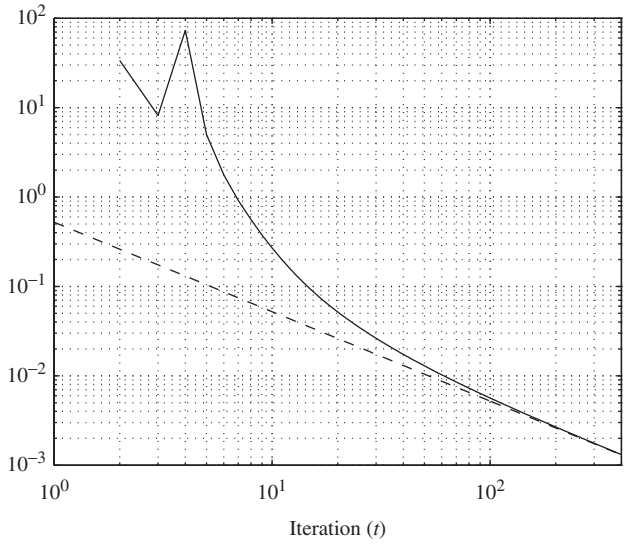

Figure 5. Third experiment $p = 1$: **MLE**(t) (dashed), **MSE**(t) (solid).

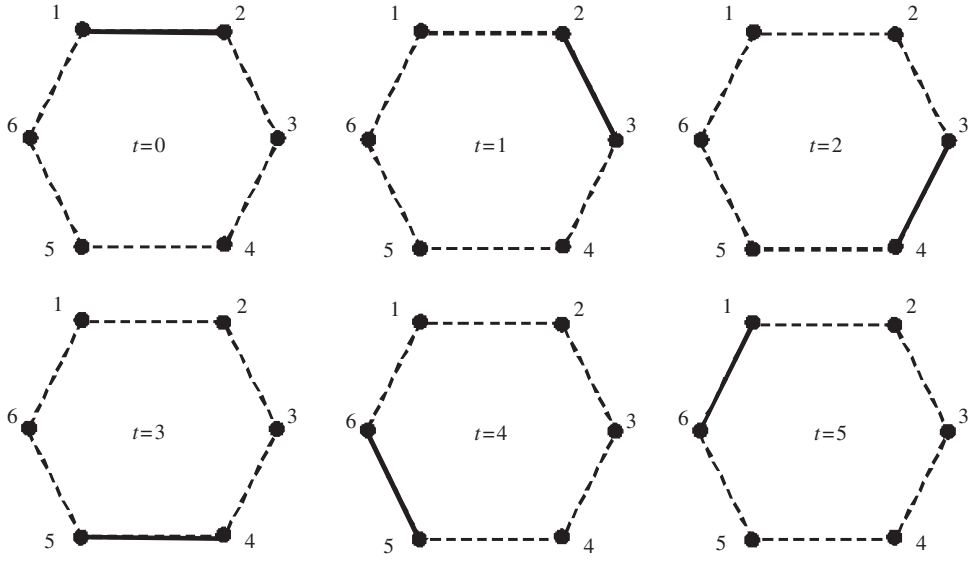

Figure 6. Time-varying ring network with six nodes.

where $e_i \in \mathbb{R}^n$ is a vector having all entries equal to zero except for the i-th position, which is equal to one. The vector of unknown parameters has dimension $m = 2$, and the scalar measurement equation for node i is

$$y_i = a_i^\top \theta + v_i$$

with $a_i = [1\ 0]$ if i is odd, and $a_i = [0\ 1]$ otherwise. The inverse measurement noise covariance is

$$\Sigma_i^{-1}(t) = \begin{cases} 1 & \text{if } i \in \{s_1(t), s_2(t)\} \\ 0 & \text{otherwise.} \end{cases}$$

Notice that the communication graph is not connected at any time instant. However, each sequence of graphs of length k has $\lfloor k/n \rfloor$ jointly connected subsequences. Therefore, the frequency of joint connectedness grows linearly with the sequence length, hence our convergence assumptions that require logarithmic growth are largely satisfied. Theorem 3.1 thus guarantees that all local estimates converge asymptotically to the true parameter value. This is indeed confirmed by the plots resulting from numerical simulations, shown in Figures 7 and 8, where it can also be observed as expected that the rate of convergence of the local estimates decreases as the number of nodes increases.

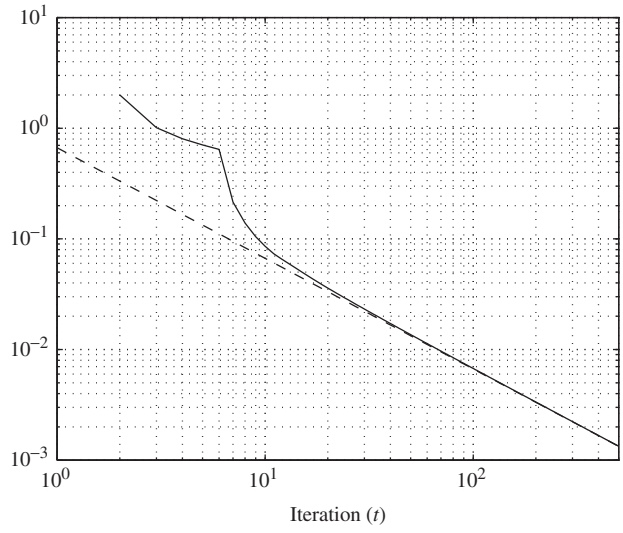

Figure 7. Ring with six nodes. **MLE**(t) (dashed), and **MSE**(t) (solid).

