A LEAST SQUARE APPROACH FOR DISTRIBUTED SENSOR FUSION IN BANDWIDTH-CONSTRAINED SENSOR NETWORKS

Shanying Zhu[1,2], *Jinming Xu*[3], *Cailian Chen**[1,2], *and Xinping Guan*[1,2]

[1]Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China
[2]Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai 200240, China
[3]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798
{shyzhu,cailianchen,xpguan}@sjtu.edu.cn, xuji0016@e.ntu.edu.sg

ABSTRACT

In this paper, we consider a simple model of distributed sensor fusion problem in sensor networks with asymmetric links, where the common goal is linear parameter estimation. For the realistic scenario of bandwidth-constrained networks, we propose a least square approach, based on distributed quantized consensus algorithms, to compute the ideal centralized sample mean estimate. Analytical results show that the proposed approach is effective in smearing out the quantization errors, and outperforms the centralized approaches with respect to the estimation performance. Simulation results are provided to validate the analytical results.

Index Terms— Distributed sensor fusion, bandwidth-constrained network, asymmetric links, consensus algorithm

1. INTRODUCTION

A fundamental problem in sensor networks is sensor fusion. In this paper we focus on a specific and simple model of a distributed sensor fusion problem in bandwidth-constrained sensor networks, where the common goal is linear parameter estimation based only on local information.

Consider a sensor network of n homogeneous nodes, each node makes observation of an unknown parameter $\theta \in \mathbb{R}$,

$$y_i = \theta + w_i,\ i = 1, 2, \ldots, n,$$

where w_i are zero mean, i.i.d. Gaussian noises with variance σ^2. If all these samples $\{y_i\}_{i=1}^n$ are collected by a fusion center perfectly, the optimal centralized sensor fusion scheme leads to the ideal sample mean estimate (ISME) $\hat{\theta} \triangleq (1/n)\sum_{i=1}^n y_i$ [1].

However, due to bandwidth limitations, the observations have to be quantized and estimation can only be made based on these quantized values. There are several works in the literature that investigate this problem in a centralized setting [2, 3]. Analytical results show that the performance is degraded by a factor with respect to the ISME $\hat{\theta}$ both for the uniform quantizer [2] and the probabilistic quantizer [3].

Recently, a number of distributed algorithms have been proposed to address the problem of sensor fusion with quantized communication [4–12]. However, most of the above works assume that the communication links between nodes are symmetric, which is in general not the case due to interference, packet collision, etc. Moreover, the results in [4–6,9] reveal that no matter which kind of static quantizers, there is always some gap between the local estimate and the ISME $\hat{\theta}$ unless some complicated dynamic mechanisms are adopted [7,10,11]. In the previous work [12], we proposed a two-stage averaging based distributed algorithm for sensor fusion over asymmetric and bandwidth-constrained sensor networks. The ISME $\hat{\theta}$ can be achieved at each node in mean square sense. But an increasing memory size of $O(n)$ is imposed, which limits its application in large-scale networks.

In [13], it is shown that distributed algorithms by exploiting additional information perform better than the centralized approaches with respect to adaption and learning over networks. Hence, a natural and interesting question is: Can distributed algorithms achieve the performance level of the ISME $\hat{\theta}$ while meeting the scalable requirement in the presence of bandwidth constraints? In the following, we will address this problem and give a positive answer. This is achieved by developing a least square approach based on consensus strategies.

2. PROBLEM FORMULATION

2.1. Network model

Let us consider a bandwidth-constrained sensor network consisting of n nodes that are linked via asymmetric links. The communication network over which nodes exchange data can then be represented by a *directed graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where

This research is partially funded by NSF of China under the grants U1405251, 61273181, 61221003 and 61521063.

$\mathcal{V} = \{1, 2, \ldots, n\}$ is the set of nodes, $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ denotes all the asymmetric links between nodes. Each node may transmit data to its *out-going neighbors* $\mathcal{N}_i^+ \triangleq \{j : (j, i) \in \mathcal{E}\}$ and receive data from *in-coming neighbors* $\mathcal{N}_i^- \triangleq \{j : (i, j) \in \mathcal{E}\}$. Let $d_i^+ = |\mathcal{N}_i^+|$ and $d_i^- = |\mathcal{N}_i^-|$ denote the *in-degree* and *out-degree*, respectively. We also define $\mathbf{A} \in \mathbb{R}^{n \times n}$ as the adjacency matrix of $\mathcal{G}$. In order to avoid any isolated nodes, we assume that $\mathcal{G}$ is *strongly connected*, i.e., each node can reach any other node via a directed path (possibly multi-hop).

