A Space-Time Diffusion Scheme for Peer-to-Peer Least-Squares Estimation

Lin Xiao
Center for the Mathematics of Information
Caltech
Pasadena, CA 91125-9300
lxiao@caltech.edu

Stephen Boyd
Department of Electrical Engineering
Stanford University
Stanford, CA 94305-9510
boyd@stanford.edu

Sanjay Lall
Department of Aeronautics and Astronautics
Stanford University
Stanford, CA 94305-4035
lall@stanford.edu

ABSTRACT

We consider a sensor network in which each sensor takes measurements, at various times, of some unknown parameters, corrupted by independent Gaussian noises. Each node can take a finite or infinite number of measurements, at arbitrary times (i.e., asynchronously). We propose a space-time diffusion scheme, that relies only on peer-to-peer communication, and allows every node to asymptotically compute the global maximum-likelihood estimate of the unknown parameters. At each iteration, information is diffused across the network by a temporal update step and a spatial update step. Both steps update each node's state by a weighted average of its current value and locally available data: new measurements for the time update, and neighbors' data for the spatial update. At any time, any node can compute a local weighted least-squares estimate of the unknown parameters, which converges to the global maximum-likelihood solution. With an infinite number of measurements, these estimates converge to the true parameter values in the sense of mean-square convergence. We show that this scheme is robust to unreliable communication links, and works in a network with dynamically changing topology.

Categories and Subject Descriptors: G.1.6 [**Numerical Analysis**]: Optimization – *least squares methods*; G.1.3 [**Numerical Analysis**]: Numerical Linear Algebra – *Linear systems (direct and iterative methods).*

General Terms: Algorithms, Theory.

Keywords: distributed algorithms, sensor networks, least-squares, estimation.

1. INTRODUCTION

With advances in integrated circuits, radio technology and mechanical miniaturization, future sensor networks will consist of nodes integrated with substantial capabilities for sensing, communication, computing and actuation. In-network signal processing and data fusion is considered a critical functionality for such *intelligent* or *autonomous* sensor networks. This opens new opportunities for the development of robust, asynchronous, distributed algorithms that work in a dynamically changing environment. The diffusion equation, as one of the most fundamental governing equations of the physical world, gives a natural algorithm for distributed computing and information fusion, with inherent robustness to various failures, and the ability to adapt to changes in the network topology. In this paper, we focus on a specific model of in-network collaborative signal processing, where the common goal is linear parameter estimation. We propose a fully distributed space-time diffusion scheme that enables every node to asymptotically compute the maximum likelihood solution.

1.1 Maximum-likelihood parameter estimation

We consider estimation of a vector of unknown parameters $\theta \in \mathbf{R}^m$ using a network of n distributed sensors. In our setup, each sensor node can take multiple measurements, each at a different time, occurring asynchronously in the network. Let $\mathcal{T}_i \subseteq \{0, 1, 2, \ldots\}$ denote the set of times that node i takes measurements. We assume all measurements are linear, with additive noise:

$$y_i(t) = A_i(t)\theta + v_i(t), \qquad t \in \mathcal{T}_i, \qquad i = 1, \ldots, n,$$

where $y_i(t) \in \mathbf{R}^{m_i(t)}$ is the measurement, $A_i(t)$ is a known matrix that relates the unknown parameter to the ith sensor measurement at time t, and $v_i(t)$ is a measurement noise. We assume $v_i(t)$ are independent Gaussian random variables with zero mean and covariance matrix $\Sigma_i(t)$. The total number of measurements taken by each sensor, $|\mathcal{T}_i|$, can be finite or infinite. For convenience of presentation, we assume for now that all $|\mathcal{T}_i|$ are finite. We will give extensions to the infinite case in §6.

With the assumption that every sensor takes a finite number of measurements, we can denote the aggregate measurement of all sensors as a vector

$$y = A\theta + v,$$

where y is a column vector obtained by stacking the measurements $y_i(t)$, A is a matrix obtained by stacking the matrices $A_i(t)$, and v is a column vector obtained by stacking

IPSN'06, April 19–21, 2006, Nashville, Tennessee, USA.

the noises $v_i(t)$. Since the noises $v_i(t)$ are independent, the covariance matrix of v is a block diagonal matrix with $\Sigma_i(t)$ on its diagonal blocks. We assume that the matrix A has full column rank. The maximum-likelihood (ML) estimate of θ, given the measurement y, is the weighted least-squares approximate solution

$$\hat{\theta}_{\mathrm{ML}} = \left(A^T\Sigma^{-1}A\right)^{-1} A^T\Sigma^{-1}y. \qquad (1)$$

Since Σ is block diagonal, we have

$$A^T\Sigma^{-1}A = \sum_{i=1}^{n}\sum_{t\in\mathcal{T}_i} A_i(t)^T\Sigma_i(t)^{-1}A_i(t)$$
$$A^T\Sigma^{-1}y = \sum_{i=1}^{n}\sum_{t\in\mathcal{T}_i} A_i(t)^T\Sigma_i(t)^{-1}y_i(t).$$

This estimate is unbiased (i.e., $\mathbf{E}\,\hat{\theta}_{\mathrm{ML}} = \theta$) and has error covariance matrix

$$Q_{\mathrm{ML}} = \mathbf{E}\left((\hat{\theta}_{\mathrm{ML}}-\theta)(\hat{\theta}_{\mathrm{ML}}-\theta)^T\right) = \left(A^T\Sigma^{-1}A\right)^{-1}. \qquad (2)$$

More generally, if the noises are not Gaussian, but are independent and have zero mean and covariances $\Sigma_i(t)$, the formula (1) gives the linear minimum-variance unbiased estimate of θ, given the measurements.

1.2 Peer-to-Peer least-squares problem

The Peer-to-Peer least-square (P2PLS) problem is to compute the maximum-likelihood estimate $\hat{\theta}_{\mathrm{ML}}$ in (1) using a distributed sensor fusion scheme. In such a scheme, there is no fusion center that collects and processes all measurement data, and the sensor nodes do not have global knowledge of the network topology. Each sensor only exchanges data with its neighbors and carries out local computation. The goal is for every sensor to eventually have a good estimate of the unknown parameters, e.g., to obtain $\hat{\theta}_{\mathrm{ML}}$. This is particularly important if the sensor nodes are carrying out multiple tasks (e.g., a team of robots) and need the estimate $\hat{\theta}_{\mathrm{ML}}$ to make local decisions.

There are many ways to do distributed sensor fusion. One straightforward method is flooding. Each sensor node broadcasts all its stored data (i.e., $y_i(t)$, $A_i(t)$ and $\Sigma_i(t)$ for some subset of the sensor nodes) to its neighbors, and stores all new data received. Eventually each node has all the data in the network, and thus can act as a fusion center to obtain $\hat{\theta}_{\mathrm{ML}}$. This method can require a large amount of data communication, storage memory, and book-keeping overhead. Much more sophisticated algorithms for distributed detection, estimation and inference in sensor networks can be found in, e.g., [10, 8, 1, 12, 18].