5. Conclusions and future work

In this article, we discussed a distributed estimation scheme for sensor networks. The nodes maintain a common data structure and can communicate with their instantaneous neighbours. At each time iteration, a node may collect a new measurement, compute a local estimate of the unknown parameter and then average its local information with the neighbours' information. We showed in Theorem 3.1 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 centralised estimate and under a rather mild hypothesis on the frequency of connectivity for the superposition (union) of subsequences of occurring communication graphs. It is worth underlining that our results on the convergence of 'consensus under persistent excitation' (the 'persistent excitation' is in our case due to the

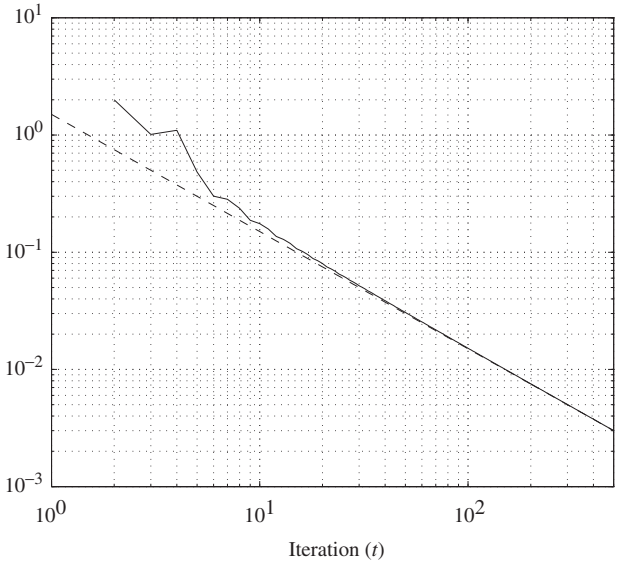

Figure 8. Ring with three nodes. **MLE**(t) (dashed), and **MSE**(t) (solid).

possible continued presence of new measurements) required an analysis somewhat different from the one usually found in the consensus literature (Jadbabaie et al. 2003; Moreau 2005; Xiao et al. 2006; Cao, Morse, and Anderson 2008). More precisely, the key technical tool required in the classical consensus approach is the convergence of certain infinite matrix products, whereas under persistent excitation we also need that the infinite sum of such products remains finite (Lemma 3.2). We expect, in turn, that the approach developed here in the context of estimation problems could be exported to other agreement-type problems such as those arising in decentralised coordination and control (Jadbabaie et al. 2003; Moreau 2005; Roy and Saberi 2005; Roy et al. 2006).

Acknowledgements

This work was supported by the MACP4Log grant (RU/02/26) from Regione Piemonte, Italy.

Notes

1. This follows from the definition (6), since $\sum_{j\in\mathcal{N}_i(t)} W_{ij}(t) = \sum_{j\in\mathcal{N}_i(t)}(1+\max\{|\mathcal{N}_i(t)|,|\mathcal{N}_j(t)|\})^{-1} \leq \sum_{j\in\mathcal{N}_i(t)}(1+|\mathcal{N}_i(t)|)^{-1} = |\mathcal{N}_i(t)|/(1+|\mathcal{N}_i(t)|) < 1$.
2. To see why this is true, consider $(i,j)\in\mathcal{E}_{a|b}$, which means that (i,j) is either in $\mathcal{E}_a$ or in $\mathcal{E}_b$, and suppose without loss of generality $(i,j)\in\mathcal{E}_a$. Then choosing $k=j$ we have that $(i,k)\in\mathcal{E}_a$ and $(k,j)\in\mathcal{E}_b$, which means indeed that $(i,j)\in\mathcal{E}_{ab}$. The second inclusion follows from an analogous reasoning, by taking $k=i$.

References

Akyildiz, I., Su, W., Sankarasubramniam, Y., and Cayirci, E. (2002), 'A Survey on Sensor Networks', *IEEE Communication Magazine*, 40, 102–114.

Alanyali, M., Venkatesh, S., Savas, O., and Aeron, S. (2004), 'Distributed Bayesian Hypothesis Testing in Sensor Networks', in *Proceedings of the American Control Conference*, pp. 5369–5374.

Cao, M., Morse, A., and Anderson, B. (2008), 'Reaching a Consensus in a Dynamically Changing Environment: A Graphical Approach', *SIAM Journal on Control and Optimization*, 47, 575–600.

Chu, M., Haussecker, H., and Zhao, F. (2002), 'Scalable Information-driven Sensor Querying and Routing for *ad hoc* Heterogeneous Sensor Networks', *International Journal of High-Performance Computing Applications*, 16, 293–313.

Delouille, V., Neelamani, R., and Baraniuk, R. (2004), 'Robust Distributed Estimation in Sensor Networks Using the Embedded Polygons Algorithm', in *Proceedings of The Third International Conference on Information Processing in Sensor Networks*, pp. 405–413.

Hartfiel, D. (2002), *Nonhomogeneous Matrix Products*, Singapore: World Scientific; 2002.

Horn, R.A., and Johnson, C.R. (1985), *Matrix Analysis*, Cambridge: Cambridge University Press; 1985.

