A General Distributed Consensus Algorithm for Wireless Sensor Networks

Jinho Choi, Shancang Li, Xinheng Wang
College of Engineering, Swansea University,
Swansea, SA2 8PP, UK
Email: {j.choi, s.li, xinheng.wang}@swansea.ac.uk

Jeongseok Ha
Deptment of Electrical Engineering,
KAIST, Korea
Email: jsha@ee.kaist.ac.kr

***Abstract*—In wireless sensor networks, distributed consensus algorithms can be employed for distributed detection. Each sensor node can compute its log-likelihood ratio (LLR) from local observations for a target event and using an iterative distributed algorithm, the average of sensors' LLRs can be available to all the sensor nodes. While the average of sensors' LLRs allows each sensor node to make a final decision as a decision statistic for an overall detection problem with all sensors' LLRs, it may be desirable if all sensors' LLRs or local observations, which form a full information vector and denoted by $\mathbf{x}$, could be available to each sensor for other purposes more than the detection of a target event. In this paper, we show that each sensor can have not only the average of local observations, but also full information vector, $\mathbf{x}$, (or its estimate) using a well-known iterative distributed algorithm. We extend the proposed approach to estimate $\mathbf{x}$ when $\mathbf{x}$ is sparse based on the notion of compressed sensing.**

I. Introduction

Wireless sensor networks (WSNs) have various applications including environmental monitoring and surveillance [1], [2]. In the context of distributed detection , the central unit is called a fusion center (FC) as all sensors' local decisions regarding a target event are collected and combined for a final decision, in which each node only makes a local decision. The detection performance at the FC depends on the number of sensors if local decisions can be correctly received at the FC. However, if local decisions from sensors are not reliably received at the FC due to wireless channel impairments, the channel conditions also affect on the performance. In general, the signals from sensor nodes that are far away from the FC may not be reliably received at the FC , which is not desirable as sensor nodes could have limited power sources. In the centralized approach with a FC, when the final decision is required at sensor nodes for further processing, the FC can broadcast it to sensors. However, in this case, the sensors that are far away may not reliably receive this final decision.

Distributed consensus algorithms (DCAs) [3], [4] can be employed for WSNs to overcome these problems. As DCAs only require local communications between neighbor sensor nodes, the transmission power could be lower than that in the centralized approach with a FC for distributed detection. Through iterative information exchanges between neighbor sensor nodes, the consensus or averaging can be achieved at all the sensor nodes. Each sensor can have the average of all sensors' log-likelihood ratios (LLRs) by an iterative distributed algorithm for distributed detection [3]. As each signal transmission for information exchange requires a certain amount of energy consumption, in general, iterative distributed algorithms of fast convergence rate are required [4], [5].

Suppose that x_l denotes the observation or local LLR at sensor l. DCAs can provide each sensor with the average of x_l's. In some applications, however, sensor nodes may need to know more than the average of sensors' LLRs or observations. In this case, DCAs could be modified. In this paper, we show that a well-known iterative distributed algorithm can be used to provide every sensors' observations to all sensor nodes without any modification. In particular, if the sensors' observations are sparse, compressed sensing based approaches [6], [7], [8] can be used to estimate spare sensors' observations with an iterative distributed algorithm. As a result, each sensor node can have the average of sensors' observations as well as an estimate of all sensors' observations.

II. Distributed Consensus Algorithms

In some WSNs, it is possible that each sensor can make a final decision as a FC does using DCAs [3], [4]. Suppose that a WSN is connected and the corresponding network topology is given by $G = (V, E)$, where $V = \{1, \ldots, L\}$ is the set of sensor nodes and $E = \{(l, m)\}$ is the set of edges. Here, L is the number of sensor nodes and an edge is the pair of two connected sensors that can communicate with each other directly.

Let N_l denote the set of the sensors that are connected with sensor l. Denote by x_l the observation at sensor l. In the context of distributed detection, x_l is the local LLR from observations of a target event at sensor l. Using a DCA, each node can have the average of x_l's. In this case, each sensor can make the final decision through an optimal decision rule as the overall LLR is available if x_l is conditionally independent.

DCAs are iterative algorithms that are based on information exchange between neighbor sensor nodes at each iteration. Let $x_l(t)$ denote the state variable at sensor l at time t with

978-1-4577-2194-6/12/$31.00 ©2012 IEEE

$x_l(0) = x_l$, where t is the index for iteration. Then, each sensor can have the average of x_l's by using the following iterative distributed algorithm [3]:

$$x_l(t+1) = x_l(t) + \mu \sum_{m \in N_l} (x_m(t) - x_l(t)) \quad (1)$$

in which $l = 1, \ldots, L$ and μ is the gain for the disagreement, $x_m(t) - x_l(t)$. We mainly consider the iterative distributed algorithm in (1) for distributed consensus or averaging in this paper. The convergence properties of the iterative distributed algorithm in (1) depends on the Laplacian matrix. The Laplacian matrix is given by

$$\mathbf{L} = \text{diag}(d_1, \ldots, d_L) - \mathbf{A},$$

where $\mathbf{A}$ is the adjacency matrix that is given by

$$[\mathbf{A}]_{l,m} = \begin{cases} 1, & \text{if } (l,m) \text{ or } (m,l) \in E; \\ 0, & \text{otherwise} \end{cases}$$