For sensor networks with limited bandwidth, each node needs to quantize its data before transmission. We adopt the probabilistic quantization scheme used in [3–5, 12]. Specifically, each node quantizes the scalar data $x \in \mathbb{R}$ in a probabilistic way

$$\mathcal{Q}(x) = \begin{cases} \left\lceil \frac{x}{\Delta} \right\rceil \Delta, & \text{with probability } p_{x,\Delta}, \\ \left\lfloor \frac{x}{\Delta} \right\rfloor \Delta, & \text{with probability } 1 - p_{x,\Delta}, \end{cases} \tag{1}$$

where $p_{x,\Delta} = x/\Delta - \lfloor x/\Delta \rfloor$, $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ denote the floor and ceiling functions, respectively. It is well known that [4, 5]

$$\mathbb{E}\{\mathcal{Q}(x)\} = x, \;\; \mathbb{E}\left\{(\mathcal{Q}(x) - x)^2\right\} \le \Delta^2/4. \tag{2}$$

We can thus rewrite $\mathcal{Q}(x) = x + u$, where u denotes the quantization error satisfying $u \in [-\Delta, \Delta]$. As claimed in [4], the quantization scheme (1) is equivalent to a substractively dithered method. It is thus natural to assume that the quantization errors are *temporally independent* and *independent from the input*.

2.2. Distributed consensus algorithm for bandwidth-constrained networks

We let its measurement y_i as the initial guess of θ for each node i. Starting from $x_i(0) = y_i$, node i updates its variable iteratively,

$$x_i(t{+}1) = x_i(t) - \alpha d_i^+ \mathcal{Q}(x_i(t)) + \alpha \sum_{j \in \mathcal{N}_i^+} \mathcal{Q}(x_j(t)) + \beta s_i(t), \tag{3}$$

where $\alpha < 1/\max_i d_i^+$, $\beta > 0$ is a tuning parameter, and $s_i(t)$ is the surplus variable. Here $s_i(t)$ is introduced to compensate for the unidirectional effects of communication links, which is updated in the following way

$$\begin{aligned} s_i(t+1) = s_i(t) - \alpha d_i^- \mathcal{Q}(s_i(t)) + \alpha \sum_{j \in \mathcal{N}_i^+} \mathcal{Q}(s_j(t)) \\ - \big[x_i(t+1) - x_i(t)\big], \end{aligned} \tag{4}$$

with the initial condition $s_i(0) = 0, \forall i = 1, 2, \ldots, n$.

The recursions (3) and (4) form the basis of our distributed sensor fusion scheme for asymmetric sensor networks with limited bandwidth. Each node i keeps track of the state $x_i(t)$ along with the surplus variable $s_i(t)$ to locally record the state changes of individual nodes. This scheme is inspired by

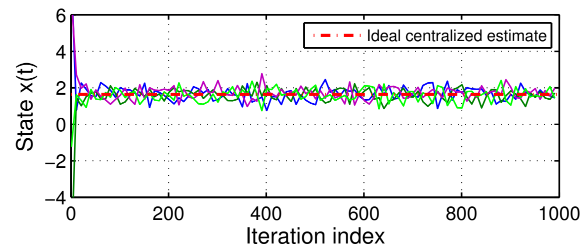

Fig. 1: States of algorithm (3) over a directed ring graph with $n = 4$ and $\Delta = 1$.

the average consensus algorithm introduced in [14] for multi-agent systems with infinite bandwidth. In order to implement (3) and (4), each node needs to know both its in-degree and out-degree. Such kind of information seems to be indispensable for distributed algorithms over networks with asymmetric links, e.g., [14–17]. We point out that the ISME $\hat{\theta}$ can be achieved at each node if the bandwidth is infinite; see [14, 17]. However, the quantization operation $\mathcal{Q}(\cdot)$ deteriorates the performance, causing the states of (3) to fluctuate around $\hat{\theta}$ with non-vanishing errors; see Fig.1 for an illustration.