Jadbabaie, A., Lin, J., and Morse, A. (2003), 'Coordination of Groups of Mobile Autonomous Agents Using Nearest Neighbour Rules', *IEEE Transactions on Automatic Control*, 48, 988–1001.

Luo, Z. (2005), 'An Isotropic Universal Decentralised Estimation Scheme for a Bandwidth Constrained *ad hoc* Sensor Network', *IEEE Journal on Selected Areas in Communication*, 23, 735–744.

Martinez, S., and Bullo, F. (2006), 'Optimal Sensor Placement and Motion Coordination For Target Tracking', *Automatica*, 42, 661–668.

Moreau, L. (2005), 'Stability of Multiagent Systems with Time-dependent Communication Links', *IEEE Transactions on Automatic Control*, 50, 169–182.

Olfati-Saber, R. (2004), 'Consensus Problems in Networks of Agents with Switching Topology', *IEEE Transactions on Automatic Control*, 49, 1520–1533.

Roy, S., and Saberi, A. (2005), 'Static Decentralised Control of a Single-integrator Network with Markovian Sensing Topology', *Automatica*, 41, 1867–1877.

Roy, S., Saberi, A., and Herlugson, K. (2006), 'A Control-theoretic Perspective on the Design of Distributed Agreement Protocols', *International Journal of Robust and Nonlinear Control*, 17, 1034–1066.

Spanos, D.P., Olfati-Saber, R., and Murray, R. (2005a), 'Approximate Distributed Kalman Filtering in Sensor Networks with Quantifiable Performance', in *Fourth International Symposium on Information Processing in Sensor Networks (IPSN 2005)*, pp. 133–139.

Spanos, D.P., Olfati-Saber, R., and Murray, R. (2005b), 'Distributed Sensor Fusion using Dynamic Consensus', in *Proceedings of the 16th IFAC World Congress*.

Tsitsiklis, J. (1993), 'Decentralised Detection', in *Advances in Signal Processing, Vol. 2*, eds. H.V. Poor, and J.B. Thomas, London, Uk: JAI Press, pp. 297–344.

Xiao, L., Boyd, S., and Lall, S. (2005), 'A Scheme for Robust Distributed Sensor Fusion Based on Average Consensus', in *International Conference on Information Processing in Sensor Networks (IPSN 2005)*, pp. 63–70.

——— (2006), '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)*, pp. 168–176.

A. Appendix

A.1 *Graphs and non-negative matrices*

A *directed graph* is a pair $\mathcal{G}=(\mathcal{V},\mathcal{E})$, where $\mathcal{V}=\{1,2,\ldots,n\}$ is a set of *nodes* and $\mathcal{E}\subseteq\mathcal{V}\times\mathcal{V}$ is a set of ordered *edges*, such that $(i,j)\in\mathcal{E}$ if node i 'is a neighbour' of node j. If $(j,i)\in\mathcal{E}$ whenever $(i,j)\in\mathcal{E}$, the graph is *undirected*. Two nodes i, j are connected if there exist a sequence of distinct edges (i.e. a *path*) leading from i to j. We consider graphs with self loops, which means that a node is always connected with itself. A graph is *connected* if every pair of nodes is connected.

A square matrix $R\geq 0$ is *primitive*, if there exist an integer $m\geq 1$ such that $R^m>0$. If R is primitive, then $\rho(R)$ is an algebraically simple eigenvalue of R and the eigen-space associated with this eigenvalue is one-dimensional. Let us now put in relation the notions of connectivity of a graph and primitiveness of a matrix. To this end, define the following set of non-negative matrices with positive diagonal entries:

$$\mathcal{M}\doteq\{R\in\mathbb{R}^{n,n}:\ R\geq 0,\ R_{ii}>0,\ i=1,\ldots,n\}. \tag{23}$$

Notice that $\mathcal{M}$ is closed under addition and multiplication.

Definition A.1: For $R\in\mathcal{M}$, we say that the matrix/graph pair $(R,\mathcal{G}(\mathcal{V},\mathcal{E}))$ is compatible if $R_{ij}>0\Leftrightarrow(i,j)\in\mathcal{E}$.

The following theorem can be readily established (using for instance Theorem 6.2.24, Theorem 8.5.2 and Lemma 8.5.5 of Horn and Johnson (1985)).

Theorem A.1: *Let $R\in\mathcal{M}$ such that $(R,\mathcal{G}(\mathcal{V},\mathcal{E}))$ is a compatible pair. Then R is primitive if and only if $\mathcal{G}$ is connected.*

Let $\mathcal{G}_a(\mathcal{V},\mathcal{E}_a)$, $\mathcal{G}_b(\mathcal{V},\mathcal{E}_b)$ be two graphs with common vertex set. The *composition* $\mathcal{G}_{ab}=\mathcal{G}_a\circ G_b$ of the two graphs is defined as the graph with vertex set $\mathcal{V}$ such that $(i,j)\in\mathcal{E}_{ab}$ if and only if $(i,k)\in\mathcal{E}_a$ and $(k,j)\in\mathcal{E}_b$ for some $k\in\mathcal{V}$. The *union* $\mathcal{G}_{a|b}=\mathcal{G}_a\cup G_b$ of the two graphs is defined as the graph with vertex set $\mathcal{V}$ such that $(i,j)\in\mathcal{E}_{a|b}$ if and only if $(i,j)\in\mathcal{E}_a$ or $(i,j)\in\mathcal{E}_b$. Notice that graph union is commutative whereas composition is not. Note also that the edge set of the union is a subset of the edge set of the composition of any permutation of the graphs, that is $\mathcal{E}_{a|b}\subseteq\mathcal{E}_{ab}$ and $\mathcal{E}_{a|b}\subseteq\mathcal{E}_{ba}$.[2]

