DISTRIBUTED ESTIMATION IN THE PRESENCE OF CORRELATED NOISES

Zhiyuan Weng and Petar M. Djurić

Department of Electrical and Computer Engineering
Stony Brook University, Stony Brook, NY 11790, USA
E-mails: {zhiyuan.weng, petar.djuric}@stonybrook.edu

ABSTRACT

We propose a method for distributed sequential estimation in the presence of correlated noises. At each time slot, a node exchanges information with its neighbors and then updates the estimate by using the received information from its neighbors and its local observation. It is assumed that the noises have the Markov property with respect to the network topology. A doubly stochastic matrix for combining information from the nodes is employed to average the sufficient statistics over the network. We show that the performance of the proposed method converges to that of the centralized optimal estimator as the iterations go on. Therefore, our algorithm approaches the Cramér-Rao bound asymptotically.

***Index Terms*—** Correlated noises, distributed estimation, least squares estimator, sequential estimation

1. INTRODUCTION

Methods for distributed processing of data have been attracting scholars from different areas for decades [1]. These methods have been of interest in various areas including wireless sensor networks, agent networks, and social learning. In this paper, we focus on a sequential learning problem, where the parameters of interest are static and at every time slot each node obtains local observations with information about the static parameters. The objective is to estimate the parameters in a distributed way. This problem has been studied in [2, 3, 4, 5, 6, 7, 8, 9, 10, 11]. In [7], the authors compare the mean-square performance of two main strategies for distributed estimation: consensus strategies and diffusion strategies. They claim that diffusion leads to faster convergence and lower mean-square deviation than consensus. Note that when the parameters of interest are dynamic, the problem becomes sequential filtering, which is another popular category of distributed estimation problems. See [12, 13, 14], for example. In this paper, we propose an efficient estimation algorithm for the case where the noises are correlated. This algorithm is neither diffusion nor consensus, but it is closer in spirits to the latter. We use a doubly stochastic matrix to combine the information from different nodes at each iteration. We prove that our algorithm approaches the optimal centralized least squares estimator asymptotically.

The paper is organized as follows. The problem is formulated in Section 2. The proposed algorithm is described in Section 3 and is analyzed in Section 4. Section 5 provides simulation results, and Section 6 concludes the work.

We use the following notation: det(A) is the determinant of a matrix A; tr(A) is the trace of A; O is a matrix with all entries equal to zero; 1_M and 0_M refer to column vectors of size $M \times 1$ with all entries equal to 1 and 0, respectively; $\lambda_i(A)$ signifies the ith eigenvalue of A, and $\lambda_{\min}(A)$ is the minimum eigenvalue of A; I_M is the identity matrix with size $M \times M$; $A^\top$ is the transpose of the matrix A; $\mathcal{N}_i$ represents the set of nodes that are neighbors of node i; $\mathbb{N}$ is the set of natural numbers; $\|A\|$ means the Frobenius norm of A,

$$\|A\| = \sqrt{\text{tr}(AA^\top)}; \tag{1}$$

$\delta_{i,j}$ is the Kronecker delta and $\otimes$ is the Kronecker product.

2. PROBLEM FORMULATION

The problem is mathematically formulated as follows. The network is represented by $G = (V, E)$, where V and E are the sets of nodes and edges, respectively. Two nodes exchange information with each other only if there is an edge between them. We assume there are N nodes in the network, namely $N = |V|$. Let $\theta \in \mathbb{R}^{L\times 1}$ be a parameter vector of interest. At time instant t, the observation at node i is modeled as

$$y_{i,t} = H_{i,t}\theta + w_{i,t}, \tag{2}$$

where $w_{i,t}, y_{i,t} \in \mathbb{R}^{M\times 1}$, $H_{i,t} \in \mathbb{R}^{M\times L}$; $H_{i,t}$ is the observation matrix; $y_{i,t}$ is the observation; and $w_{i,t}$ is a Gaussian noise vector, where

$$\text{E}[w_{i,t}] = 0_M \tag{3}$$

$$\text{E}[w_{i,t}w_{j,s}^\top] = \delta_{t,s}\Sigma_{i,j}, \tag{4}$$

and $H_{i,t}$ is full rank. Let

This work was supported by NSF under Award CCF-1018323.