Our goal in this paper is to *design* an appropriate approach to mitigate the quantization effects in (3) and (4) such that the ISME $\hat{\theta}$ can be asymptotically achieved at each node.

3. MITIGATING THE QUANTIZATION EFFECT: A LEAST SQUARE FORMULATION

One of the key properties of the quantization scheme (1) is that the quantization errors u's are zero-mean, temporally independent random noises. This temporal information has been used in [8, 12] to investigate the consensus seeking over symmetric and asymmetric networks, respectively. However, a slow convergence during the transient period is observed, which will consume much energy to achieve an estimate with desired accuracy.

In this paper, we adopt a different approach than [8, 12] to mitigate the quantization effects. Our main idea is motivated by the following result.

Theorem 1. *Assume that $\mathcal{G}$ is strongly connected, for sufficiently small $\beta > 0$, the state $\mathbf{x}(t)$ of (3) under the quantization scheme (1) converges to $\hat{\theta}$ in mean, i.e.,*

$$\lim_{t \to \infty} \mathbb{E}\{\mathbf{x}(t)\} = 0.$$

Proof. For each i, write $\mathcal{Q}(x_i(t)) = x_i(t) + u_i(t)$ and $\mathcal{Q}(s_i(t)) = s_i(t) + v_i(t)$, where $u_i(t)$, $v_i(t)$ are the quantization errors. Stack $x_i(t)$, $s_i(t)$, $u_i(t)$ and $v_i(t)$ into column vectors, we can express (3) and (4) in a compact form

$$\mathbf{z}(t+1) = \mathbf{P}\mathbf{z}(t) + \alpha \mathbf{L}_{\text{aug}} \mathbf{w}(t), \tag{5}$$

where $\mathbf{z}(t) = [\mathbf{x}(t)^T, \mathbf{s}(t)^T]^T$, $\mathbf{w}(t) = [\mathbf{u}(t)^T, \mathbf{v}(t)^T]^T$ and

$$\mathbf{P} \triangleq \begin{bmatrix} \mathbf{I} - \alpha\mathbf{L} & \beta\mathbf{I} \\ \alpha\mathbf{L} & (1-\beta)\mathbf{I} - \alpha\mathbf{L}^- \end{bmatrix}, \mathbf{L}_{\text{aug}} \triangleq \begin{bmatrix} -\mathbf{L} & \mathbf{0} \\ \mathbf{L} & -\mathbf{L}^- \end{bmatrix},$$

in which $\mathbf{L} \triangleq \text{diag}\{d_i^+\}_i - \mathbf{A}$ and $\mathbf{L}^- \triangleq \text{diag}\{d_i^-\}_i - \mathbf{A}$. By properties of (2), we have $\mathbb{E}\{\mathbf{u}(t)\} = \mathbb{E}\{\mathbf{v}(t)\} = \mathbf{0}$. Hence,

$$\mathbb{E}\{\mathbf{z}(t+1)\} = \mathbf{P}\mathbb{E}\{\mathbf{z}(t)\}. \tag{6}$$

Applying [14, Theorem 4] to (6) shows that for small $\beta > 0$, we have $\lim_{t\to\infty} \mathbb{E}\{\mathbf{x}(t)\} = \hat{\theta}\mathbf{1}$ and $\lim_{t\to\infty} \mathbb{E}\{\mathbf{s}(t)\} = \mathbf{0}$. Thus, $\lim_{t\to\infty} \mathbb{E}\{x_i(t)\} = \hat{\theta}$, $\forall i$, completing the proof. □