Let now $A\in\mathcal{M}$ be compatible with $\mathcal{G}_a$, and $B\in\mathcal{M}$ be compatible with $\mathcal{G}_b$. Then, it is a matter of simple matrix algebra to verify that the product AB is compatible with the composition graph $\mathcal{G}_{ab}$ (to this end, just notice that $[AB]_{ij}=\sum_{k=1}^{n}A_{ik}B_{kj}>0$ if and only if $A_{ik}>0$ and $B_{kj}>0$ for some k). Similarly, one can show that the sum $A+B$ is compatible with the union graph $\mathcal{G}_{a|b}$. These notions of composition and union of two graphs can obviously be extended to sequences of an arbitrary number of graphs. We thus have the following result.

Lemma A.1: *Let $\{\mathcal{G}_k(\mathcal{V},\mathcal{E}_k),k=1,\ldots,q\}$ be a sequence of graphs, and let $\{R_k\in\mathcal{M},k=1,\ldots,q\}$ be any sequence of matrices such that R_k is compatible with $\mathcal{G}_k$, for $k=1,\ldots,q$. Then*:

- *The product $R_1R_2\cdots R_q$ is primitive if and only if the composition graph $\mathcal{G}_1\circ\mathcal{G}_2\circ\cdots\circ\mathcal{G}_q$ is connected;*
- *The sum $\alpha_1R_1+\alpha_2R_2+\cdots+\alpha_qR_q$, $\alpha_i>0$, is primitive if and only if the union graph $\mathcal{G}_1\cup\mathcal{G}_2\cup\cdots\cup\mathcal{G}_q$ is connected.*

As we have observed previously, the edge set of the union of a sequence of graphs is a subset of the edge set of the composition of any permutation of the graph sequence. This means in turn that if the union graph $\mathcal{G}_1\cup\mathcal{G}_2\cup\cdots\cup\mathcal{G}_q$ is connected, then any composition $\mathcal{G}_{p_1}\circ\mathcal{G}_{p_2}\circ\cdots\circ\mathcal{G}_{p_q}$, where $\{p_1,\ldots,p_q\}$ is a permutation of $\{1,\ldots,q\}$, is also connected. From this reasoning we obtain the following corollary of Lemma A.1.

Corollary A.1: *Let $\{\mathcal{G}_k(\mathcal{V},\mathcal{E}_k),\ k=1,\ldots,q\}$ be a sequence of graphs, and let $\{R_k\in\mathcal{M},\ k=1,\ldots,q\}$ be any sequence of matrices such that R_k is compatible with $\mathcal{G}_k$, for $k=1,\ldots,q$.*

If the union graph $\mathcal{G}_1\cup\mathcal{G}_2\cup\cdots\cup\mathcal{G}_q$ is connected then the product $R_{p_1}R_{p_2}\cdots R_{p_q}$ is primitive, for any permutation $\{p_1,\ldots,p_q\}$ of $\{1,\ldots,q\}$.

A.2 *Stochastic non-negative matrices*

Consider a subset of $\mathcal{M}$ composed of matrices in which the sum over each row is equal to one (such matrices are usually called (row) *stochastic*):

$$\mathcal{M}_{\mathrm{s}}\doteq\{R\in\mathcal{M}:\ R\mathbf{1}=\mathbf{1}\}. \tag{24}$$

The set $\mathcal{M}_{\mathrm{s}}$ is convex and closed under multiplication. For $R\in\mathcal{M}_{\mathrm{s}}$ we have that 1 is an eigenvalue of R. Observe that the spectral radius of a matrix is no larger than any norm of the matrix (see Theorem 5.6.9 of Horn and Johnson (1985)), hence by taking the ℓ_∞-induced norm we have that for $R\in\mathcal{M}_{\mathrm{s}}$ it holds that

$$\rho(R)\leq\|R\|_\infty=\max_{i=1,\ldots,n}\sum_{j=1}^{n}|R_{ij}|=1.$$

Since 1 is an eigenvalue of R it therefore must be $\rho(R)=1$. It also follows that if $R\in\mathcal{M}_{\mathrm{s}}$ is primitive then R has a unique (i.e. an algebraically simple) eigenvalue at 1, hence all other eigenvalues have modulus strictly smaller than 1 and the fixed-point subspace

$$\mathcal{I}(R)\doteq\{x\in\mathbb{R}^n:\ Rx=x\}$$

is one dimensional. The following result holds (see Theorem 8.5.1 of Horn and Johnson (1985)).

Theorem A.2: *Let $R\in\mathcal{M}_{\mathrm{s}}$ and let $\lambda_1(R)$, $\lambda_2(R),\ldots,\lambda_n(R)$ denote the eigenvalues of R ordered with non-increasing modulus, with $\lambda_1(R)=\rho(R)=1$. If R is primitive, then $\lambda_i(R)<1$ for $i=2,\ldots,n$. Moreover*

$$\mathcal{I}(R)=\mathrm{span}\{\mathbf{1}\},$$

and

$$\lim_{k\to\infty}R^k=\frac{1}{n}\mathbf{1}v^\top>0,$$

where $v>0$ is a left eigenvector of R associated with $\lambda_1(R)=1$: $v^\top R = v^\top$. In the particular case also when the sum over each column of R is one (R is doubly stochastic), we simply have that $v=\mathbf{1}$.