In [23] we proposed a simple iterative scheme for distributed sensor fusion based on average consensus (e.g., [22, 17]). It can be considered as an algorithm for a special case of the P2PLS problem where each sensor makes a single measurement at time $t = 0$ (i.e., $\mathcal{T}_i = \{0\}$ for all i). This scheme does not involve explicit point-to-point message passing or routing; instead, it diffuses information across the network by updating each node's data with a weighted average of its neighbors' data. At each iteration, every node can compute a local weighted least-square estimate, which eventually converges to the global maximum-likelihood solution. This scheme is robust to unreliable communication links (e.g., due to mobility, fading or power constraints in wireless sensor networks). We showed that it works in a network with dynamically changing topology, provided that the set of infinitely occurring communication graphs is jointly connected. Related work on distributed sensor fusion based on average consensus includes [14, 5, 6].

In this paper, we present a *space-time diffusion* scheme for the general P2PLS problem. This approach can be thought of as an extension of the distributed average consensus approach (*space diffusion*) presented in [23]. More specifically, at each iteration, it diffuses information by a *temporal update* step and a *spatial update* step. In the temporal update step, each node takes a weighted average of its local data and the new measurement if available. In the spatial update step, each node updates its data with a weighted average of its own and neighbors' data (they maintain the same data structure). The averaging weights are time varying and depend on the number of measurements each node has made. If every node takes a finite number of measurements, we show that the space-time diffusion scheme converges to the maximum likelihood solution under a similar condition to that required for convergence of space diffusion. In the case of an infinite number of measurements, we are able to show convergence of the nodes that take measurements most frequently, but we conjecture that convergence occurs under much weaker assumptions.

1.3 Outline

The rest of this paper is organized as follows. In §2 we review the space diffusion scheme (distributed average consensus) and the related convergence result. In §3 we give the space-time diffusion scheme for the scalar case and state its convergence result for a finite number of measurements. In §4 we show how the space-time diffusion scheme is constructed and prove its convergence when there are finitely many measurements. In §5 we apply the space-time diffusion scheme to solve the general P2PLS problem. In §6 we give a convergence result which holds when there are infinitely many measurements. In §7 we demonstrate the space-time diffusion scheme with numerical examples.

2. REVIEW OF SPACE DIFFUSION

First we introduce some notation. We represent the time-varying communication graphs of the sensor network by undirected graphs $\mathcal{G}(t) = (\mathcal{V}, \mathcal{E}(t))$, where $\mathcal{V} = \{1, 2, \ldots, n\}$ is the set of nodes and $\mathcal{E}(t)$ is the set of active links at time t. The sequence of communication graphs $\{\mathcal{G}(t)\}_{t=0}^{\infty}$ can be either deterministic or stochastic. In the latter case, we assume that they are independent of the measurement noises. Let $\mathcal{N}_i(t) = \{j \in \mathcal{V} \mid \{i,j\} \in \mathcal{E}(t)\}$ denote the set of neighbors of node i at time t. In a peer-to-peer network, each node is only allowed to communicate with its instantaneous neighbors in $\mathcal{N}_i(t)$, and may carry out local computation.

The space diffusion scheme works for the special case of $|\mathcal{T}_i| = 1$ for $i = 1, \ldots, n$. For convenience, we also assume that every sensor takes its measurement at $t = 0$. Alternatively, we can assume that a node that takes its first measurement at a later time is not connected to any other nodes until its measurement is taken. We demonstrate the basic algorithm with the scalar case, where the measurements have the form

$$y_i = \theta + v_i, \qquad i = 1, \ldots, n.$$

Here θ is a scalar to be estimated and the noises v_i are independent Gaussian random variables with zero mean and unit variance. In this case the maximum-likelihood estimate is simply the average of the measurements:

$$\hat{\theta}_{\mathrm{ML}} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

In the space diffusion algorithm, each node maintains a state $x_i(t)$ initialized as $x_i(0) = y_i$. Then each node updates its state according to the equation

$$x_i(t+1) = x_i(t) + \sum_{j\in\mathcal{N}_i(t)} w_{ij}(t)\,(x_j(t) - x_i(t)) \qquad (3)$$

for $i = 1, \ldots, n$, where the weights $w_{ij}(t)$ are determined as

$$w_{ij}(t) = \frac{1}{1+\max\{d_i(t),\ d_j(t)\}}, \qquad \{i,j\} \in \mathcal{E}(t). \qquad (4)$$

We call these the *Metropolis weights*. Equivalently, each node updates its state with a weighted average of its previous value and the states at its instantaneous neighbors.

The following theorem concerning the convergence of (3) was established in [23]:

THEOREM 1. *If the set of communication graphs that occur infinitely often is jointly connected, then the states in the iteration (3) converge to the average of their initial values.*

For a finite collection of graphs with the same vertex set $\{(\mathcal{V}, \mathcal{E}_k)\}_{k=1}^{p}$, we say that they are *jointly connected* if the graph $(\mathcal{V}, \cup_{k=1}^{p} \mathcal{E}_k)$ is connected. Since the number of nodes n is finite, there can be only a finite number of possible communication graphs, and so there is at most a finite number of infinitely occurring communication graphs. An equivalent statement of the condition in Theorem 1 is that the graph $(\mathcal{V},\ \cup_{s\geq t}\mathcal{E}(s))$ is connected for all $t \geq 0$ [3].

Distributed average consensus belongs to a more general class of *distributed agreement* problems [20, 21], in which the final agreement value does not need to be the average. Algorithms for distributed consensus find applications in, e.g., distributed load balancing in parallel computers [7, 4], and multi-agent coordination and flocking [11, 14, 9, 13, 16].

3. SPACE-TIME DIFFUSION

As in the previous section, we first consider a scalar version of the P2PLS problem. Suppose all measurements taken by the sensors have the form

$$y_i(t) = \theta + v_i(t), \qquad t \in \mathcal{T}_i, \qquad i = 1, \ldots, n, \qquad (5)$$

where θ is a scalar to be estimated, and $v_i(t)$ are independent Gaussian random variables with zero mean and unit variance. We assume for now that all $|\mathcal{T}_i|$ are finite. In this case, the maximum-likelihood solution (1) is simply the average of all measurements,

$$\hat{\theta}_{\mathrm{ML}} = \frac{\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_i} y_i(t)}{\sum_{i=1}^{n}|\mathcal{T}_i|}. \qquad (6)$$

In the P2PLS problem, we want to asymptotically compute $\hat{\theta}_{\mathrm{ML}}$ at each node i by a distributed iterative algorithm.

The *space-time diffusion* scheme we propose is a nontrivial generalization of *space diffusion* with Metropolis weights. In this scheme, each node i maintains a scalar state $x_i(t)$, and keeps track of its *time degree* $d_i(t)$, the total number of measurements it has taken up to time t (the space degree is $|\mathcal{N}_i(t)|$). At $t = 0$, the initial states $x_i(0)$ are left unspecified. However, we initialize the time degrees by $d_i(-1) = 0$ (the negative index is for notational convenience). For $t = 0, 1, 2, \ldots$, each iteration of this algorithm includes two steps: *temporal update* and *spatial update*.