and d_l is the degree of node l, i.e., $d_l = |N_l|$. The iterative distributed algorithm is now rewritten as

$$\mathbf{x}(t+1) = \mathbf{x}(t) - \mu \mathbf{L}\mathbf{x}(t) = (\mathbf{I} - \mu \mathbf{L})\mathbf{x}(t) \quad (2)$$

where $\mathbf{x}(t) = [x_1(t) \ \ldots \ x_L(t)]^{\mathrm{T}}$ is the state vector. For a connected and undirected graph, $G = (V, E)$, the minimum eigenvalue of the Laplacian matrix, $\mathbf{L}$, is 0 and its corresponding eigenvector is $\frac{1}{\sqrt{L}}[1 \ \ldots \ 1]^{\mathrm{T}}$. Thus, the largest eigenvalues of $\mathbf{I} - \mu \mathbf{L}$ is 1, the magnitudes of the other eigenvalues are less than 1 if

$$0 < \mu < \frac{1}{\max_l d_l}. \quad (3)$$

Since the eigenvector corresponding to the largest eigenvalue, 1, of $\mathbf{I} - \mu \mathbf{L}$ is $\frac{1}{\sqrt{L}}[1 \ \ldots \ 1]^{\mathrm{T}}$, it can be shown that

$$\lim_{t \to \infty} x_l(t) = \frac{1}{L} \sum_{l=1}^{L} x_l = \bar{x}, \ l = 1, \ldots, L. \quad (4)$$

That is, $x_l(t)$ of each sensor node can reach the consensus or average of x_l's as $t \to \infty$.

Other distributed consensus algorithms are available from (1). With a weight matrix, $\mathbf{W}$, the iterative distributed algorithm in (1) can be generalized as

$$x_l(t+1) = W_{l,l} x_l(t) + \sum_{m \in N_l} W_{l,m} x_m(t), \ l = 1, \ldots, L, \quad (5)$$

where $W_{l,m} = W_{m,l}$, or $\mathbf{x}(t+1) = \mathbf{W}\mathbf{x}(t)$. In [4], the conditions for the convergence can be found. For the consensus over averaging in (4), a necessary condition for $\mathbf{W}$ is

$$\mathbf{W}\mathbf{1} = \mathbf{1} \ (\ \mathbf{1}^{\mathrm{T}}\mathbf{W} = \mathbf{1}^{\mathrm{T}} \text{ as } \mathbf{W} \text{ is symmetry}). \quad (6)$$

In addition to (6), if the following condition is satisfied,

$$\rho\left(\mathbf{W} - \frac{1}{L}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right) < 1, \quad (7)$$

where $\rho(\cdot)$ denotes the spectral radius of a matrix, the consensus or averaging in (4) can be achieved.

III. More Than Average

Using a DCA, distributed detection can be performed as the average of x_l's is available at every sensor nodes after some iterations. If we need more than the average of x_l's, distributed consensus algorithms may not be suitable or require some modification. For example, it would be required that each sensor is to know $\mathbf{x}$, which is referred to as the full information vector throughout the paper. A simple approach with the iterative distributed algorithm in (1) is based on making use of an iterative distributed algorithm L times. For the lth phase, the initial value of state variable is set to x_l at sensor l (i.e., $x_l(0) = x_l$) and set to zero at the other sensors (i.e., $x_m(0) = 0$ for $m \neq l$). In the lth phase, the average value becomes $\frac{x_l}{L}$ and after L phases, each sensor knows $\mathbf{x}$.

A. Main Result I: Full Information Acquisition

Let $\mathbf{D} = \mathbf{I} - \mu \mathbf{L}$ and $\mathbf{D}^m = (\mathbf{I} - \mu \mathbf{L})^m = [\mathbf{d}_{m,1}^{\mathrm{T}}, \mathbf{d}_{m,2}^{\mathrm{T}}, \cdots, \mathbf{d}_{m,L}^{\mathrm{T}}]^T$ i.e., $\mathbf{d}_{m,l}^{\mathrm{T}}$ is the lth row vector of $(\mathbf{I} - \mu \mathbf{L})^m$. For a undirected graph, $\mathbf{L}$ is symmetric. Thus, $\mathbf{d}_{m,l}$ is also the lth column vector of $(\mathbf{I} - \mu \mathbf{L})^m$. At sensor l, we have

$$\mathbf{z}_l = [x_l(0) \ x_l(1) \ \ldots \ x_l(M)]^{\mathrm{T}} = [\mathbf{d}_{0,l}^{\mathrm{T}}, \mathbf{d}_{1,l}^{\mathrm{T}}, \cdots, \mathbf{d}_{M,l}^{\mathrm{T}}]^{\mathrm{T}} \quad (8)$$

where $\mathbf{x} = \mathbf{x}(0) = [x_1 \ \ldots \ x_L]^{\mathrm{T}}$.

Lemma 1: The rank of the following matrix

$$\mathbf{C}_l = [\mathbf{d}_{0,l} \ \mathbf{d}_{1,l} \ \ldots \ \mathbf{d}_{M,l}] \quad (9)$$

is full if and only if the rank of the controllability matrix of $\{\mathbf{D}, \mathbf{b}_l\}$, $[\mathbf{b}_l \ \mathbf{D}\mathbf{b}_l \ \mathbf{D}^2\mathbf{b}_l \ \ldots \ \mathbf{D}^M\mathbf{b}_l]$ is full. Here, $\mathbf{b}_l$ stands for an $L \times 1$ column vector of zeros except the lth element whose value is 1, i.e., $\mathbf{b}_l$ is the lth standard basis vector.