$$H_t = \begin{bmatrix} H_{1,t} \\ \vdots \\ H_{N,t} \end{bmatrix}, \quad y_t = \begin{bmatrix} y_{1,t} \\ \vdots \\ y_{N,t} \end{bmatrix}, \quad w_t = \begin{bmatrix} w_{1,t} \\ \vdots \\ w_{N,t} \end{bmatrix}, \tag{5}$$

and

$$\Sigma = \begin{bmatrix} \Sigma_{1,1} & \cdots & \Sigma_{1,N} \\ \vdots & \ddots & \vdots \\ \Sigma_{N,1} & \cdots & \Sigma_{N,N} \end{bmatrix}. \tag{6}$$

The matrix Σ is assumed to be strictly positive definite. Then the entire model can be expressed as

$$y_t = H_t\theta + w_t, \tag{7}$$

where w_t is zero mean white Gaussian noise with covariance matrix Σ. From all the observations received at t, the least squares estimator is [15]

$$\hat{\theta}_t = \left(H_t^\top \Sigma^{-1} H_t\right)^{-1} H_t^\top \Sigma^{-1} y_t. \tag{8}$$

We assume that $w_{i,t}$ satisfies the Markov property with respect to the graph G, i.e., the noises of any pair of nonadjacent nodes are conditionally independent given the remaining noise values,

$$p(w_{i,t}, w_{j,t}|w_{V\setminus i,j}) = p(w_{i,t}|w_{V\setminus i,j})p(w_{j,t}|w_{V\setminus i,j}) \\ \text{for all } \{i,j\} \notin E, \text{and for all } t \in \mathbb{N}. \tag{9}$$

Let $K = \Sigma^{-1}$, and since it is the inverse of the covariance matrix, we call it the precision matrix. Since $w_{i,t}$ is Gaussian, we have [16]

$$K_{i,j} = O \quad \text{for all } \{i,j\} \notin E, \tag{10}$$

where $K_{i,j}$ is the (i,j)th block of K.

Given the observations from the beginning to time instant t, the least squares estimate $\tilde{\theta}_t$ can be expressed as

$$\tilde{\theta}_t = \left(\sum_{s=1}^{t} H_s^\top K H_s\right)^{-1} \sum_{s=1}^{t} H_s^\top K y_s, \tag{11}$$

where $\sum_{s=1}^{t} H_s^\top K y_s$ represents the sufficient statistics of the model. Our objective is to calculate $\tilde{\theta}_t$ in a distributed way.

3. THE PROPOSED ESTIMATION ALGORITHM

In this section, we describe the proposed algorithm. We assume node i has access to $H_{j,t}$ for $j \in \mathcal{N}_i$ through communication at time instant t. Let $Q \in \mathbb{R}^{N\times N}$ be a doubly stochastic matrix, which satisfies

$$Q1_N = 1_N, \quad 1_N^\top Q = 1_N^\top. \tag{12}$$

Denote by $Q_{i,j}$ and $Q_{i,j}^t$ the (i,j)th entry of Q and Q^t, respectively. We note that $Q_{i,j} = 0$ if nodes i and j are not connected. Such Q can be constructed by letting $Q = I_N - \epsilon\Xi$, where Ξ is the Laplacian matrix of the graph G; ϵ is a coefficient satisfying $\epsilon < 1/\max_i(\deg(i))$, with $\deg(i)$ denoting the degree of node i. Note that

$$\lim_{t\to\infty} Q_{i,j}^t = \frac{1}{N} \quad \text{for } i,j \in \{1,\cdots,N\}. \tag{13}$$

This is the principle we use behind the averaging of the sufficient statistics.

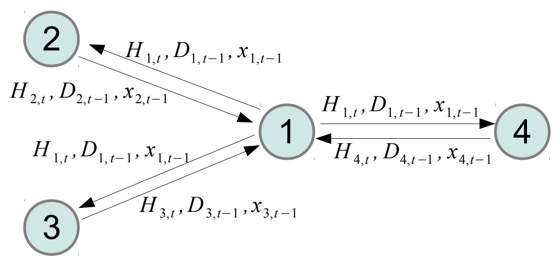

Fig. 1. Information exchange at time instant t.