- **Temporal update** (also called *measurement update*). For each node $i = 1, \ldots, n$, if $t \in \mathcal{T}_i$, let

$$\begin{aligned} &d_i(t) = d_i(t-1) + 1 \\ &x_i(t_+) = (1 - 1/d_i(t))\,x_i(t) + (1/d_i(t))y_i(t); \end{aligned} \qquad (7)$$

  otherwise let

$$d_i(t) = d_i(t-1), \qquad x_i(t_+) = x_i(t). \qquad (8)$$

  Here we use the time label t_+ to signify the states after temporal update. We assume that temporal updates are finished instantaneously at each node, thus t_+ and t are essentially the same integer.

- **Spatial update.** *This step is executed at node i only if $d_i(t) > 0$, i.e., only if it has taken at least one measurement.* Each node communicates with its neighbors about their time degrees and calculates its own space-time degree

$$d_i^{\mathrm{ST}}(t) = d_i(t) + \sum_{j\in\mathcal{N}_i(t)} d_j(t). \qquad (9)$$

  Then it broadcasts its current state $x_i(t_+)$ and space-time degree to its neighbors. Based on received information from its neighbors, each node calculates the Metropolis weights

$$W_{ij}(t) = \min\left\{1/d_i^{\mathrm{ST}}(t), 1/d_j^{\mathrm{ST}}(t)\right\}, \quad \{i,j\}\in\mathcal{E}(t), \qquad (10)$$

  and updates its state using the equation

$$x_i(t+1) = x_i(t_+) + \sum_{j\in\mathcal{N}_i(t)} d_j(t) W_{ij}(t)(x_j(t_+) - x_i(t_+)). \qquad (11)$$

THEOREM 2. *Suppose every sensor makes a finite number of measurements. If the set of communication graphs that occur infinitely often is jointly connected, then the states in the space-time diffusion scheme all converge to the space-time average; i.e.,*

$$\lim_{t\to\infty} x_i(t) = \frac{\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_i} y_i(t)}{\sum_{i=1}^{n}|\mathcal{T}_i|}, \qquad i = 1, \ldots, n. \qquad (12)$$

We prove this theorem in the next section. The proof also shows how the space-time diffusion scheme is constructed.

4. CONVERGENCE WITH A FINITE NUMBER OF MEASUREMENTS

The convergence of the space-time diffusion scheme with a finite number of measurements can be proved by considering distributed average consensus on *space-time graphs*.

We construct the space-time graph at time t as follows. First note that the time-degree $d_i(t)$ is the total number of measurements taken at node i up to time t. In addition to the physical node i, we add $d_i(t) - 1$ virtual nodes at node i (see Figure 1). We associate each measurement with one of

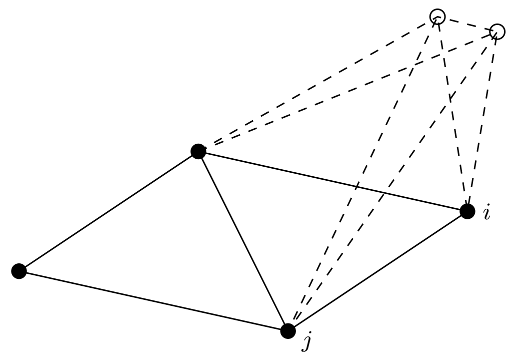

Figure 1: A simple space-time graph with two virtual nodes added at node i. The physical nodes are drawn as solid dots and the virtual nodes are drawn as small circles; The physical links are drawn as solid lines and the virtual links are drawn as dashed lines.

the $d_i(t)$ nodes. Let $\mathcal{V}_i(t)$ denote the set of these $d_i(t)$ nodes at node i. We connect every pair of nodes within the set $\mathcal{V}_i(t)$, thus forming a complete subgraph with $d_i(t)$ nodes. We denote this complete subgraph by $\mathcal{G}_i(t)$. Recall that $\mathcal{E}(t)$ is the set of active links of the communication graph. For every $\{i,j\} \in \mathcal{E}(t)$, we add the link $\{i',j'\}$ for every pair of $i' \in \mathcal{V}_i(t)$ and $j' \in \mathcal{V}_j(t)$, except the link $\{i,j\}$ which already exists. All these added links connect to one or two virtual nodes, and they are called *virtual links* (see Figure 1). We denote this space-time graph at time t by $\mathcal{G}^{\mathrm{ST}}(t)$.

In the space-time graph, each node is associated with a single measurement. Therefore we can allocate a state at every node (physical or virtual), initialize it by the associated measurement, and then conduct distributed average consensus on space-time graphs. More specifically, let $x_i^v(t)$ be the state associated with a node $v \in \mathcal{V}_i(t)$. The space diffusion algorithm (3) for distributed average consensus on space-time graphs has the form

$$x_i^v(t+1) = x_i^v(t) + \sum_{u \in \mathcal{V}_i(t)} W_{ii}(t)\Big(x_i^u(t) - x_i^v(t)\Big) + \sum_{j \in \mathcal{N}_i(t)} \sum_{u \in \mathcal{V}_j(t)} W_{ij}(t)\Big(x_j^u(t) - x_i^v(t)\Big), \quad (13)$$

where $W_{ii}(t)$ and $W_{ij}(t)$ are the Metropolis weights for the space-time graph $\mathcal{G}^{\mathrm{ST}}(t)$. To obtain these weights, we note that the degree of every node belonging to the set $\mathcal{V}_i(t)$ is

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

Applying the formula for Metropolis weights (4) yields the weights in (10). Note that every link between two different sets $\mathcal{V}_i(t)$ and $\mathcal{V}_j(t)$, physical or virtual, have the same Metropolis weight given in equation (10). For links within the complete subgraph $\mathcal{G}_i(t)$, the Metropolis weight is given by $W_{ii}(t) = 1/d_i^{\mathrm{ST}}(t)$.

Applying Theorem 1 to the equation (13) shows that every state $x_i^v(t)$ converges to the space-time average $\hat{\theta}_{\mathrm{ML}}$ given in (6) if the collection of space-time graphs that occur infinitely often is jointly connected. By our construction of the space-time graphs, this is equivalent to the condition that the collection of infinitely occurring physical graphs is jointly connected.

The major disadvantage of this basic averaging scheme is that it requires each physical node to keep track of multiple states, associated with additional virtual nodes. We have already seen how the constructed symmetry in the space-time graph reduces the number of different Metropolis weights: The weight on $\{u,v\}$ with $u \in \mathcal{V}_i(t)$ and $v \in \mathcal{V}_j(t)$ only depends on i and j, and not on the specific identification of the nodes u and v. Next we introduce a simple trick to make all states at the same physical node completely equivalent, so we only need to maintain a single copy of it.