Theorem 1: Sensor l can have $\mathbf{x}$ with $M = L-1$ iterations of (2) if the rank of the controllability matrix of $\{\mathbf{D}, \mathbf{b}_l\}$ is full.

In Theorem 1, sensors can have the full information vector, $\mathbf{x}$, not just the average of $\mathbf{x}$, $\bar{x}$, with $L - 1$ iterations of (1). Theorem 1 is referred to as the full information acquisition (FIA) process in this paper as it can provide $\mathbf{x}$ through the iterative distributed algorithm in (1). Since the applicability of the FIA process at sensor l depends on the rank of $\mathbf{C}_l$, we need to see more details for the conditions of full rank $\mathbf{C}_l$.

Lemma 2: If *i)* the elements of the lth row vector $\mathbf{E}$ are not zero and *ii)* $\{1 - \mu\lambda_l\}$ are distinct, $\mathbf{C}_l$ is full rank.

Consider a WSN in Fig. 1. The corresponding eigenvalues are $\{0, 2, 2, 4\}$. Thus, this Laplacian matrix does not satisfy the conditions in Lemma 2. Furthermore, it can be verified that $\mathbf{C}_l$ is not full rank ($rank(\mathbf{C}_l) = 3$ for all l). On the other hand, if we consider the case where the connection between nodes 1 and 2 disappears, all the elements of the eigenvectors of this Laplacian matrix are not zero and the eigenvalues are distinct. Thus, $\{\mathbf{C}_l\}$ are full rank for all l.

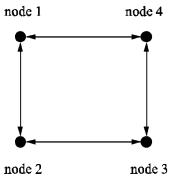

Fig. 1. An illustration of a WSN of 4 nodes as an undirected graph.

B. Main Result II: Partial Information Acquisition

If $\mathbf{C}_l$ is full rank, $\mathbf{x}$ is available to sensor l using the FIA process. However, if $\mathbf{C}_l$ is not full rank, the full information vector, $\mathbf{x}$, may not be available. Thus, it is important to find any limitations to extract $\mathbf{x}$ from $\mathbf{z}_l$, which are strongly dependent on the rank of $\mathbf{C}_l$.

Lemma 3: Suppose that the number of the distinct eigenvalues of $\mathbf{L}$, i.e., $\{\lambda_l\}$, is $L_d \leq L$. Then, we have $rank(\mathbf{C}_l) \leq \min(M+1, L_d)$.

The result in Lemma 3 shows that the multiplicity of the eigenvalues of the Laplacian matrix, $\mathbf{L}$, decides the rank of $\mathbf{C}_l$. For strongly regular graphs, the number of distinct eigenvalues of the Laplacian matrix is 3 , i.e., $L_d = 3$. However, we can have more than the average of $\mathbf{x}$ by using the FIA process. When $\mathbf{C}_l$ is not full rank, since $\mathbf{x}$ is not available, we may consider its estimate from the FIA process, which is referred to as the partial information vector. For example, from Theorem 1, if $\mathbf{C}_l^{\mathrm{T}}$ is rank deficient, we can have

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} ||\mathbf{z}_l - \mathbf{C}_l^{\mathrm{T}}\mathbf{x}||^2 = (\mathbf{C}_l^{\mathrm{T}})^{\dagger}\mathbf{z}_l \tag{10}$$

where the superscript $\dagger$ denotes the pseudo-inverse. In this case, $\hat{\mathbf{x}}$ is a partial information vector. Furthermore, $\hat{\mathbf{x}} = \bar{x}\mathbf{1}$ is also a partial information vector. The ratio of the rank of $\mathbf{C}_l$ to L, denoted by β_l, is referred to as the information acquisition ratio (IAR). The order of a partial information vector of $\mathbf{x}$, $\hat{\mathbf{x}}$, is defined as the rank of $\mathbf{T}$ if $\mathbf{x}$ and $\hat{\mathbf{x}}$ are related as follows: $\hat{\mathbf{x}} = \mathbf{T}\mathbf{x}$.

Thus, the order of a partial information vector, $\hat{\mathbf{x}}$, at sensor l is $\beta_l L$ and upper-bounded by L_d. As the order increases, the partial information vector $\hat{\mathbf{x}}$ can approach the full information vector $\mathbf{x}$.

In order to see what information we can extract from $\mathbf{z}_l$ as a partial information vector, we need to define new variables. From (8) and Lemma 3, we have

$$z_l(m) = \mathbf{v}_l^{\mathrm{T}}\mathbf{\Gamma}^m\mathbf{y} = \sum_{p=1}^{L} v_{l,p}\gamma_p^m y_p \tag{11}$$

where $\mathbf{y} = \mathbf{E}^{\mathrm{T}}\mathbf{x}$ and $y_p = \mathbf{e}_p^{\mathrm{T}}\mathbf{x}$. For convenience, let $\gamma_1 \geq \gamma_2 \geq \ldots \geq \gamma_L$. Clearly, $\gamma_1 = 1$ and $\mathbf{e}_1 = \frac{1}{\sqrt{L}}[1 \ \ldots \ 1]^{\mathrm{T}}$. As $m \to \infty$, we can confirm that

$$\lim_{m\to\infty} z_l(m) = \lim_{m\to\infty} \sum_{p=1}^{L} v_{l,p}\gamma_p^m y_p = \frac{1}{L}\sum_{j=1}^{L} x_j$$