A.2.1 *Symmetric paracontractions*

A matrix $R\in\mathbb{R}^{n,n}$ is called *paracontractive* (Hartfiel 2002) with respect to the Euclidean norm if

$$\|Rx\| < \|x\|, \quad \forall x \neq 0,\ x \notin \mathcal{I}(R),$$

that is, a paracontractive matrix is contractive for all non-null vectors lying outside its fixed-point subspace. It is easy to verify that a symmetric matrix is paracontractive if and only if its spectrum belongs to the semi-open interval $(-1, 1]$. Now consider the subset of $\mathcal{M}_s$ formed by symmetric matrices:

$$\mathcal{M}_{ss} \doteq \{R \in \mathcal{M}_s:\ R = R^T\}. \tag{25}$$

(a matrix $R\in\mathcal{M}_{ss}$ is symmetric, non-negative with strictly positive diagonal elements and doubly stochastic, that is $R\mathbf{1}=\mathbf{1}, \mathbf{1}^\top R=\mathbf{1}^\top$). For $R\in\mathcal{M}_{ss}$ we clearly have that $\rho(R)\equiv\|R\|=1$. Moreover, the following fact holds, see, for instance, Xiao, Boyd, and Lall (2005).

Lemma A.2: *If $R\in\mathcal{M}_{ss}$ then $\sigma(R)\subset(-1,1]$, hence R is paracontractive.*

We are interested in developing conditions under which a sequence of matrices from $\mathcal{M}_{ss}$ forms products that are primitive and paracontractive, i.e. that are contractive for all vectors not lying in span$\{\mathbf{1}\}$. We preliminarily recall from Theorem A.2 that if $R\in\mathcal{M}_s$ is primitive then $\rho(R)=1$ is a simple eigenvalue of R and the corresponding eigenspace $\mathcal{I}(R)$ is one dimensional, that is

$$\mathcal{I}(R) = \text{span}\{\mathbf{1}\}, \quad \text{for } R \in \mathcal{M}_s \text{ primitive.} \tag{26}$$

We are now in position to state the following result, which specialises Corollary A.1 to the case of products of symmetric stochastic matrices.

Corollary A.2: *Let $\{R_k\in\mathcal{M}_{ss}, k=1,\dots,q\}$ be a sequence of matrices and $\{\mathcal{G}_k(\mathcal{V},\mathcal{E}_k), k=1,\dots,q\}$ a sequence of graphs such that R_k is compatible with $\mathcal{G}_k$, for $k=1,\dots,q$.*

If the union graph $\mathcal{G}_1\cup\mathcal{G}_2\cup\cdots\cup\mathcal{G}_q$ is connected then the product $R_{p_1}R_{p_2}\cdots R_{p_q}$ is primitive and paracontractive, for any permutation $\{p_1,\dots,p_q\}$ of $\{1,\dots,q\}$.

Proof: Let us denote for ease of notation with $P\doteq R_{p_1}R_{p_2}\cdots R_{p_q}$ the product of a generic permutation of matrices $R_1,\dots,R_q\in\mathcal{M}_{ss}$. We have that $P\in\mathcal{M}_s$, and from Corollary A.1 it follows that if the union graph $\mathcal{G}_1\cup\mathcal{G}_2\cup\cdots\cup\mathcal{G}_q$ is connected then P is primitive. Therefore, using (26) we have that:

$$\mathcal{I}(P) = \text{span}\{\mathbf{1}\}. \tag{27}$$

We next show that

$$\|Pv\| < \|v\| \quad \text{for all } v \notin \mathcal{I}(P), \tag{28}$$

which indeed means that P is paracontractive. To this end, let $v\notin\mathcal{I}(P)$ and notice that in forming the product $Pv = R_{p_1}R_{p_2}\cdots R_{p_q}v$ only two cases might arise: either $v\in\mathcal{I}(R_{p_k})$ for all $k=1,\dots,q$, or not. But $v\in\mathcal{I}(R_{p_k})$ for all $k=1,\dots,q$, would imply that $Pv = R_{p_1}R_{p_2}\cdots R_{p_q}v = R_{p_1}R_{p_2}\cdots R_{p_{q-1}}v = \cdots = R_{p_1}v = v$, which would mean that $v\in\mathcal{I}(P)$, and this is not possible since we selected $v\notin\mathcal{I}(P)$. Therefore there must exist an index $z\in\{1,\dots,q\}$ such that $v\in\mathcal{I}(R_{p_k})$ for all $k=z+1,\dots,q$ and $v\notin\mathcal{I}(R_{p_z})$. Hence, we have that

$$Pv = R_{p_1}R_{p_2}\cdots R_{p_z}v, \quad \text{where } R_{p_z}v \neq v.$$

Since R_{p_z} is paracontractive we have that $\|R_{p_z}v\|<\|v\|$, and therefore by sub-multiplicativity of norms it follows that:

$$\begin{aligned}\|Pv\| &= \|R_{p_1}R_{p_2}\cdots R_{p_z}v\| \le \|R_{p_1}R_{p_2}\cdots R_{p_{z-1}}\|\cdot\|R_{p_z}v\| \\ &\le \|R_{p_z}v\| < \|v\|,\end{aligned}$$

which proves that P is paracontractive. □

Consider now any q matrices $R_k\in\mathcal{M}_{ss}, k=1,\dots,q$. Since each $R_k\in\mathcal{M}_{ss}$ is symmetric, it is unitarily diagonalisable, that is, it admits a set of orthogonal eigenvectors. $\mathbf{1}/\sqrt{n}$ is always an eigenvector of R_k associated with the largest-modulus eigenvalue $\lambda_1(R_k)=1$, and we may write $R_k\in\mathcal{M}_{ss}$ in the form:

$$R_k = \frac{1}{n}\mathbf{1}\mathbf{1}^\top + Z_k, \quad Z_k = V_k D_k V_k^\top, \tag{29}$$

where $V_k\in\mathbb{R}^{n,n-1}$ is such that

$$V_k^\top V_k = I_{n-1}, \quad V_k V_k^\top = I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^\top, \quad \mathbf{1}^\top V_k = 0, \tag{30}$$

and $D_k=\text{diag}(\lambda_2(R_k),\dots,\lambda_n(R_k))\in\mathbb{R}^{n-1,n-1}$ is a diagonal matrix containing the last $n-1$ eigenvalues of R_k arranged in order of non-increasing modulus. Since $\mathbf{1}\mathbf{1}^\top$ is orthogonal to all Z_k's, we can write the product $(R_1\cdots R_q)$ in the form:

$$(R_1\cdots R_q) = \frac{1}{n}\mathbf{1}\mathbf{1}^\top + (Z_1\cdots Z_q). \tag{31}$$

The following lemma holds.

Lemma A.3: *Let $R_k\in\mathcal{M}_{ss}, Z_k\doteq R_k-(1/n)\mathbf{1}\mathbf{1}^\top, k=1,\dots,q$. If the product $(R_1\cdots R_q)$ is primitive and paracontractive, then*

$$\|Z_1\cdots Z_q\| < 1,$$

that is, the product $(Z_1\cdots Z_q)$ is contractive with respect to the spectral norm.

Proof: Let $P=(R_1\ \cdots\ R_q)$, $Q=(Z_1\ \cdots\ Z_q)$. From (31), we have

$$Q = P - \frac{1}{n}\mathbf{1}\mathbf{1}^\top.$$

Since P is row stochastic, primitive and paracontracting, we have that $Pv=v$ if $v\in\text{span}\{\mathbf{1}\}$ and $\|Pv\|<\|v\|$ otherwise. Then, let a generic vector $z\in\mathbb{R}^n$ be represented as the sum of orthogonal components:

$$z = x + y, \quad x = \alpha\mathbf{1} \in \text{span}\{\mathbf{1}\},\ y \in \text{span}\{\mathbf{1}\}^\perp,$$

and note that

$$\|z\|^2 = \|x\|^2 + \|y\|^2 \quad \Rightarrow \quad \|y\| \le \|z\|.$$

We have

$$\begin{aligned}Qz &= \left(P - \frac{1}{n}\mathbf{1}\mathbf{1}^\top\right)(\alpha\mathbf{1}+y) \\ &= \alpha P\mathbf{1} + Py - \alpha\frac{1}{n}\mathbf{1}\mathbf{1}^\top\mathbf{1} - \frac{1}{n}\mathbf{1}\mathbf{1}^\top y \\ &= \alpha\mathbf{1} + Py - \alpha\mathbf{1} - 0 \\ &= Py,\end{aligned}$$

hence

$$\|Qz\| = \|Py\| < \|y\| \leq \|z\|, \quad \forall z \in \mathbb{R}^n,\ z \neq 0$$

and therefore

$$\|Q\| = \sup_{z \neq 0} \frac{\|Qz\|}{\|z\|} < 1,$$

which proves that Q is contractive with respect to the spectral norm. □

A.3 *Proof of Lemma 3.1*

Since matrices $\{R_{t-k}, \ldots, R_{t-1}\}$ belong to $\mathcal{M}_{ss}$ and are compatible with the graph sequence $\{\mathcal{G}_{t-k}, \ldots, \mathcal{G}_{t-1}\}$, which has $N(k)$ jointly connected subsequences, by Corollary A.2 each subproduct $(R_{e_i}, R_{e_i-1}, \ldots, R_{s_i})$ is primitive and paracontracting. Therefore, using the representation in (29), (30) and applying Lemma A.3, each subproduct $(Z_{e_i}, Z_{e_i-1}, \ldots, Z_{s_i})$ is contractive with respect to the spectral norm. Since such subproducts are in number of $N(k)$, we may define

$$\lambda(t-k, t-1) \doteq \sup_{i=1,\ldots,N(k)} \|(Z_{s_{i+1}}, \ldots, Z_{s_i+1})\| < 1.$$