Instead of conducting average consensus on the sequence of space-time graphs $\mathcal{G}^{\mathrm{ST}}(t)$, we insert a graph $\mathcal{G}^{\mathrm{T}}(t)$ immediately before every $\mathcal{G}^{\mathrm{ST}}(t)$. The graph $\mathcal{G}^{\mathrm{T}}(t)$ is the union of the complete subgraphs $\mathcal{G}_i(t)$ at all of the nodes. Now we conduct average consensus on the sequence of space-time graphs

$$\mathcal{G}^{\mathrm{T}}(0),\ \mathcal{G}^{\mathrm{ST}}(0),\ \mathcal{G}^{\mathrm{T}}(1),\ \mathcal{G}^{\mathrm{ST}}(1),\ \mathcal{G}^{\mathrm{T}}(2),\ \mathcal{G}^{\mathrm{ST}}(2),\ \ldots. \quad (14)$$

We note that the graph $\mathcal{G}^{\mathrm{T}}(t)$ contains only links $\{u,v\}$ such that u and v belong to the same set $V_i(t)$ for some i. Thus the iterations on the graphs $\mathcal{G}^{\mathrm{T}}(t)$ are implemented at each node independently; the last term in the equation (13) vanishes. The equation becomes

$$x_i^v(t+1) = x_i^v(t) + \sum_{u \in \mathcal{V}_i(t)} W'_{ii}(t)\Big(x_i^u(t) - x_i^v(t)\Big).$$

Moreover, since the subgraphs $\mathcal{G}_i(t)$ are complete and they are isolated from each other, we have $W'_{ii}(t) = 1/d_i(t)$, and the above equation further simplifies to

$$x_i^v(t+1) = \frac{1}{d_i(t)} \sum_{u \in \mathcal{V}_i(t)} x_i^u(t). \quad (15)$$

Thus all the states at the nodes in $\mathcal{V}_i(t)$ become identical, equal to their average, after this iteration. Therefore we only need to keep a single copy of the states afterward, which we denote by $x_i(t_+)$.

The next iteration is on the graph $\mathcal{G}^{\mathrm{ST}}(t)$. Substituting all $x_i^v(t)$ by $x_i(t_+)$ in equation (13), we see that the second term on the right hand side vanishes, and the equation simplifies to

$$x_i^v(t+1) = x_i(t_+) + \sum_{j \in \mathcal{N}_i(t)} d_j(t) W_{ij}(t)\Big(x_j(t_+) - x_i(t_+)\Big). \quad (16)$$

So it turns out that $x_i^v(t+1)$ are the same for $v \in \mathcal{V}_i(t)$. Thus we only need to keep a single copy of it and denote it by $x_i(t+1)$. We see that the above iteration (16) is precisely the spatial update (11). Now it is also clear that the iteration (15) on $\mathcal{G}^{\mathrm{T}}(t)$ is equivalent to the temporal update, being equation (7) with new measurement and equation (8) without new measurement. In summary, distributed average consensus on the sequence of graphs in (14) is equivalent to the space-time diffusion algorithm (7)-(11).

Finally, we note that the graphs that occur infinitely often in the sequence $\mathcal{G}^{\mathrm{ST}}(t)$ also occur infinitely often in the sequence (14). Therefore the same convergence condition on the physical graphs applies here.

5. P2PLS BY SPACE-TIME DIFFUSION

Now we explain how the space-time diffusion scheme can be used to compute $\hat{\theta}_{\mathrm{ML}}$ in the general setup described in §1.1. Throughout this section, we assume that the collection of infinitely occurring (physical) communication graphs is jointly connected.

5.1 The space-time diffusion scheme

As in the scalar case, this is a distributed iterative algorithm. In this scheme, each sensor node maintains a local *composite information matrix* $P_i(t) \in \mathbf{R}^{m \times m}$ and a local *composite information state* $q_i(t) \in \mathbf{R}^m$. Each node also keeps track of its time degree $d_i(t)$. At $t = 0$ let $d_i(-1) = 0$ and the values of $P_i(0)$ and $q_i(0)$ can be left unspecified. For $t = 0, 1, 2, \ldots$, each iteration consists of two steps: temporal update and spatial update.

- **Temporal update** (or *Measurement update*). For every node $i = 1, \ldots, n$, if $t \in \mathcal{T}_i$, let
$$d_i(t) = d_i(t-1) + 1,$$
and compute
$$P_i(t_+) = \left(1 - \frac{1}{d_i(t)}\right) P_i(t) + \frac{1}{d_i(t)} A_i(t)^T \Sigma_i(t)^{-1} A_i(t)$$
$$q_i(t_+) = \left(1 - \frac{1}{d_i(t)}\right) q_i(t) + \frac{1}{d_i(t)} A_i(t)^T \Sigma_i(t)^{-1} y_i(t);$$
Otherwise, let
$$d_i(t) = d_i(t-1), \quad P_i(t_+) = P_i(t), \quad q_i(t_+) = q_i(t).$$
- **Spatial update.** For every node i with $d_i(t) > 0$, compute the space-time degree using (9) and also the Metropolis weights W_{ij} using (10). Then it updates its local data using the equations
$$P_i(t+1) = P_i(t_+) + \sum_{j \in \mathcal{N}_i(t)} d_j(t) W_{ij}(t) (P_j(t_+) - P_i(t_+))$$
$$q_i(t+1) = q_i(t_+) + \sum_{j \in \mathcal{N}_i(t)} d_j(t) W_{ij}(t) (q_j(t_+) - q_i(t_+)).$$

In other words, this algorithm conducts space-time diffusion *entry-wise* for the information matrices $A_i(t)^T \Sigma_i(t)^{-1} A_i(t)$ and information vectors $A_i(t)^T \Sigma_i(t)^{-1} y_i(t)$. By Theorem 2, we have
$$\lim_{t \to \infty} P_i(t) = \frac{1}{\sum_{i=1}^n |\mathcal{T}_i|} \sum_{i=1}^n \sum_{t \in \mathcal{T}_i} A_i(t)^T \Sigma_i(t)^{-1} A_i(t),$$
$$\lim_{t \to \infty} q_i(t) = \frac{1}{\sum_{i=1}^n |\mathcal{T}_i|} \sum_{i=1}^n \sum_{t \in \mathcal{T}_i} A_i(t)^T \Sigma_i(t)^{-1} y_i(t).$$
Therefore, each node asymptotically computes the ML estimate (1) via
$$\hat{\theta}_{\mathrm{ML}} = \lim_{t \to \infty} P_i(t)^{-1} q_i(t), \qquad i = 1, \ldots, n.$$

This scheme has many nice properties that make it particularly attractive for distributed data fusion in ad hoc networks with dynamically changing topology. In particular, it doesn't involve explicit point-to-point message passing or routing. Instead, it iteratively diffuses information in the network by the temporal update and spatial update steps. Some key features of the algorithm are:

- *Universal data structure.* All the nodes keeps track of its time degree $d_i(t)$, and maintain the same, fixed data structure: $P_i(t) \in \mathbf{R}^{m \times m}$ and $q_i(t) \in \mathbf{R}^m$, which are independent of local measurement dimensions $m_i(t)$.
- *Isotropic protocol for communication and computing.* All the nodes communicate with their instantaneous neighbors, and use the same rule to compute space-time degree $d_i^{\mathrm{ST}}(t)$ and the Metropolis weights $W_{ij}(t)$.
- *Robust to link failures and topology changes.* With a finite number of measurements at every node, this scheme works provided the infinitely occurring communication graphs is jointly connected. As we will see in §6, the connectivity condition can be further weakened with an infinite number of measurements.