Letting $\mathbf{u}_l = [(v_{l,1}y_1) \ \ldots \ (v_{l,L}y_L)]^{\mathrm{T}}$, from (11), we have

$$\mathbf{z}_l = \mathbf{Q}\ \mathbf{u}_l \tag{12}$$

where $\mathbf{Q}$ is a Vandermonde matrix . Let $M = L_d - 1$. Consider the set of the distinct eigenvalues $\{\bar{\gamma}_l,\ l = 1, \ldots, L_d\}$ from $\{\gamma_l,\ l = 1, \ldots, L\}$. That is, $\bar{\gamma}_j$ denotes the jth largest distinct eigenvalue among $\{\gamma_l,\ l = 1, \ldots, L\}$. Then, we have

$$\mathbf{z}_l = \bar{\mathbf{Q}}\bar{\mathbf{u}}_l, \tag{13}$$

where

$$[\bar{\mathbf{u}}_l]_p = \sum_{j\in\mathcal{M}_p} [\mathbf{u}_l]_j.$$

Here, $\mathcal{M}_p = \{l|\gamma_l = \bar{\gamma}_p\}$. Since $\bar{\mathbf{Q}}$ is full rank, we can obtain $\bar{\mathbf{u}}_l$ once $\mathbf{z}_l$ is available with $M = L_d - 1$ at sensor l.

Now, we will focus on the relation between $\bar{\mathbf{u}}_l$ and $\mathbf{x}$. Since $\mathbf{u}_l = \mathrm{diag}(\mathbf{v}_l)\mathbf{y}$, we have

$$\mathbf{u}_l = \mathrm{diag}(\mathbf{v}_l)\mathbf{E}^{\mathrm{T}}\mathbf{x}. \tag{14}$$

The relationship between $\bar{\mathbf{u}}_l$ and $\mathbf{u}_l$ is given by

$$\bar{\mathbf{u}}_l = \mathbf{J}\mathbf{u}_l, \tag{15}$$

where

$$[\mathbf{J}]_{p,l} = \begin{cases} 1, & l \in \mathcal{M}_p; \\ 0, & l \notin \mathcal{M}_p. \end{cases}$$

From (14) and (15), we have

$$\bar{\mathbf{u}}_l = \mathbf{P}_l\mathbf{E}^{\mathrm{T}}\mathbf{x}. \tag{16}$$

where $\mathbf{P}_l = \mathbf{J}\mathrm{diag}(\mathbf{v}_l)$ of size $L_d \times L$ plays a crucial role in deciding the rank of $\mathbf{C}_l$ as follows. Letting $\bar{\mathbf{u}}_l$ be the partial information vector, the order of $\bar{\mathbf{u}}_l$ is the rank of $\mathbf{P}_l$ (note that $\mathbf{E}$ is full rank). Thus, we can see that as the rank of $\mathbf{P}_l$ increases, $\bar{\mathbf{u}}_l$ can approximate $\mathbf{x}$ better.

Lemma 4: The rank of $\mathbf{C}_l$ when $M = L_d - 1$ is equal to the rank of $\mathbf{P}_l$ and $rank(\mathbf{P}_l) \geq 1$.

From (16), we can show that the minimum number of iterations for the consensus of (1) as follows.

Theorem 2: For a graph whose Laplacian matrix has L_d distinct eigenvalues, only $L_d - 1$ iterations of the iterative distributed algorithm in (1) are required to obtain $\bar{x}$ at every sensor nodes.

Theorem 3: For a connected and undirected graph G, the diameter, denoted by $\mathrm{diam}(G)$, is upper-bounded as $\mathrm{diam}(G) \leq L_d - 1$.

C. Learning Phase

Note that each sensor should know network topology or $\mathbf{L}$ to have $\mathbf{C}_l$. For a WSN, a learning phase could be required so that each sensor can have $\mathbf{C}_l$. We can modify (1) to allow that each sensor can have $\mathbf{C}_l$. From (8), we have

$$\mathbf{z}_l = \mathbf{C}_l^{\mathrm{T}}\mathbf{x}. \tag{17}$$

Here, we assume that $M = L - 1$. Suppose that the learning phase consists of L sub-phases. In each sub-phase, the sensors' initial state vector is different and denoted by $\mathbf{s}_p$ for the pth sub-phase. We assume that each sensor knows the L initial signal vectors, $\{\mathbf{s}_1, \ldots, \mathbf{s}_L\}$. Let $\mathbf{S} = [\mathbf{s}_1 \ \ldots \ \mathbf{s}_L]$. We assume that $\mathbf{S}$ has full rank (thus, $\mathbf{S}^{-1}$ exists). During the pth sub-phase, after $L-1$ iterations, $\mathbf{z}_l$ at sensor l is given by

$$\mathbf{z}_l = \mathbf{r}_{p,l} = \mathbf{C}_l^{\mathrm{T}} \mathbf{s}_p, \ p = 1, \ldots, L. \tag{18}$$

After L sub-phases, sensor l can have

$$\mathbf{R}_l = [\mathbf{r}_{1,l} \ \ldots \ \mathbf{r}_{L,l}] = \mathbf{C}_l^{\mathrm{T}} \mathbf{S}. \tag{19}$$

As $\mathbf{S}$ is known at each sensor, we have

$$\mathbf{C}_l^{\mathrm{T}} = \mathbf{R}_l \mathbf{S}^{-1}. \tag{20}$$

This learning phase requires $L(L-1)$ iterations of the iterative algorithm in (1) or (2). The significance of (20) is that each sensor does not need to know the Laplacian matrix, $\mathbf{L}$, or topology of the WSN to have $\mathbf{C}_l$. Furthermore, we only need to know $\mathbf{C}_l$ (not $\mathbf{L}$) for the FIA process in Theorem 1.