Let now consider the supremum of $\lambda(t-k, t-1)$ over $t \geq 1$, $k \geq \bar{k}$. Since we assume that the set of possible R_τ matrices is of finite cardinality, and that the number of factors in the product $(Z_{e_i}, Z_{e_i-1}, \ldots, Z_{s_i})$ is finite by the RJC hypothesis, also the set of all possible products $(Z_{e_i}, Z_{e_i-1}, \ldots, Z_{s_i})$ is finite, hence (18) follows by defining

$$\lambda \doteq \sup_{t \geq 1, k \geq 1} \lambda(t-k, t-1) < 1.$$

□

A.4 *Proof of Lemma 3.2*

Let $t \geq 1$ and consider any subsequence of S of length k, $S(t-k, t-1) = \{\mathcal{G}_{t-k}, \ldots, \mathcal{G}_{t-1}\}$, and a corresponding matrix subsequence $\mathcal{R}(t-k, t-1) = \{R_{t-k}, \ldots, R_{t-1}\}$. Since S is log-RJC with index $\bar{\lambda} < 1$ with respect to $\mathcal{R}$, we have from Definition 3.3 that there exist $\alpha > 0$ and a finite $\bar{k} \geq \alpha$ such that $N(k) \geq N_{lb}(k)$ for all $k > \bar{k}$. Moreover, for all $k > \bar{k}$, sequence $S(t-k, t-1)$ contains $N(k)$ jointly connected subsequences. Let $[s_1, e_1], [s_2, e_2], \ldots, [s_{N(k)}, e_{N(k)}]$, $s_1 \geq t-k$, $e_{N(k)} \leq t-1$, denote the indices delimiting such subsequences. Then, the product $(Z_{t-1}Z_{t-2}\cdots Z_{t-k})$ can be written as $(F_{N(k)}Q_{N(k)}F_{N(k)-1}Q_{N(k)-1}\cdots F_1Q_1F_0)$, where Q_i, $i = 1, \ldots, N(k)$, are the partial products of factors corresponding to the jointly connected subsequences $Q_i = (Z_{e_i}, Z_{e_i-1}\ldots Z_{s_i})$, and where F_i, $i = 0, \ldots, N(k)$, denote the products of Z factors corresponding to the 'filling' terms between jointly connected subsequences. Therefore, according to Lemma 3.1 and the definition of joint connectivity index, it holds that $\|Q_i\| \leq \bar{\lambda} < 1$, $i = 1, \ldots, N(k)$. Then, we have that

$$\begin{aligned}
\|Z_{t-1}Z_{t-2}\cdots Z_{t-k}\| &= \|F_{N(k)}Q_{N(k)}F_{N(k)-1}Q_{N(k)-1}\cdots F_1Q_1F_0\| \\
&\leq \|F_{N(k)}\| \cdot \|Q_{N(k)}\| \cdots \|F_1\| \cdot \|Q_1\| \cdot \|F_0\| \\
&\leq 1 \cdot \bar{\lambda} \cdots 1 \cdot \bar{\lambda} \cdot 1 \\
&= \bar{\lambda}^{N(k)} \leq \bar{\lambda}^{N_{lb}(k)} \leq \bar{\lambda}^{(2\log(k/\alpha)/\log(1/\bar{\lambda}))} \\
&= \frac{\alpha^2}{k^2}, \quad \text{for } k > \bar{k}.
\end{aligned}$$

If instead $k \leq \bar{k}$, we may just write

$$\|Z_{t-1}Z_{t-2}\cdots Z_{t-k}\| \leq \|Z_{t-1}\| \cdots \|Z_{t-k}\| \leq 1.$$

Therefore,

$$\begin{aligned}
&\lim_{t\to\infty} \sum_{k=1}^{t} \|Z_{t-1}Z_{t-2}\cdots Z_{t-k}\| \\
&\quad = \sum_{k=1}^{\bar{k}} \|Z_{t-1}Z_{t-2}\cdots Z_{t-k}\| + \lim_{t\to\infty} \sum_{k=\bar{k}+1}^{t} \|Z_{t-1}Z_{t-2}\cdots Z_{t-k}\| \\
&\quad \leq \sum_{k=1}^{\bar{k}} 1 + \lim_{t\to\infty} \sum_{k=\bar{k}+1}^{t} \frac{\alpha^2}{k^2} = \bar{k} + \lim_{t\to\infty} \left(\sum_{k=1}^{t} \frac{\alpha^2}{k^2} - \sum_{k=1}^{\bar{k}} \frac{\alpha^2}{k^2} \right) \\
&\quad = \left(\bar{k} - \sum_{k=1}^{\bar{k}} \frac{\alpha^2}{k^2} \right) + \frac{\alpha^2\pi}{6},
\end{aligned}$$

where the last statement follows by recalling the sum of infinite series $\sum_{k=1}^{\infty} 1/k^2 = \pi/6$. □

A.5 *Proof of Proposition 2.1*

Notice from (6) that matrix $W(t)$ is symmetric and doubly stochastic (i.e. $\mathbf{1}^\top W(t) = \mathbf{1}^\top$, $W(t)\mathbf{1} = \mathbf{1}$). Moreover, $W_{ij}(t) \in [0, 1]$ and $W_{ii}(t)$ are strictly positive, see, e.g. Xiao et al. (2005); § II.A. Then, Equation (7) writes

$$\begin{aligned}
P_i(t+1) &= \left(1 - \sum_{j \in \mathcal{N}_i(t)} W_{ij}(t)\right) P_i(t_+) + \sum_{j \in \mathcal{N}_i(t)} W_{ij}(t) P_j(t_+) \\
&= W_{ii}(t) P_i(t_+) + \sum_{j \in \mathcal{N}_i(t)} W_{ij}(t) P_j(t_+) \\
&= \sum_{j=1}^{n} W_{ij}(t) P_j(t_+).
\end{aligned}$$