Other distributed algorithms have been proposed to solve the normal equation
$$\left(A^T \Sigma^{-1} A\right) \hat{\theta}_{\mathrm{ML}} = A^T \Sigma^{-1} y$$
directly by either iterative methods such as block Gauss-Seidel iteration [8], or by Gauss elimination [15]. However, they must wait to start until all measurements are taken, and may not enjoy all the nice properties above. The space-time diffusion scheme can start at a node whenever it has taken one measurement. Moreover, as we will show next, it allows every node to generate a useful intermediate estimate before convergence.

5.2 Properties of intermediate estimates

The true maximum likelihood estimate can be found at every node only in the limit as $t \to \infty$. In this section we study the properties of the intermediate estimates,
$$\hat{\theta}_i(t) = P_i(t)^{-1} q_i(t), \qquad i = 1, \ldots, n, \tag{17}$$
available at node i as soon as $P_i(t)$ is invertible. The following theorem shows that these intermediate estimates have some nice properties.

THEOREM 3. *Assume that the communication graph $\mathcal{G}(t)$, as a random variable, is independent of the measurement noise v. Then all the intermediate estimates $\hat{\theta}_i(t)$ are* unbiased *estimates of the true parameter, i.e.,*
$$\mathbf{E}\, \hat{\theta}_i(t) = \theta, \qquad i = 1, \ldots, n, \tag{18}$$
whenever $\hat{\theta}_i(t)$ is defined (i.e., when $P_i(t)$ is invertible); and the error covariance matrix at each node converges to that of the global ML solution, i.e.,
$$\lim_{t \to \infty} \mathbf{E}\left[(\hat{\theta}_i(t) - \theta)(\hat{\theta}_i(t) - \theta)^T\right] = \left(A^T \Sigma^{-1} A\right)^{-1}. \tag{19}$$
Here the expectation is taken with respect to both the measurement noise v and the sequence of random graphs $\{\mathcal{G}(t)\}_{t=0}^{\infty}$. These results also hold if the sequence of time-varying communication graphs is a given deterministic sequence.

This theorem has the same form as Theorem 3 in [23], which was proved for the space-diffusion case. Here the proof is omitted since it is essentially the same as in [23]. The only difference is that we should consider distributed average consensus on the constructed space-time graph. As in the space diffusion case, the intermediate estimates at each node can be interpreted as a *block-scaled* weighted least-squares solution. They all converge to the maximum likelihood solution as time goes to infinity.

6. CONVERGENCE WITH AN INFINITE NUMBER OF MEASUREMENTS

If the number of measurements $|\mathcal{T}_i|$ is infinite for some or all the nodes, then we need to show that all the local estimates converge to the true parameter, in the sense that the error covariance matrix converges to zero. To simplify the presentation, we focus on the scalar case in §3. In this case, we expect that every scalar state converges to the true parameter; i.e.,

$$\lim_{t\to\infty} x_i(t) = \theta, \qquad i = 1, \ldots, n.$$

This result looks like a *distributed* version of the law of large numbers. In fact, if every node takes an infinite number of measurements and they never communicate with each other, then the space-time diffusion scheme only has the temporal average step (7) at each iteration (in this case we let $x_i(t+1) = x_i(t_+)$). Then the law of large numbers can be directly applied at every node to show that its state $x_i(t)$ converges to the true parameter θ almost surely.

However, when there is interprocess communication, the spatial update step couples the states at different nodes, so we *cannot* use the (simple version of the) law of large numbers. One may argue that since the estimation error variance for a finite number of measurements is

$$\sigma^2_{\mathrm{ML}} = \frac{1}{\sum_{i=1}^n |\mathcal{T}_i|}$$

(assuming all measurements have unit variance), it converges to zero if some $|\mathcal{T}_i|$ goes to infinity. The flaw in this argument is that the above variance is *asymptotic* and it can only be achieved as time goes to infinity even if all $\mathcal{T}_i$ are finite. Nevertheless, we are able to show the following convergence result with an infinite number of measurements.

Theorem 4. *In the space-time diffusion scheme described in §3, let $d_{\mathrm{tot}}(t) = \sum_{i=1}^n d_i(t)$ be the total number of measurements taken at time t. If a node i satisfies the condition*

$$\liminf_{t\to\infty} \frac{d_i(t)}{d_{\mathrm{tot}}(t)} > 0, \tag{20}$$

Then the estimate $x_i(t)$ converges to θ in mean-square sense.

The proof of Theorem 4 is given in the appendix. The condition (20) is satisfied, e.g., if a node takes at least one measurement in every T steps for some $T > 0$. In this case, we have $\liminf_{t\to\infty} d_i(t)/d_{\mathrm{tot}}(t) > 1/T$. Note that in the above theorem, we do not assume the joint connectedness of the set of infinitely occurring communication graphs, because every node that has an infinite number of measurements has sufficient information for its estimate to converge to the exact parameter.

Theorem 4 does not tell the full story. In particular, it does not specify the convergence behavior of the rest of the nodes. We would expect that the state at every node that takes an infinite number of measurements converges to θ. For nodes that only take a finite number of measurements, we would expect their convergence with some additional assumptions on the communication graphs, similar to that in Theorem 1. Thus we have the following conjecture:

Conjecture 5. *Let $\mathcal{G}(\infty) = (\mathcal{V}, \cup_{s\geq t}\mathcal{E}(s))$. Suppose every connected component of $\mathcal{G}(\infty)$ that contains a node with a finite number of measurements also contains at least one node with an infinite number of measurements. Then every state $x_i(t)$ in the space-time diffusion scheme converges (in mean square) to the true parameter θ.*

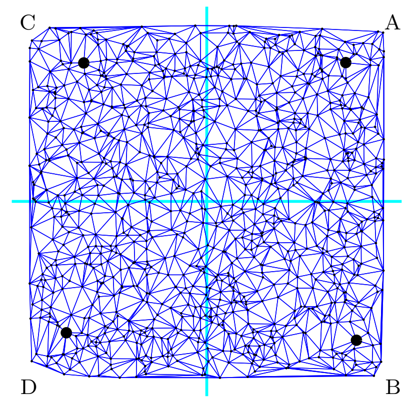

Figure 2: A sensor network with 1000 nodes. Nodes in the four square regions labeled A, B, C and D take measurements at different paces.

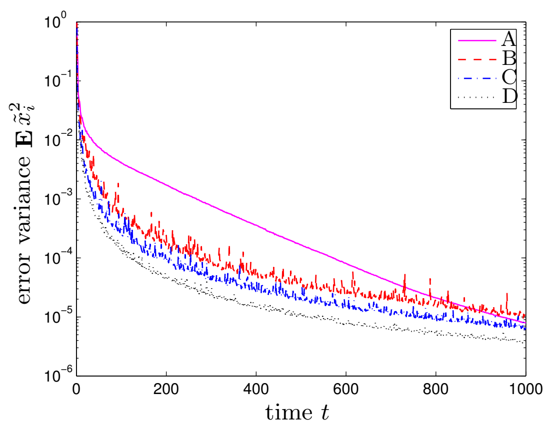

Figure 3: Estimation error variance at 4 nodes, each being a representative from the four regions A, B, C and D (big black dots in Figure 2).