IV. Compressed Sensing for Sparse Samples

For some applications of WSNs, a fraction of sensor nodes may have some significant observations, while the others' observations could be insignificant . In this case, $\mathbf{x}$ becomes sparse. In this section, we consider the application of compressed sensing (CS) [7], [8] to estimate sparse $\mathbf{x}$ at each sensor node through the iterative distributed algorithm.

A. Compressed Sensing Problem

Suppose that $\mathbf{x}$ is an S-sparse vector. A vector is called S-spare if its elements are zero excepts at most S elements. From (8), based on [7], [9], we can formulate the following constrained ℓ_1-minimization problem:

$$\min_{\mathbf{x}} ||\mathbf{x}||_1 \quad s.t. \ \ \mathbf{z}_l = \mathbf{C}_l^{\mathrm{T}} \mathbf{x} \tag{21}$$

which is the sparse signal recovery problem from undersampled data when $rank(\mathbf{C}_l^{\mathrm{T}}) < L$. In CS, $\mathbf{C}_l^{\mathrm{T}}$ becomes the sensing matrix. Conditions to recover an S-sparse vector from (21) are found in [7], in particular, when the sensing matrix is random. From (13), (14), and (15), we have

$$\mathbf{z}_l = \bar{\mathbf{Q}} \mathbf{P}_l \mathbf{E}^{\mathrm{T}} \mathbf{x}. \tag{22}$$

Denote by ψ_i the ith column vector of $\mathbf{E}^{\mathrm{T}}$ (i.e., $\mathbf{\Psi} = [\psi_1 \ \psi_2 \ \ldots \ \psi_L] = \mathbf{E}^{\mathrm{T}}$). In CS, $\mathbf{y} = \mathbf{E}^{\mathrm{T}} \mathbf{x}$ can be considered a signal vector that can be expressed by a linear combination of a few basis vectors of $\{\psi_i\}$, which is called the representation basis. Using (13), the constraint in (22) can be replaced by

$$\bar{\mathbf{u}}_l = \bar{\mathbf{Q}}^{-1} \mathbf{z}_l = \mathbf{\Phi}_l^{\mathrm{T}} \mathbf{\Psi} \mathbf{x}, \tag{23}$$

where $\mathbf{\Phi}_l = [\phi_1 \ \phi_2 \ \ldots \ \phi_{L_d}] = \mathbf{P}_l^{\mathrm{T}}$. It is easy to see that the column vectors of $\mathbf{\Phi}_l$ are orthogonal. This means that $\mathbf{\Phi}_l$ is the (unnormalized) sensing basis. Since $\mathbf{v}_l = \mathbf{E}^{\mathrm{T}} \mathbf{b}_l = \mathbf{\Psi} \mathbf{b}_l = \psi_l$, the sensing basis, $\mathbf{\Phi}_l = \mathrm{diag}(\mathbf{v}_l) \mathbf{J}^{\mathrm{T}} = \mathrm{diag}(\psi_l) \mathbf{J}^{\mathrm{T}}$, is not independent of the representation basis, $\mathbf{\Psi}$. In order to have a good performance for the estimation of $\mathbf{x}$ by (21), it is usually required that the sensing and representation bases are incoherent. As most elements of a column vector of $\mathbf{J}$ are zero, the sensing basis is a spike basis. Thus, if $\mathbf{\Psi}$ is the Fourier basis, the incoherence between $\mathbf{\Psi}$ and $\mathbf{\Phi}_l$ could be maximized. The Laplacian matrix of a special class of graphs is a circulant matrix. For example, the Laplacian matrices of the ring and complete graphs are circulant. Since the eigenvectors of a circulant matrix are the column vectors of a discrete Fourier transform (DFT) matrix, $\mathbf{\Psi}$ of those graphs is the Fourier basis. In order to characterize the graphs whose $\mathbf{\Psi}$ is the Fourier basis, when $V = \{0, 1, \ldots, L-1\}$, we consider the following $2d$-regular graphs whose set of edges are given by

$$E = \{(u, k), \ k = \{(u + l)_L, \ l = \pm 1, \ldots, \pm d\}, \tag{24}$$

where $(x)_L$ denotes the modulo-L of x. Then, the Laplacian matrix of such a $2d$-regular graph is the circulant matrix whose first row vector is given by

$$[\mathbf{L}]_{1,:} = [(2d) \ \underbrace{-1 \ \ldots \ -1}_{d\text{-times}} \ 0 \ \ldots \ 0 \ \underbrace{-1 \ \ldots \ -1}_{d\text{-times}}]. \tag{25}$$

This graph is referred to as a circulant $2d$-regular graph. The eigenvalues of $\mathbf{L}$ are

$$\lambda_l = 2 \left(d - \sum_{k=1}^{d} \cos \left(\frac{2\pi}{L} kl \right) \right), \ l = 0, 1, \ldots, L-1. \tag{26}$$

We can see that the eigenvalues are not distinct. For example, $\lambda_l = \lambda_{L-l}$, $l = 1, 2, \ldots, \lfloor L/2 \rfloor$. This implies that the number of the distinct eigenvalues is bounded as $L_d \leq \lfloor L/2 \rfloor + 1$.