An analogous expression can be derived for $q_i(t+1)$, therefore the spatial update equations take the following simplified form

$$P_i(t+1) = \sum_{j=1}^{n} W_{ij}(t) P_j(t_+) \tag{32}$$

$$q_i(t+1) = \sum_{j=1}^{n} W_{ij}(t) q_j(t_+). \tag{33}$$

Substituting (4) and (5) into these latter expressions, we obtain the final space-time recursions for the composite information matrices $P_i(t)$ and composite information states $q_i(t)$:

$$\begin{aligned}
P_i(t+1) &= \frac{t}{t+1} \sum_{j=1}^{n} W_{ij}(t) P_j(t) \\
&\quad + \frac{1}{t+1} \sum_{j=1}^{n} W_{ij}(t) A_j^\top(t) \Sigma_j^{-1}(t) A_j(t)
\end{aligned} \tag{34}$$

$$\begin{aligned}
q_i(t+1) &= \frac{t}{t+1} \sum_{j=1}^{n} W_{ij}(t) q_j(t) \\
&\quad + \frac{1}{t+1} \sum_{j=1}^{n} W_{ij}(t) A_j^\top(t) \Sigma_j^{-1}(t) y_j(t).
\end{aligned} \tag{35}$$

Note that in these recursions the initial values of $P_i(0)$, $q_i(0)$ are irrelevant, since they are multiplied by $t=0$. From the above recursions, it is not difficult to find a general expression for $P_i(t)$, $q_i(t)$, $t=1,2,\ldots$ To this end, let us work on the recursion (35), the case of (34) being completely analogous. Defining

$$q(t) \doteq \begin{bmatrix} q_1(t) \\ q_2(t) \\ \vdots \\ q_n(t) \end{bmatrix}, \quad u(t) \doteq \begin{bmatrix} A_1^\top(t)\Sigma_1^{-1}(t)y_1(t) \\ A_2^\top(t)\Sigma_2^{-1}(t)y_2(t) \\ \vdots \\ A_n^\top(t)\Sigma_n^{-1}(t)y_n(t) \end{bmatrix}$$

$$\mathcal{W}(t) \doteq \begin{bmatrix} W_{11}(t)I & \cdots & W_{1n}(t)I \\ W_{21}(t)I & \cdots & W_{2n}(t)I \\ \vdots & \cdots & \vdots \\ W_{n1}(t)I & \cdots & W_{nn}(t)I \end{bmatrix} = W(t) \otimes I,$$

we write (35), $i=1,\ldots,n$, in compact vector form

$$q(t+1) = \frac{t}{t+1}\mathcal{W}(t)q(t) + \frac{1}{t+1}\mathcal{W}(t)u(t). \tag{36}$$

Applying (36) recursively for $t=0,1,\ldots$ we obtain

$$\begin{aligned}
q(1) &= \mathcal{W}(0)u(0) \\
q(2) &= \frac{1}{2}\mathcal{W}(1)q(1) + \frac{1}{2}\mathcal{W}(1)u(1) \\
&= \frac{1}{2}\mathcal{W}(1)\mathcal{W}(0)u(0) + \frac{1}{2}\mathcal{W}(1)u(1) \\
q(3) &= \frac{2}{3}\mathcal{W}(2)q(2) + \frac{1}{3}\mathcal{W}(2)u(2) \\
&= \frac{2}{3}\frac{1}{2}\mathcal{W}(2)\mathcal{W}(1)\mathcal{W}(0)u(0) + \frac{2}{3}\frac{1}{2}\mathcal{W}(2)\mathcal{W}(1)u(1) + \frac{1}{3}\mathcal{W}(2)u(2) \\
&= \frac{1}{3}\mathcal{W}(2)\mathcal{W}(1)\mathcal{W}(0)u(0) + \frac{1}{3}\mathcal{W}(2)\mathcal{W}(1)u(1) + \frac{1}{3}\mathcal{W}(2)u(2) \\
&\vdots \quad \vdots \quad \vdots \\
q(t) &= \frac{1}{t}(\mathcal{W}(t-1)\cdots\mathcal{W}(0)u(0) + \mathcal{W}(t-1)\cdots\mathcal{W}(1)u(1) + \cdots \\
&\qquad + \mathcal{W}(t-1)u(t-1)),
\end{aligned}$$

that is

$$\begin{aligned}
q(t) &= \frac{1}{t}\sum_{k=0}^{t-1}[\mathcal{W}(t-1)\cdots\mathcal{W}(k)]u(k) \\
&= \frac{1}{t}\sum_{k=0}^{t-1}[(W(t-1)\cdots W(k)) \otimes I]u(k) \\
&= \frac{1}{t}\sum_{k=0}^{t-1}[\Phi(t-1;k) \otimes I]u(k), \quad t=1,2,\ldots,
\end{aligned}$$

where we defined

$$\Phi(t-1;k) \doteq W(t-1)\cdots W(k). \tag{37}$$

The i-th vector component in $q(t)$ hence writes

$$q_i(t) = \frac{1}{t}\sum_{k=0}^{t-1}\sum_{j=1}^{n}\Phi_{ij}(t-1;k)u_j(k), \quad t=1,2,\ldots,$$

which concludes the proof. □