In the proposed algorithm, each node keeps two matrix variables, D_i and x_i, which approximate $\sum_{s=1}^{t} H_s^\top K H_s$ and $\sum_{s=1}^{t} H_s^\top K y_s$, respectively. The method is based on the following formulas:

$$D_{i,t} = \sum_{j\in\mathcal{N}_i} Q_{i,j} D_{j,t-1} + \sum_{j\in\mathcal{N}_i} H_{j,t}^\top K_{j,i} H_{i,t} \tag{14}$$

$$= \sum_{s=1}^{t}\sum_{j=1}^{N} Q_{i,j}^{t-s} \sum_{k\in\mathcal{N}_j} H_{k,s}^\top K_{k,j} H_{j,s}, \tag{15}$$

$$x_{i,t} = \sum_{j\in\mathcal{N}_i} Q_{i,j} x_{j,t-1} + \sum_{j\in\mathcal{N}_i} H_{j,t}^\top K_{j,i} y_{i,t} \tag{16}$$

$$= \sum_{s=1}^{t}\sum_{j=1}^{N} Q_{i,j}^{t-s} \sum_{k\in\mathcal{N}_j} H_{k,s}^\top K_{k,j} y_{j,s},, \tag{17}$$

$$\tilde{\theta}_{i,t} = D_{i,t}^{-1} x_{i,t}, \tag{18}$$

and where Q^0 is defined to be the identity matrix. The information a node transmits to its neighbors includes $H_{i,t}, D_{i,t}$ and $x_{i,t}$ (see Fig. 1).

Let

$$Q_{(i)}^t = \begin{bmatrix} Q_{i,1}^t I_M & O & \cdots & O \\ O & Q_{i,2}^t I_M & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & Q_{i,N}^t I_M \end{bmatrix}. \tag{19}$$

Then (18) can be written as

$$\tilde{\theta}_{i,t} = \left(\sum_{q=1}^{t} H_q^\top K Q_{(i)}^{t-q} H_q\right)^{-1} \sum_{s=1}^{t} H_s^\top K Q_{(i)}^{t-s} y_s. \tag{20}$$

4. ANALYSIS

In this section, we first prove that the proposed estimator (20) is unbiased. Secondly, we prove that (20) asymptotically approaches the centralized estimator (11).

Let $\hat{\theta}_{i,t}$ be the least squares estimator based on local observations at the current time instant,

$$\hat{\theta}_{i,t} = \left(\sum_{j\in\mathcal{N}_i} H_{j,t}^\top K_{j,i} H_{i,t} \right)^{-1} \sum_{j\in\mathcal{N}_i} H_{j,t}^\top K_{j,i} y_{i,t}. \quad (21)$$

After substituting (2) into (21), we obtain

$$\mathrm{E}[\hat{\theta}_{i,t}] = \theta. \quad (22)$$

Also, by the property in (12), we can see that (20) is a linear combination of unbiased estimates with the sum of weighting coefficients being the identity matrix. Therefore, (20) is unbiased.

Next we compare the proposed estimator (20) and the centralized estimator (11). Before we proceed, we make the following assumptions about the model:

1. The sequence $\{w_t\}_{t\in\mathbb{N}}$ is bounded,
2. The matrix $\{H_t\}_{t\in\mathbb{N}}$ is bounded, and
3. The matrix $\{H_t\}_{t\in\mathbb{N}}$ is full rank and does not converge to a rank deficit matrix.

Let Y_t denote all the observations to time instant t,

$$Y_t = [y_1, \cdots, y_t]. \quad (23)$$

Consider the centralized and distributed estimators as two functions, $f_c(Y_t)$ and $f_d(Y_t)$ with observations as variables. Note that the randomness is temporarily ignored here. We prove that the proposed algorithm is asymptotically equivalent to the centralized estimator.

Theorem 1. *Assume y_t is bounded for all t, and the matrix sequence H_q does not converge to a rank deficient matrix. Then*

$$\lim_{t\to\infty} \left(f_d(Y_t) - f_c(Y_t)\right) = 0. \quad (24)$$

We will use the following lemma in the proof of Theorem 1 [17]:

Lemma 1. *Let A, B be Hermitian matrices of size $n \times n$ and let the eigenvalues $\lambda_i(A)$, $\lambda_i(B)$ and $\lambda_i(A+B)$ be arranged in increasing order. For each $i = 1, \cdots, n$ we have*

$$\lambda_i(A) + \lambda_{\min}(B) \le \lambda_i(A+B). \quad (25)$$

Proof of Theorem 1: Let

$$A_t = \frac{1}{N} \sum_{q=1}^{t} H_q^\top K H_q, \quad (26)$$

$$B_t = \sum_{p=1}^{t} H_p^\top K \left(Q_{(i)}^{t-p} - \frac{1}{N} I_{MN} \right) H_p, \quad (27)$$

$$C_t = \frac{1}{N} \sum_{s=1}^{t} H_s^\top K y_s, \quad (28)$$

$$D_t = \sum_{r=1}^{t} H_r^\top K \left(Q_{(i)}^{t-r} - \frac{1}{N} I_{MN} \right) y_r, \quad (29)$$

then (11) becomes

$$\tilde{\theta}_t = f_c(Y_t) = A_t^{-1} C_t, \quad (30)$$

whereas (20) turns into

$$\tilde{\theta}_{i,t} = f_d(Y_t) = (A_t + B_t)^{-1}(C_t + D_t). \quad (31)$$

By using the Woodbury matrix identity

$$(A_t + B_t)^{-1} = A_t^{-1} - A_t^{-1}(B_t^{-1} + A_t^{-1})^{-1} A_t^{-1}, \quad (32)$$

we can write (31) as

$$\begin{aligned} \tilde{\theta}_{i,t} =& A_t^{-1} C_t + A_t^{-1} D_t - A_t^{-1}(B_t^{-1} + A_t^{-1})^{-1} A_t^{-1} C_t \\ & - A_t^{-1}(B_t^{-1} + A_t^{-1})^{-1} A_t^{-1} D_t \quad (33) \\ =& \left(I - A_t^{-1}(B_t^{-1} + A_t^{-1})^{-1}\right) A_t^{-1} C_t + A_t^{-1} D_t \\ & - A_t^{-1}(B_t^{-1} + A_t^{-1})^{-1} A_t^{-1} D_t. \quad (34) \end{aligned}$$

Therefore, we shall prove that $A_t^{-1}(B_t^{-1} + A_t^{-1})^{-1}$ and the last two terms in (34) go to zero as t goes to infinity. It would be sufficient to show that B_t and D_t are bounded and A_t^{-1} goes to zero as t goes to infinity. We first show that B_t and D_t are bounded. The underlying principle we are using is that any geometric series with the absolute value of the common ratio less than 1 converges.

Note that

$$B_t = \sum_{s=1}^{t} \sum_{j=1}^{N} \left(Q_{i,j}^{t-s} - \frac{1}{N} \right) \sum_{k\in\mathcal{N}_j} H_{k,s}^\top K_{k,j} H_{j,s}, \quad (35)$$

$$D_t = \sum_{s=1}^{t} \sum_{j=1}^{N} \left(Q_{i,j}^{t-s} - \frac{1}{N} \right) \sum_{k\in\mathcal{N}_j} H_{k,s}^\top K_{k,j} y_{j,s}. \quad (36)$$

Let

$$b_t = \max_{s\in\{1,\cdots,t\}, j\in\{1,\cdots,N\}} \left\| \sum_{k\in\mathcal{N}_j} H_{k,s}^\top K_{k,j} H_{j,s} \right\|, \quad (37)$$

$$d_t = \max_{s\in\{1,\cdots,t\}, j\in\{1,\cdots,N\}} \left\| \sum_{k\in\mathcal{N}_j} H_{k,s}^\top K_{k,j} y_{j,s} \right\|. \quad (38)$$

Since Q is symmetric, it can be decomposed as

$$Q = U\Lambda U^{\top}, \tag{39}$$

where U is an orthonormal matrix and Λ is a diagonal matrix with eigenvalues being the diagonal entries. Then

$$Q^t = U\Lambda^t U^{\top}. \tag{40}$$

Denote by u_{ij} the (i,j)th entry of U. Then

$$Q_{i,j}^{t-s} = \sum_{l=1}^{N} u_{jl}u_{il}\lambda_l^{t-s}. \tag{41}$$

Since Q is a doubly stochastic matrix, we know that one of the eigenvalues is 1 with the corresponding eigenvector being $(1/\sqrt{N})1_N$, and the rest of the eigenvalues are strictly less than 1 in magnitude. Let $\lambda_N = 1$; we have $u_{iN}u_{jN}\lambda_N^{t-s} = \frac{1}{N}$ and

$$Q_{i,j}^{t-s} - \frac{1}{N} = \sum_{l=1}^{N-1} u_{jl}u_{il}\lambda_l^{t-s}. \tag{42}$$