Consequently, for a circulant $2d$-regular graph, since $L_d \leq \lfloor L/2 \rfloor + 1$, the number of measurements is at most $\lfloor L/2 \rfloor + 1$. From [10, Eq. (7.45)], an upper-bound on S for the sparse signal recovery with a sufficiently high probability is given by

$$S \leq \alpha \frac{L}{2(\log L)^6}, \tag{27}$$

where $\alpha > 0$ is a small constant.

While circulant $2d$-regular graphs are deterministic and structured, most graphs of WSNs could be random. Since $\mathbf{D} = \mathbf{I} - \mu \mathbf{L}$, most elements of $\mathbf{D}$ will be zero or μ. Thus, the elements of $\mathbf{C}_l$ may not be sufficiently random although a graph can be random. In this case, the performance of the sparse signal recovery in (21) may not be satisfactory. To avoid this problem, the iterative distributed algorithm in (1) can be modified as (5) and the weights, $W_{l,m}$, could be random. Due to (6), we have

$$W_{l,l} = 1 - \sum_{m \in N_l} W_{l,m}. \tag{28}$$

Thus, $W_{l,m} = W_{m,l}$, $m \neq l$, can be randomly generated, but $W_{l,l}$ will be decided by (28). For achieving consensus or averaging, (7) should be satisfied. However, if the $W_{l,m}$'s are random, (7) would be satisfied with a certain probability. Since $\rho(\mathbf{W} - \frac{1}{L}\mathbf{1}\mathbf{1}^{\mathrm{T}}) < 1$ means that $\mathbf{s}^{\mathrm{T}}(\mathbf{W} - \frac{1}{L}\mathbf{1}\mathbf{1}^{\mathrm{T}})\mathbf{s} < 1$ for any $\mathbf{s}$ of unit-norm, assuming $\mathbf{s}$ is also random, we can find a condition for $W_{l,m}$'s that guarantees $\mathbb{E}[\mathbf{s}^{\mathrm{T}}(\mathbf{W} - \frac{1}{L}\mathbf{1}\mathbf{1}^{\mathrm{T}})\mathbf{s}] < 1$ as follows.

Theorem 4: Suppose that $W_{l,m}$, $(l,m) \in E$, are independent and identically distributed (iid) with $\mathbb{E}[W_{l,m}] = 0$ and $\mathbb{E}[|W_{l,m}|^2] = \sigma_W^2$. Let $\mathbf{s} = [s_1 \ \ldots \ s_L]^{\mathrm{T}}$ and assume that the s_l's are iid with $\mathbb{E}[s_l] = 0$ and $\mathbb{E}[|s_l|^2] = \frac{1}{L}$ (so that $\mathbb{E}[||\mathbf{s}||^2] = 1$). Furthermore, we assume that $\mathbb{E}[|s_l|^4] = \mu_4 < \frac{1}{L^2}$. If

$$\sigma_W^2 < \frac{1}{4\bar{d}}, \tag{29}$$

where $\bar{d} = \frac{\sum_{l=1}^{L} d_l}{L}$, we have

$$\mathbb{E}[\mathbf{s}^{\mathrm{T}}\left(\mathbf{W} - \frac{1}{L}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right)\mathbf{s}] < 1. \tag{30}$$

Since the condition in (29) only guarantees that $\mathbb{E}[\mathbf{s}^{\mathrm{T}}\left(\mathbf{W} - \frac{1}{L}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right)\mathbf{s}] < 1$, it would be a necessary condition for $\mathbf{s}^{\mathrm{T}}\left(\mathbf{W} - \frac{1}{L}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right)\mathbf{s} < 1$ for any $\mathbf{s}$ of $||\mathbf{s}|| = 1$.

B. Numerical Results

In this subsection, we present numerical results when the iterative distributed algorithm is used to estimate sparse signals. The number of iterations is set to $M = L_d - 1$. With $L = 100$, $2d$-regular graphs with Laplacian matrices in (25) are considered for simulations. Two different types of $\mathbf{x}$ are considered. For the first type, each element of $\mathbf{x}$ is independent and $x_l \sim \mathcal{N}(0,1)$. This type of $\mathbf{x}$, denoted by $\mathbf{x}_{(1)}$, is not sparse. The second type of $\mathbf{x}$, denoted by $\mathbf{x}_{(2)}$, is S-spare:

$$[\mathbf{x}_{(2)}]_l = \begin{cases} [\mathbf{x}_{(1)}]_l, & \text{if } |[\mathbf{x}_{(1)}]_l| \text{ is the largest signals} \\ 0, & \text{otherwise.} \end{cases} \tag{31}$$