Theorem 1 states that the expectation $\mathbb{E}\{x_i(t)\}$ at each node i does converge to $\hat{\theta}$, although its realizations have a fluctuating behavior around $\hat{\theta}$ as observed in Fig. 1. This result motivates us to express the state $x_i(t)$ as

$$x_i(t) = \hat{\theta} + noise_i(t),\ \forall i = 1, 2, \ldots, n,$$

where $nosie_i(t)$ is the error capturing the fluctuation from the desired $\hat{\theta}$, which satisfies $\mathbb{E}\{noise_i(t)\} \to 0$ as $t \to \infty$. In this way, we may regard $x_i(t)$ as a noisy measurement of $\hat{\theta}$ at node i. Since no further distributional knowledge of $noise_i(t)$ is available, a popular approach to filter out $noise_i(t)$ in this case is the least square method [1]. It can be obtained by minimizing the following cost function

$$J_i(t) = \min_{\hat{x}\in\mathbb{R}} \sum_{s=0}^{t} b_s(x_i(s) - \hat{x})^2,$$

where $\{b_s\}_{s=0}^t$ are weighting factors that emphasize the contributions of the data $\{x_i(s)\}_{s=0}^t$. Following the procedures described in [1, Chap. 8], one obtains the recursive form

$$\hat{x}_i(t+1) = \hat{x}_i(t) + \gamma_t\big[x_i(t+1) - \hat{x}_i(t)\big], \tag{7}$$

where $\gamma_t > 0$ is the gain factor determined by $\{b_s\}_{s=0}^t$. To start the recursion, we let the initial guess $\hat{x}_i(0) = x_i(0)$, $\forall i$.

In the above framework of least square formuation, $\hat{x}_i(t)$ of (7) rather than $x_i(t)$ of (3) is adopted as the estimate of $\hat{\theta}$ at node i. This is different from the consensus algorithms in [14], where the focus is the convergence of $\mathbf{x}(t)$. We summarize the proposed Distributed Least Square Estimation (D-LSE) algorithm through consensus strategies for asymmetric bandwidth-constrained sensor networks in Algorithm 1, where its t-th iteration run by node i is presented.

4. CONVERGENCE PROPERTIES OF D-LSE ALGORITHM

In the section, we show that the proposed D-LSE algorithm turns out to be a convergent algorithm in the mean square sense. We need some preliminaries.

The next result establishes the mean square boundedness of the state $\mathbf{x}(t)$ of the algorithm (5). To this end, let us denote the error by $\mathbf{e}(t) \triangleq [\mathbf{x}(t)^T, \mathbf{s}(t)^T]^T - [\hat{\theta}\mathbf{1}^T, \mathbf{0}]^T$.

Algorithm 1 D-LSE algorithm

Input: $\alpha, \beta, \gamma_t, y_i$.
Output: $\hat{x}_i$.
1: **Initialization:** $x_i(0) = y_i$, $s_i(0) = 0$.
2: Receive data from its in-neighbors $\mathcal{N}_i^+$: $\mathcal{Q}(x_j(t))$, $\mathcal{Q}(s_j(t))$, $j \in \mathcal{N}_i^+$.
3: Update the states $x_i(t)$ and $s_i(t)$ following (3) and (4).
4: Update the estimate $\hat{x}_i(t)$ of $\hat{\theta}$ following (7).

Lemma 1. *Let $\mathcal{G}$ be strongly connected and $\beta > 0$ be sufficiently small, then there is a constant $c_{n,\mathbf{Q}} > 0$ depending only on n and $\mathbf{Q}$ so that*

$$\limsup_{t\to\infty} \mathbb{E}\{\|\mathbf{e}(t)\|^2\} \le \frac{n\alpha c_{n,\mathbf{Q}}\Delta^2\|\mathbf{L}_{aug}\|_2^2}{2}. \tag{8}$$

Proof. Let $\mathbf{Q} \triangleq \mathbf{P} - \mathbf{P}_\infty$, where $\mathbf{P}_\infty \triangleq \begin{bmatrix} \frac{\mathbf{1}\mathbf{1}^T}{n} & \frac{\mathbf{1}\mathbf{1}^T}{n} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$. Noting that the sum $\mathbf{1}^T(\mathbf{x}(t)+\mathbf{s}(t)) = \mathbf{1}^T\mathbf{x}(0)$ of (5) is preserved for all $t \ge 0$, and $\hat{\theta} = \mathbf{1}^T\mathbf{x}(0)/n$, we have