We hope to prove the conjecture in future work. While the proof of convergence with an infinite number of measurements is more challenging than the case with a finite number of measurements, our goal in practice is often to obtain good intermediate estimates, rather than waiting for the asymptotic results. From this perspective, the space-time diffusion scheme has many nice properties. For example, the unbiased property of the intermediate estimates (18) is not affected by whether the sets $\mathcal{T}_i$ are finite or infinite, since there are only a finite number of measurements taken whenever an intermediate estimate is queried.

7. NUMERICAL EXAMPLES

Consider a sensor network with its communication graph shown in Figure 2. This graph is generated as follows. We first randomly generate $n = 1000$ sensor nodes, uniformly

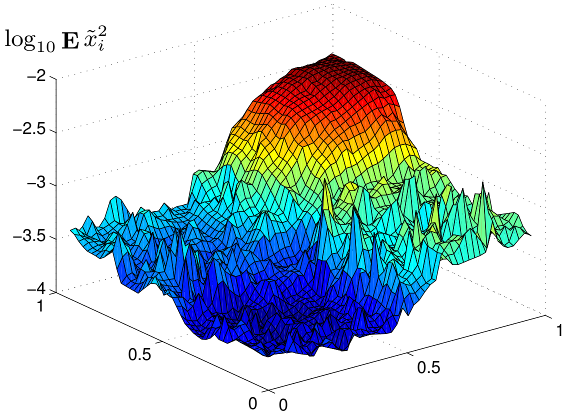

Figure 4: Error variance profile of the whole network at time $t = 100$.

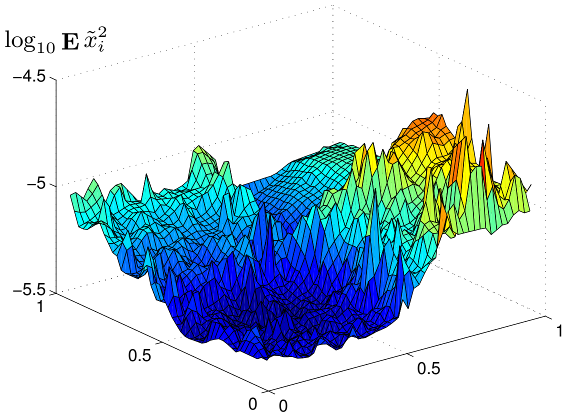

Figure 5: Error variance profile of the whole network at time $t = 1000$.

distributed on the unit square $[0, 1] \times [0, 1]$; then we use *Delaunay triangularization* to generate the edges [19] (such that no nodes are contained in any triangle's circumcircle). The graph has 1000 nodes and 2975 edges, with maximum degree 11 and minimum degree 3.

For simplicity, we use the scalar model in §3 to demonstrate the convergence behavior of space-time diffusion. In this case, each measurement is corrupted by a Gaussian noise with zero mean and unit variance; see equation (5). The estimation error at each node is

$$\tilde{x}_i(t) = x_i(t) - \theta, \qquad i = 1, \ldots, n$$

where $x_i(t)$ is the state at node i in the space-time diffusion scheme (7)-(11). We will focus on the error variances $\mathbf{E}\, \tilde{x}_i(t)^2$.

In our simulation, the communication graph is dynamically changing: at each time t, each edge is available with probability $1/2$, independent of other edges and all previous steps. We let every node take a measurement at time $t = 0$; then they take new measurements at different paces as follows. We partition the network in four square regions A, B, C and D, as shown in Figure 2. Nodes in region A do not take any more measurements for $t > 0$. At each time $t > 0$, every node in region B takes a new measurement with probability $1/4$, every node in region C does so with probability $1/2$, and every node in region D always takes a new measurement. All the random events are independent of each other.

Figure 3 shows the estimation error variances at four different nodes, each being a representative from the four regions (identified as big black dots in Figure 2). Figures 4 and 5 show the error variance profile of the whole network at time $t = 100$ and $t = 1000$, respectively. (These are interpolated surface plots that pass through the data points.) From these plots, we can see a clear tradeoff between the convergence speeds of temporal update and spatial update. In the beginning, new measurements account for most of the decrease in error variances through temporal update. Thus the four regions have quite different error variance profiles; the more frequently measurements are taken, the smaller the error variances are (see Figure 4). After a while, the smoothing effect of the spatial update becomes more obvious. For example, Figure 5 shows that the error variance profile of region A, where no measurements are taken after $t = 0$, is lower than that of regions B and C. This transition occurs at around $t = 900$, as indicated in Figure 3. We also notice that region A has a smoother profile than the other three regions, regardless of their average heights.

We should note that the tradeoff between the convergence speeds of temporal and spatial updates depends on the particular sensor network setup and parameters used. For example, if the communication graphs have rapid mixing properties (e.g., expander graphs) and the nodes take new measurements at a relatively slow pace, then the convergence of the spatial update can dominate the convergence of temporal update. In this case, the error variance profile can be rather flat at most times.

APPENDIX

In this appendix, we prove Theorem 4. Let's start by writing a single equation for the space-time diffusion scheme. The measurement update (7) and time update (11) equations can be combined as

$$\begin{aligned} x(t+1) = {} & \widetilde{W}(t) D(t)^{-1} D(t-1) x(t) \\ & + \widetilde{W}(t) \left(I - D(t)^{-1} D(t-1) \right) y(t), \end{aligned} \qquad (21)$$

where the matrix $\widetilde{W}(t)$ is given by

$$\widetilde{W}_{ij}(t) = \begin{cases} d_j(t)\ \min\left\{1/d_i^{\mathrm{ST}}(t),\ 1/d_j^{\mathrm{ST}}(t)\right\} & \text{if } \{i, j\} \in \mathcal{E}(t) \\ 1 - \sum_{k \in \mathcal{N}_i(t)} \widetilde{W}_{ik}(t) & \text{if } i = j \\ 0 & \text{otherwise,} \end{cases}$$

the matrix $D(t)$ is a diagonal matrix whose diagonal elements are the time degrees $d_i(t)$. The components of the vector $y(t)$ is the measurement $y_i(t)$ if $t \in \mathcal{T}_i$ and can be arbitrary if $t \notin \mathcal{T}_i$ (in this case $d_i(t) = d_i(t-1)$ and $1 - d_i^{-1}(t) d_i(t-1) = 0$). For convenience of later derivation, we set $y_i(t) = \theta$ if $t \notin \mathcal{T}_i$, which can be thought of as a measurement with zero noise).