For $\mathbf{x} = \mathbf{x}_{(1)}$, we use the ℓ_2-minimization to estimate $\mathbf{x}$, i.e., (10). For $\mathbf{x} = \mathbf{x}_{(2)}$, we use both the ℓ_2- and ℓ_1-minimization (i.e., (21)). The following normalized mean squared error (NMSE) is used as a performance measure:

$$\text{NMSE} = \frac{\mathbb{E}[||\hat{\mathbf{x}} - \mathbf{x}||^2]}{\mathbb{E}[||\mathbf{x}||^2]}.$$

Fig. 2 shows the NMSE values when both types of $\mathbf{x}$ are to be estimated using the iterative distributed algorithm for a WSN of circulant $2d$-regular graph. It is assumed that $L = 100$ and $d = 8$. Each NMSE value is obtained by averaging the NMSE values at $L = 100$ sensor nodes and 1000 runs (in each run, independent $\mathbf{x}_{(1)}$ and $\mathbf{x}_{(2)}$ are generated). The estimation of $\mathbf{x}$ by the ℓ_2-minimization in (10) is considered for both the first and second types of $\mathbf{x}$. It is shown that NMSE is independent of S when $\mathbf{x} = \mathbf{x}_{(2)}$ if the ℓ_2-minimization is used. On the other hand, for $\mathbf{x} = \mathbf{x}_{(2)}$ (i.e., sparse signals),

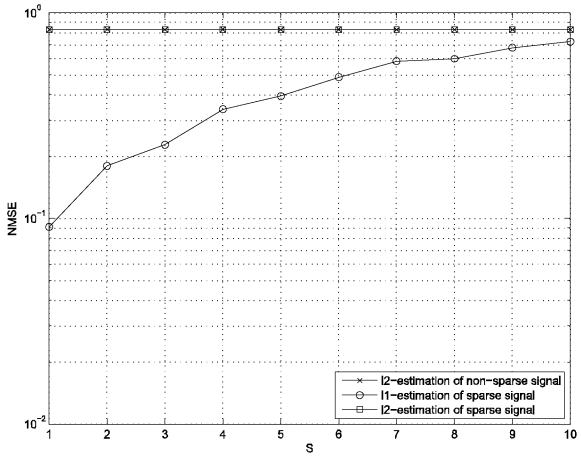

Fig. 2. NMSE versus S: ℓ_1- and ℓ_2-minimization to estimate $\mathbf{x}$ (a circulant 16-regular graph of $L = 100$ sensor nodes).

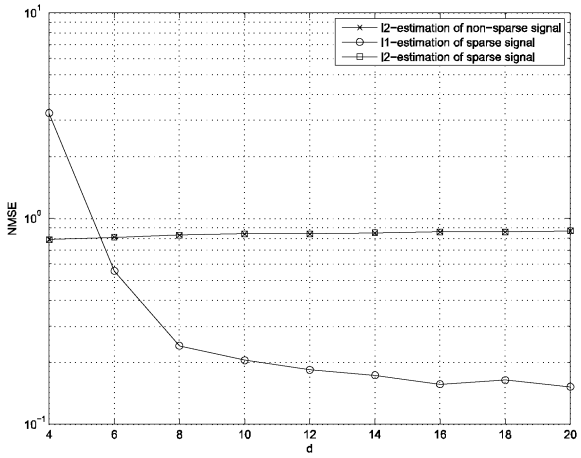

Fig. 3. NMSE versus d: ℓ_1- and ℓ_2-minimization to estimate $\mathbf{x}$ (a circulant $2d$-regular graph of $L = 100$ sensor nodes and $S = 4$).

the ℓ_1-minimization in (21) provides a lower NMSE than the ℓ_2-minimization and NMSE increases with S.

In order to see the impact of d, we consider simulations with circulant $2d$-regular graphs of $L = 100$ sensor nodes and $S = 4$ for sparse signals. The NMSE values obtained by simulations of 1000 runs are shown in Fig. 3. It is shown that if sensors are more connected (i.e., d increases), the ℓ_1-minimization can provide a better estimate of $\mathbf{x} = \mathbf{x}_{(2)}$ as the NMSE decreases with d. On the other hand, the performance of the ℓ_2-minimization is almost independent of d.

For WSNs with random graphs, we consider a normalized area of 1×1 with L sensors located randomly. Sensors are connected if they are within a range of $0 < R < 1$. If the graph is not connected, sensors' locations are re-generated randomly until the resulting graph is connected. This random graph is

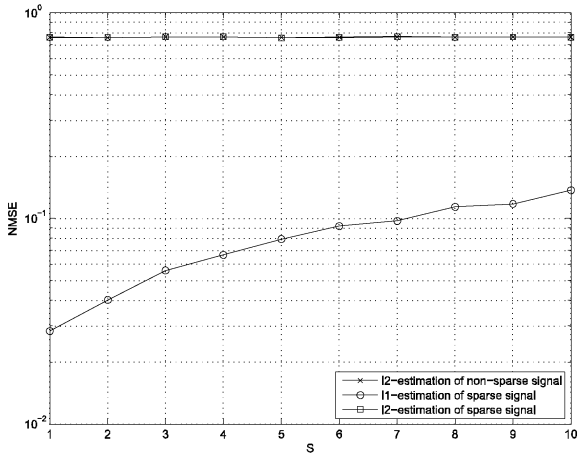

Fig. 4. NMSE versus S: ℓ_1- and ℓ_2-minimization to estimate $\mathbf{x}$ (random R-graphs of $L = 100$ sensor nodes and $R = 0.4$).