$$\mathbf{e}(t+1) = \mathbf{Q}\mathbf{e}(t) + \alpha\mathbf{L}_{\text{aug}}\mathbf{w}(t), \tag{9}$$

where we use the facts that $\mathbf{L}\mathbf{1} = \mathbf{0}$, $\mathbf{1}^T\mathbf{L}^- = \mathbf{0}$ and $\mathbf{Q}[\mathbf{x}(0)^T\mathbf{1}\mathbf{1}^T/n, \mathbf{0}]^T = \mathbf{0}$. It thus follows that

$$\begin{aligned} \mathbb{E}\{\|\mathbf{e}(t)\|^2\} &\overset{(a)}{=} \left\|\mathbf{Q}^t \begin{bmatrix}\mathbf{x}(0) \\ \mathbf{0}\end{bmatrix}\right\|^2 \\ &\quad + \alpha\sum_{k=0}^{t-1}\mathbb{E}\left\{\left\|\mathbf{Q}^{t-1-k}\mathbf{L}_{\text{aug}}\mathbf{w}(s)\right\|^2\right\} \\ &\overset{(b)}{\le} \left\|\mathbf{Q}^t\right\|_2^2\|\mathbf{x}(0)\|^2 + \frac{n\alpha\Delta^2\|\mathbf{L}_{\text{aug}}\|_2^2}{2}\sum_{k=0}^{t-1}\|\mathbf{Q}^k\|_2^2, \end{aligned} \tag{10}$$

where (a) is due to the independence assumption on the quantization errors $\mathbf{u}(t)$ and $\mathbf{v}(t)$, and (b) follows from the properties of the quantization scheme $\mathcal{Q}(\cdot)$.

By [14, Proposition 6], we can show that $\lim_{k\to\infty}\mathbf{P}^k = \mathbf{P}_\infty$ yielding $\rho(\mathbf{Q}) < 1$. Thus, by taking limits on both sides of (10) and using some known inequalities for matrix powers in [18, 19], one can obtain

$$\limsup_{t\to\infty} \frac{2\mathbb{E}\{\|\mathbf{e}(t)\|^2\}}{n\alpha\Delta^2\|\mathbf{L}_{\text{aug}}\|_2^2} \le c'_{n,\mathbf{Q}} + c''_{n,\mathbf{Q}}\sum_{k=n}^{\infty} k^{2(n-1)}\rho(\mathbf{Q})^{2k},$$

where $c'_{n,\mathbf{Q}}$ and $c''_{n,\mathbf{Q}}$ are two positive constants depending on n and $\mathbf{Q}$.

Following a similar argument as in [12], we have

$$\sum_{k=n}^{\infty} k^{2(n-1)}\rho(\mathbf{Q})^{2k} < \infty,$$

from which the lemma follows. □

Theorem 2. *Assume that $\mathcal{G}$ is strongly connected, and the gain factors $\{\gamma_t\}_{t\geq 0}$ satisfy*

$$\lim_{t\to\infty} \gamma_t = 0, \text{ and } \sum_{t=0}^{\infty} \gamma_t = \infty,$$

then for sufficiently small $\beta > 0$, the proposed D-LSE algorithm under the probabilistic quantization scheme (1) *converges to the centralized estimate $\hat{\theta}$ in mean square sense, i.e.,*

$$\lim_{t\to\infty} \mathbb{E}\{\|\hat{\mathbf{x}}(t) - \hat{\theta}\mathbf{1}\|^2\} = 0.$$

The proof is rather technical and due to page limitation, we do not provide it here. A sketch is as follows: We first introduce an auxiliary recursion of $\hat{s}_i(t)$ in the similar form of (7) to get an augmented system of $\hat{\mathbf{e}}(t) = [\hat{e}_1(t), \dots, \hat{e}_n(t)]^T$, where $\hat{e}_i(t) = [\hat{x}_i(t) - \hat{\theta}, \hat{s}_i(t)]^T, \forall i$. A recurrent inequality on $\mathbb{E}\{\|\hat{\mathbf{e}}(t)\|^2\}$ is then established by using Lemma 1, namely, for large t,

$$\mathbb{E}\{\|\hat{\mathbf{e}}(t+1)\|^2\} \leq (1-\gamma_t)^2 \mathbb{E}\{\|\hat{\mathbf{e}}(t)\|^2\} + \mathcal{O}(\gamma_t^2) + o(\gamma_t),$$

from which we can conclude that $\mathbb{E}\{\|\hat{\mathbf{e}}(t)\|^2\} \to 0$ as $t \to \infty$. The mean square convergence of $\mathbb{E}\{\|\hat{\mathbf{x}}(t)\|^2\}$ thus follows.