Notice that $\widetilde{W}(t)$ is stochastic, thus $\theta \mathbf{1} = \widetilde{W}(t) \theta \mathbf{1}$. Subtracting $\theta \mathbf{1}$ from both sides of the equation (21) yields

$$\begin{aligned} x(t+1) - \theta\mathbf{1} = {} & \widetilde{W}(t) D(t)^{-1} D(t-1) (x(t) - \theta\mathbf{1}) \\ & + \widetilde{W}(t) \left(I - D(t)^{-1} D(t-1) \right) (y(t) - \theta\mathbf{1}). \end{aligned}$$

Letting $\tilde{x}(t) = x(t) - \theta\mathbf{1}$, the above equation becomes

$$\begin{aligned} \tilde{x}(t+1) &= \widetilde{W}(t)D(t)^{-1}D(t-1)\tilde{x}(t) \\ &+\widetilde{W}(t)\left(I - D(t)^{-1}D(t-1)\right)v(t), \end{aligned} \tag{22}$$

where $v_i(t)$ is the measurement noise if $t \in \mathcal{T}_i$ and zero otherwise. From the above equation, we further obtain

$$\tilde{x}(t) = \sum_{\tau=0}^{t-1} A(t-1)A(t-2)\cdots A(\tau+1)B(\tau)v(\tau),$$

where

$$\begin{aligned} A(t) &= \widetilde{W}(t)D(t)^{-1}D(t-1) \\ B(t) &= \widetilde{W}(t)\left(I - D(t)^{-1}D(t-1)\right). \end{aligned}$$

Since $A(t)$ and $B(t)$ are independent of the measurement noise $v(t)$, it is easy to show that $\mathbf{E}\ \tilde{x}(t) = 0$. In other words,

$$\mathbf{E}\, x_i(t) = \theta, \qquad i = 1, \ldots, n, \qquad t = 0, 1, \ldots.$$

Thus $x_i(t)$ is always an unbiased estimate of θ. This result holds regardless of the sizes of $\mathcal{T}_i$ and how frequent every node makes measurements.

Next we show that the error variances converge to zero; i.e.,

$$\lim_{t\to\infty} \mathbf{E}\, \tilde{x}_i^2(t) = 0, \qquad i = 1, \ldots, n.$$

First we need to introduce some notation. Let

$$\pi_i(t) = \frac{d_i(t)}{d_{\text{tot}}(t)}, \qquad i = 1, \ldots, n$$

and let $\Pi(t)$ denote the diagonal matrix whose diagonal entries are $\pi_1(t), \ldots, \pi_n(t)$. So we have $\Pi(t) = (1/d_{\text{tot}}(t))D(t)$. We define a new state vector

$$z(t) = \Pi(t-1)^{1/2}\ \tilde{x}(t) = \Pi(t-1)^{1/2}\ (x(t) - \theta\mathbf{1}),$$

and let

$$\begin{aligned} P(t) &= \Pi(t)^{1/2}\ \widetilde{W}(t)\ \Pi(t)^{-1/2} \\ u(t) &= \Pi(t)^{1/2}\ v(t). \end{aligned}$$

We note that the matrix $P(t)$ is symmetric because for $\{i, j\} \in \mathcal{E}(t)$,

$$\begin{aligned} P_{ij}(t) &= \pi_i(t)^{1/2}\widetilde{W}_{ij}(t)\pi_j(t)^{-1/2} \\ &= d_i(t)^{1/2}d_j(t)\ \min\left\{1/d_i^{\text{ST}}(t),\ 1/d_j^{\text{ST}}(t)\right\} d_j(t)^{-1/2} \\ &= d_i(t)^{1/2}d_j(t)^{1/2}\ \min\left\{1/d_i^{\text{ST}}(t),\ 1/d_j^{\text{ST}}(t)\right\} \\ &= P_{ji}(t). \end{aligned}$$

We have the following equation by left multiplying $\Pi(t)^{1/2}$ on both sides of the equation (22):

$$\begin{aligned} z(t+1) &= \sqrt{\frac{d_{\text{tot}}(t-1)}{d_{\text{tot}}(t)}}P(t)D(t)^{-1/2}D(t-1)^{1/2}z(t) \\ &+P(t)\left(I - D(t)^{-1}D(t-1)\right)u(t). \end{aligned}$$

Apparently $z(t)$ has zero mean. We consider its covariance matrix

$$Q(t) = \mathbf{E}\, z(t)z(t)^T,$$

which satisfies the equation

$$\begin{aligned} Q(t+1)\ =\ & \\ \frac{d_{\text{tot}}(t-1)}{d_{\text{tot}}(t)}&P(t)D(t)^{-1/2}D(t-1)^{1/2}Q(t)D(t-1)^{1/2}D(t)^{-1/2}P(t) \\ &+P(t)\left(I - D(t)^{-1}D(t-1)\right)\Sigma_u(t)\left(I - D(t)^{-1}D(t-1)\right)P(t) \end{aligned}$$

where $\Sigma_u(t) = \mathbf{E}\ u(t)u(t)^T$. Next we take the spectral norm of both sides of the above equation. Using standard inequalities for the spectral norm, we have

$$\begin{aligned} \|Q(t+1)\| \le \frac{d_{\text{tot}}(t-1)}{d_{\text{tot}}(t)}&\|P(t)\|^2 \left\|D(t)^{-1/2}D(t-1)^{1/2}\right\|^2 \|Q(t)\| \\ &+ \left\|I - D(t)^{-1}D(t-1)\right\|^2 \|P(t)\|^2 \|\Sigma_u(t)\|. \end{aligned}$$

We can simplify the above inequality by making the following observations. First, we have $\|P(t)\| = 1$ because $P(t)$ is symmetric and is similar to the stochastic matrix $\widetilde{W}(t)$. Second, we have $\left\|D(t)^{-1/2}D(t-1)^{1/2}\right\| \le 1$ because $d_i(t-1) \le d_i(t)$ for $i = 1, \ldots, n$. Therefore, we obtain the inequality

$$\|Q(t+1)\| \le \frac{d_{\text{tot}}(t-1)}{d_{\text{tot}}(t)}\|Q(t)\| + \left\|I - D(t)^{-1}D(t-1)\right\|^2 \|\Sigma_u(t)\|. \tag{23}$$

From this we can show that the sequence $\|Q(t)\|$ converges to zero. We need the following lemma (see, e.g., [2]).

Lemma 6. *Let Y_t, W_t, Z_t be three sequences such that W_t is nonnegative for all t. Assume that*

$$Y(t+1) \le Y(t) - W(t) + Z(t), \qquad t = 0, 1, \ldots,$$

and that the series $\sum_{t=0}^{T} Z(t)$ converges as $T \to \infty$. Then either $Y(t) \to -\infty$ or $Y(t)$ converges to a finite value and in the later case $\sum_{t=0}^{\infty} W(t) < \infty$.