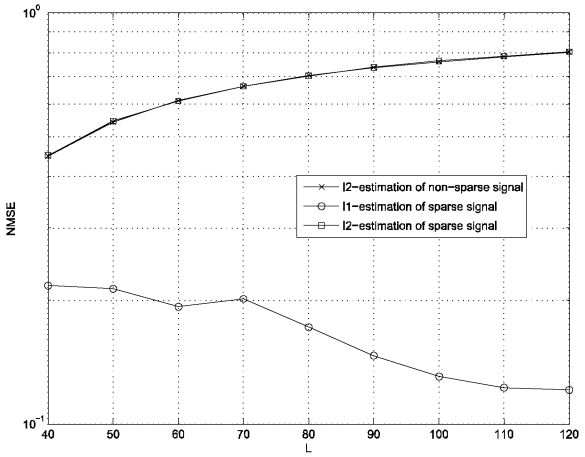

Fig. 5. NMSE versus L: ℓ_1- and ℓ_2-minimization to estimate $\mathbf{x}$ (random R-graphs with $R = 0.4$ and $S = L/10$).

referred to as a random R-graph. For simulations in Fig. 4 and 5, we assume that the elements of $\mathbf{W}$ for the iterative distributed algorithm are randomly generated with $\sigma_W^2 = \frac{1}{L}$. If $\mathbf{W} - \frac{1}{L}\mathbf{1}\mathbf{1}^{\mathrm{T}}$ has a spectral radius greater than 1, the weights, $W_{l,m}$, are re-generated until the spectral radius becomes less than 1. Fig. 4 shows the NMSE values when both types of $\mathbf{x}$ are to be estimated using the iterative distributed algorithm for random R-graphs of $R = 0.4$ and $L = 100$. Each NMSE value is obtained by averaging the NMSE values at $L = 100$ sensor nodes and 1000 runs. It is shown that the ℓ_1-minimization in (21) provides a lower NMSE than the ℓ_2-minimization and the NMSE increases with S. For example, when $S = 5$, the NMSE obtained from the ℓ_1-minimization is 10 times smaller than that obtained from the ℓ_2-minimization.

In order to see the impact of the number of sensor nodes in a WSN, NMSE is obtained with different values of L for random R-graphs of $R = 0.4$. It is assumed that S is proportional to L. That is, $S = L/10$. Simulation results are shown in Fig. 5. When $\mathbf{x}$ is sparse (i.e., $\mathbf{x} = \mathbf{x}_{(2)}$), the NMSE of the estimate obtained by the ℓ_1-minimization decreases with L. On the other hand, the NMSE of the estimate obtained by the ℓ_2-minimization increases with L for both $\mathbf{x}_{(1)}$ and $\mathbf{x}_{(2)}$.

V. Concluding Remarks

An iterative distributed algorithm was proposed in this works, which was originally derived for achieving consensus or averaging, to provide an estimate of the full information vector $\mathbf{x}$ at each sensor node. It was shown that the iterative distributed algorithm can provide not only the average of sensors' observations $\{x_l\}$, but also an estimate of $\mathbf{x} = [x_1 \ \ldots \ x_L]^{\mathrm{T}}$ under a certain condition. We have found that this condition is related to a controllability matrix and the order of a partial information vector, as an estimate of $\mathbf{x}$, depends on the number of distinct eigenvalues of a graph. From the latter result, we had two immediate results: i) the minimum number of iterations of the iterative distributed algorithm for achieving consensus is $L_d - 1$, where L_d is the number of distinct eigenvalues; ii) the diameter of a connected graph is upper-bounded by $L_d - 1$.

Acknowledgment

This work was partially supported by EPSRC-DSTL under grant No. $EP/H011919/1$.

References

[1] A. Swami, Q. Zhao, Y.-W. Hong, and L. Tong, *Wireless Sensor Networks*, John Wiley & Sons, 2007.

[2] Q. Zhao, A. Swami, and L. Tong, "The interplay between signal processing and networking in sensor networks," *IEEE Signal Proc. Mag.*, vol. 23, no. 4, pp. 84-93, July 2006.

[3] R. O. Saber and R. M. Murray, "Consensus protocols for networks of dynamic agents," in *Proc. IEEE American Control Conf.*, 2003, pp. 951-956.

[4] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems and Control Letters*, vol. 53, pp. 65-78, 2004.

[5] R. L. G. Cavalcante and B. Mulgrew, "Adaptive filter algorithms for accelerated discrete-time consensus," *IEEE Trans. Signal Proc.*, vol. 58(3), pp. 1049-1058, March 2010.

[6] J. Haupt, W. U. Bajwa, M. Rabbat, and R. Nowak, "Compressed sensing for networked data," *IEEE Signal Proc. Magazine*, vol. 25(2), pp. 92-101, March 2008.

[7] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. on Infor. Theory*, vol. 52(2), pp. 489-509, February 2006.

[8] D. Donoho, "Compressed sensing," *IEEE Trans. on Infor. Theory*, vol. 52(4), pp. 1289-1306, April 2006.

[9] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Proc. Mag.*, vol. 21, no. 3, pp. 21-30, March 2008.

[10] E. Candès and T. Tao, "Near-optimal signal recovery from random projections: universal encoding strategies," *IEEE Trans. on Infor. Theory*, vol. 52(12), pp. 5406-5425, December 2006.