Remark 1. *Theorem 2 states that the gap between the local estimate $\hat{x}_i(t)$ and the centralized ISME $\hat{\theta}$ can be arbitrarily small, which implies that the distributed approach D-LSE would soon outperform the centralized approaches in [2, 3] for bandwidth-constrained sensor networks.*

5. SIMULATION RESULTS

We consider a random network of $N = 20$ nodes to monitor an unknown $\theta = 8$. Each node observes $y_i = \theta + w_i$, where w_i is the Gaussian noise with variance 4. To simulate certain pairs of asymmetric links between nodes, we first generate an undirected network using the random geometric graph model: Nodes are placed uniformly at random over $[0, 1] \times [0, 1]$ and two are connected by two unidirectional links if the distance is less than $\sqrt{\log N/N}$. We then remove 30% of the unidirectional links to generate a asymmetric network. We put $\alpha = 1/(1 + \max_i d_i^+)$ and $\beta = 0.2$ for (3), and $\gamma_t = \sqrt{0.5 \log\log(t+10)}/t$ for (7). It is easy to check that such γ_t satisfies the conditions of Theorem 2. The quantization step-size is set $\Delta = 1$.

Fig. 2 depicts the intermediate state $\mathbf{x}(t)$ of (3), and the estimate $\hat{\mathbf{x}}(t)$ generated by the proposed D-LSE algorithm. It is clear that D-LSE has the ability of mitigating the quantization effect, achieving the centralized ISME $\hat{\theta}$ asymptotically. This corroborates the theoretical result given in Theorem 2. The performance of D-LSE is measured by the mean square error $(1/n)\sum_{i=1}^{n}(\hat{x}_i(t) - \hat{\theta})^2$, which is averaged over 100 independent runs. Fig. 3 shows the mean square errors of different algorithms, from which we can easily identify that our D-LSE algorithm achieves a superior estimation performance to other distributed algorithms. Moreover, it outperforms the centralized approaches QSME and DES within a moderate number of iterations, which is a direct implication of Theorem 2. This shows the effectiveness of the least square approach of our D-LSE algorithm in solving the distributed sensor fusion problem in sensor networks with bandwidth constraints.

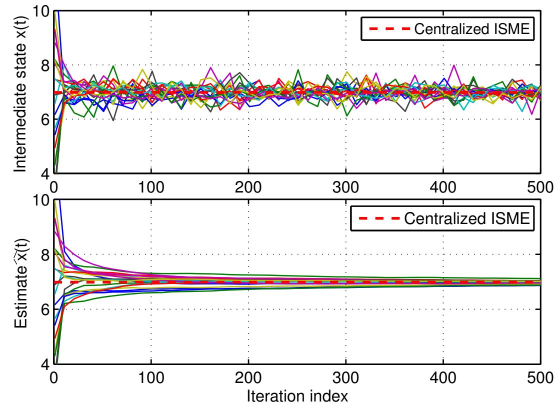

Fig. 2: Intermediate states $\mathbf{x}(t)$ of algorithm (3) (top) and estimates $\hat{\mathbf{x}}(t)$ generated by D-LSE (bottom) with $\Delta = 1$.

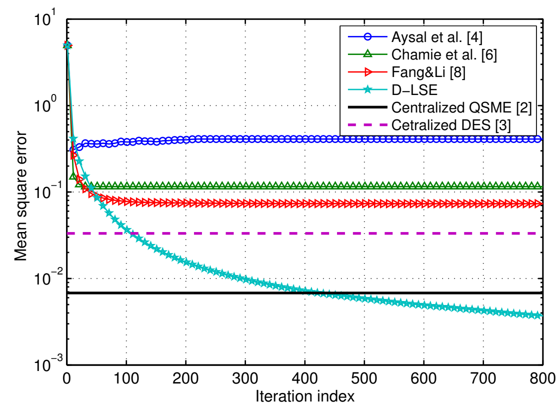

Fig. 3: Comparison results of mean square error between different algorithms with respect to the centralized estimate $\hat{\theta}$.