In correspondence with the notation of the lemma, let

$$\begin{aligned} Y(t) &= \|Q(t)\| \\ W(t) &= \left(1 - \frac{d_{\text{tot}}(t-1)}{d_{\text{tot}}(t)}\right)\|Q(t)\| \\ Z(t) &= \left\|I - D(t)^{-1}D(t-1)\right\|^2 \|\Sigma_u(t)\|. \end{aligned}$$

Now let's take a closer look at the sequence

$$\left\|I - D(t)^{-1}D(t-1)\right\| = \begin{cases} 0 & \text{if } t \notin \cup_{i=1}^{n}\mathcal{T}_i \\ 1/\min\limits_{i:t\in\mathcal{T}_i} d_i(t) & \text{otherwise.} \end{cases}$$

Note that the possible values in the sequence are $0, 1, 1/2, 1/3, \ldots$, and each nonzero value can occur at most n times (they may not occur in order). Therefore,

$$\sum_{t=0}^{\infty} \left\|I - D(t)^{-1}D(t-1)\right\|^2 \le n\sum_{k=1}^{\infty}\frac{1}{k^2} < \infty.$$

By assumption $\|\Sigma_u(t)\|$ is bounded for all t. Thus by Lemma 6 the inequality (23) implies that the sequence $\|Q(t)\|$ converge to a finite value and

$$\sum_{t=1}^{\infty}\left(1 - \frac{d_{\text{tot}}(t-1)}{d_{\text{tot}}(t)}\right)\|Q(t)\| < \infty. \tag{24}$$

Let the limit of $\|Q(t)\|$ as $t \to \infty$ be c, which is obviously nonnegative. Suppose $c > 0$. Then there exist a T such that

$\|Q(t)\| \geq (1/2)c$ for all $t \geq T$. Therefore

$$\begin{aligned}\sum_{t=T}^{\infty}\left(1-\frac{d_{\mathrm{tot}}(t-1)}{d_{\mathrm{tot}}(t)}\right)\|Q(t)\| &\geq \frac{c}{2}\sum_{t=T}^{\infty}\left(1-\frac{d_{\mathrm{tot}}(t-1)}{d_{\mathrm{tot}}(t)}\right)\\ &\geq \frac{c}{2}\sum_{k=1}^{\infty}\frac{1}{d_{\mathrm{tot}}(T)+kn}\\ &= \infty.\end{aligned}$$

We have the last inequality by the observation that the sequence $1 - d_{\mathrm{tot}}(t-1)/d_{\mathrm{tot}}(t)$ is nonzero whenever $d_{\mathrm{tot}}(t)$ increases (new measurements are made), and it can increase at most by n in one step. This causes contradiction with equation (24). So we must have $c = 0$; i.e.,

$$\lim_{t\to\infty}\|Q(t)\| = 0.$$

In particular, as the diagonal entries of $Q(t)$, we have

$$\lim_{t\to\infty}\pi_i(t-1)\,\mathbf{E}\,\tilde{x}_i(t)^2 = 0, \qquad i = 1,\ldots,n.$$

For those nodes that take measurements most frequently, i.e., those satisfy

$$\liminf_{t\to\infty}\pi_i(t) = \liminf_{t\to\infty}\frac{d_i(t)}{d_{\mathrm{tot}}(t)} > 0,$$

we must have

$$\lim_{t\to\infty}\mathbf{E}\,\tilde{x}_i^2(t) = 0.$$

This completes the proof of Theorem 4.

A. REFERENCES

[1] M. Alanyali, S. Venkatesh, O. Savas, and S. Aeron. Distributed Bayesian hypothesis testing in sensor networks. In *Proceedings of American Control Conference*, pages 5369–5374, Boston, June 2004.

[2] D. P. Bertsekas and J. N. Tsitsiklis. Gradient convergence in gradient methods with errors. *SIAM Journal on Optimization*, 10(3):627–642, 2000.

[3] V. D. Blondel, J. M. Hendrickx, A. Olshevsky, and J. N. Tsitsiklis. Convergence in multiagent coordination, consensus, and flocking. Proceedings of IEEE Conference on Decision and Control, 2005.

[4] J. Boillat. Load balancing and Poisson equation in a graph. *Concurrency: Practice and Experience*, 2:289–313, 1990.

[5] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Gossip algorithms: design, analysis and applications. to appear in Proceedings of IEEE INFOCOM, 2005.

[6] J.-Y. Chen, G. Pandurangan, and D. Xu. Robust computation of aggregates in wireless sensor networks: distributed randomized algorithms and analysis. In *Proceedings of The 4th International Symposium on Information Processing in Sensor Networks (IPSN)*, pages 348–355, UCLA, Los Angeles, April 2005.

[7] G. Cybenko. Load balancing for distributed memory multiprocessors. *Journal of Parallel and Distributed Computing*, 7:279–301, 1989.

[8] 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, Berkeley, California, USA, April 2004.

[9] Y. Hatano and M. Mesbahi. Agreement over random networks. In *Proceedings of 43rd IEEE Conference on Decision and Control*, Atlantis, Bahamas, December 2004.

[10] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann, and F. Silva. Directed diffusion for wireless sensor networking. *IEEE/ACM Transactions on Networking*, 11:2–16, 2003.

[11] A. Jadbabaie, J. Lin, and A. S. Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions Automatic Control*, 48(6):988–1001, June 2003.

[12] Z.-Q. Luo. An isotropic universal decentralized estimation scheme for a bandwidth constrained ad hoc sensor network. *IEEE Journal on Selected Areas in Communications*, 2005. Special issue on Self-Organizing Distributed Collaborative Sensor Networks, to appear.

[13] L. Moreau. Stability of multiagent systems with time-dependent communication links. *IEEE Transactions on Automatic Control*, 50:169–182, 2005.

[14] R. Olfati-Saber and R. M. Murray. Consensus problems in networks of agents with switching topology and time-delays. *IEEE Transactions on Automatic Control*, 49(9):1520–1533, September 2004.

[15] K. Plarre and P. R. Kumar. Extended message passing algorithm for inference in loopy gaussian graphical models. *Ad Hoc Networks*, 2:153–169, 2004.

[16] W. Ren and R. W. Beard. Consensus seeking in multi-agent systems under dynamically changing interaction topologies. *IEEE Transactions on Automatic Control*, 2005. to appear.

[17] D. S. Scherber and H. C. Papadopoulos. Locally constructed algorithms for distributed computations in ad-hoc networks. In *Proceedings of 3rd International Symposium on Information Processing in Sensor Networks*, pages 11–19, Berkeley, CA, April 2004.

[18] D. P. Spanos, R. Olfati-Saber, and R. M. Murray. Approximate distributed Kalman filtering in sensor networks with quantifiable performance. In *Proceedings of The Fourth International Conference on Information Processing in Sensor Networks*, pages 133–139, Los Angeles, California, USA, April 2005.

[19] S.-H. Teng, C. W. Wong, and D. T. Lee. Unstructured mesh generation: theory, practice, and perspectives. *International Journal of Computational Geometry and Applicatons*, 10(3):227–266, 2000.

[20] J. N. Tsitsiklis. *Problems in Decentralized Decision Making and Computation*. PhD thesis, Massachusetts Institute of Technology, 1984.

[21] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Transactions on Automatic Control*, 31(9):803–812, September 1986.

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

[23] L. Xiao, S. Boyd, and S. Lall. A scheme for robust distributed sensor fusion based on average consensus. In *Proceedings of The 4th International Symposium on Information Processing in Sensor Networks (IPSN)*, pages 63–70, UCLA, Los Angeles, April 2005.