6. CONCLUSIONS

In this paper, we proposed a least square approach for distributed sensor fusion in bandwidth-constrained sensor networks with asymmetric links. In the scheme, a surplus variable is introduced to compensate for asymmetric links, and all the nodes only communicate with their one-hop neighbors, and keep track of local states. A least square reformulation of these local states was then performed to generate local estimates. It is shown that all such estimates converge to the ideal centralized estimate in mean square sense. We also provide some simulation studies to validate the theoretical results.

7. REFERENCES

[1] S. M. Kay, *Fundamentals of Statistical Processing: Estimation Theory*, Prentice Hall, Upper Saddle River, NJ, 1993.

[2] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks-Part I: Gaussian case," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1131–1143, 2006.

[3] J.-J. Xiao and Z.-Q. Luo, "Decentralized estimation in an inhomogeneous sensing environment," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3564–3575, 2005.

[4] T. C. Aysal, M. J. Coates, and M. G. Rabbat, "Distributed average consensus with dithered quantization," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4905–4918, 2008.

[5] R. Carli, F. Fagnani, P. Frasca, and S. Zampieri, "Gossip consensus algorithms via quantized communication," *Automatica*, vol. 46, no. 1, pp. 70–80, 2010.

[6] M. El Chamie, J. Liu, and T. Başar, "Design and analysis of distributed averaging with quantized communication," in *Proc. 53rd IEEE Conf. Decision and Control(CDC)*, Los Angeles, USA, Dec. 2014, pp. 3860–3865.

[7] J. Fang and H. Li, "An adaptive quantization scheme for distributed consensus," in *Proc. 34th IEEE Int. Conf. Acoust., Speech, Signal Process.(ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 2777–2780.

[8] J. Fang and H. Li, "Distributed consensus with quantized data via sequence averaging," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 944–948, 2010.

[9] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks: Quantized data and random link failures," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1383–1400, 2010.

[10] T. Li, M. Fu, L. Xie, and J.-F. Zhang, "Distributed consensus with limited communication data rate," *IEEE Trans. Autom. Control*, vol. 56, no. 2, pp. 279–292, 2011.

[11] D. Thanou, E. Kokiopoulou, Y. Pu, and P. Frossard, "Distributed average consensus with quantization refinement," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 194–205, 2013.

[12] S. Zhu, Y. C. Soh, and L. Xie, "Distributed parameter estimation with quantized communication via running average," *IEEE Trans. Signal Process.*, vol. 63, no. 17, pp. 4634–4646, 2015.

[13] A.H. Sayed, S.-Y. Tu, J. Chen, X. Zhao, and Z. J. Towfic, "Diffusion strategies for adaptation and learning over networks: An examination of distributed strategies and network behavior," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 155–171, 2013.

[14] K. Cai and H. Ishii, "Average consensus on general strongly connected digraphs," *Automatica*, vol. 48, no. 11, pp. 2750–2761, 2012.

[15] M. Franceschelli, A. Giua, and C. Seatzu, "Distributed averaging in sensor networks based on broadcast gossip algorithms," *IEEE Sensors J.*, vol. 11, no. 3, pp. 808–817, 2011.

[16] F. Iutzeler, P. Ciblat, and W. Hachem, "Analysis of sum-weight-like algorithms for averaging in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 61, no. 11, pp. 2802–2814, 2013.

[17] S. Wu and M. G. Rabbat, "Broadcast gossip algorithms for consensus on strongly connected digraphs," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 3959–3971, 2013.

[18] M. Gil, "Estimate for the norm of matrix-valued functions," *Linear and Multilinear Algebra*, vol. 35, no. 1, pp. 65–73, 1993.

[19] P. Henrici, "Bounds for iterates, inverses, spectral variation and fields of values of non-normal matrices," *Numer. Math.*, vol. 4, no. 1, pp. 24–40, 1962.