Then

$$\|B_t\| = \sum_{s=1}^{t}\sum_{j=1}^{N}\left(Q_{i,j}^{t-s} - \frac{1}{N}\right)\left\|\sum_{k\in\mathcal{N}_j} H_{k,s}^{\top}K_{k,j}H_{j,s}\right\| \tag{43}$$

$$\leq \sum_{s=1}^{t}\sum_{j=1}^{N}\left(Q_{i,j}^{t-s} - \frac{1}{N}\right)b_t \tag{44}$$

$$= b_t\sum_{j=1}^{N}\sum_{l=1}^{N-1}u_{jl}u_{il}\sum_{s=1}^{t}\lambda_l^{t-s} \tag{45}$$

$$= b_t\sum_{j=1}^{N}\sum_{l=1}^{N-1}u_{jl}u_{il}\frac{1-\lambda_l^t}{1-\lambda_l}. \tag{46}$$

Similarly,

$$\|D_t\| \leq d_t\sum_{j=1}^{N}\sum_{l=1}^{N-1}u_{jl}u_{il}\frac{1-\lambda_l^t}{1-\lambda_l}. \tag{47}$$

$\|B_t\|$ and $\|D_t\|$ will be bounded for as long as b_t and d_t are bounded, which is true according to our assumption.

Next, we prove that A_t^{-1} goes to zero in the long run. Since every $H_q^{\top}KH_q$ is a positive definite matrix, it seems obvious that A_t^{-1} goes to zero. But the proof is not trivial. Since A_t is symmetric, the eigenvalue decomposition can be expressed as $A_t = U_t\Lambda_tU_t^{\top}$, where Λ_t is a diagonal matrix with diagonal entries being the eigenvalues; U is an orthonormal matrix. The inverse becomes $A_t^{-1} = U_t\Lambda_t^{-1}U_t^{\top}$. We only need to prove that each eigenvalue of A_t goes to infinity. Since

$$A_t = A_{t-1} + H_t^{\top}KH_t, \tag{48}$$

and by Lemma 1, we have

$$\lambda_j(A_{t-1}) + \lambda_{\min}(H_t^{\top}KH_t) \leq \lambda_j(A_t), \quad j = 1, \cdots, M. \tag{49}$$

Because $H_t^{\top}KH_t$ is strictly positive definite, $\lambda_{\min}(H_t^{\top}KH_t)$ is greater than 0. Since it is assumed that H_t does not converge to a rank deficient matrix, $\{\lambda_{\min}(H_t^{\top}KH_t)\}_{t\in\mathbb{N}}$ will not converge to 0. Then A_t^{-1} converges to zero as t goes to infinity. This completes the proof. □

Note that w_t is a Gaussian process, and y_t is bounded with probability 1. Therefore, the next theorem follows immediately.

Theorem 2. *Under the assumption of Theorem 1, the proposed estimator* (20) *converges to that of the centralized estimator* (11) *with probability 1 in the long run.*

5. SIMULATION

In this section, we test the algorithms in a network with 20 nodes. The topology is shown in Figure 2. We let $L = 2$, $M = 3$ and $\theta = [1, -1]^{\top}$. All the entries in H are i.i.d. standard Gaussian variables, and $\epsilon = 0.1$. Also let the precision matrix K equal to $(2I_N - \epsilon\Xi) \otimes S$ where

$$S = \begin{bmatrix} 2 & 0.5 & -1 \\ 0.5 & 2 & 1 \\ -1 & 1 & 2 \end{bmatrix}. \tag{50}$$

This is an easy way to specify a positive definite matrix that satisfies the Markov property. Figure 3 shows the mean

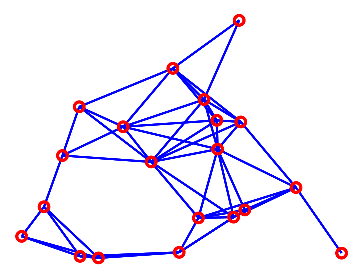

Fig. 2. Topology of the network.

square error performance of one realization for the setting. We can see that the performance of the proposed method approaches the centralized estimator after 300 iterations. Figure 4 shows the averaged results of 500 runs. The performance of the proposed estimator converges quickly to that of the centralized estimator.

6. CONCLUSION

In this paper, we proposed a distributed sequential algorithm for the case where the noises are correlated. We assume that the noises have the conditional independence property, We show that the proposed algorithm is asymptotically equivalent

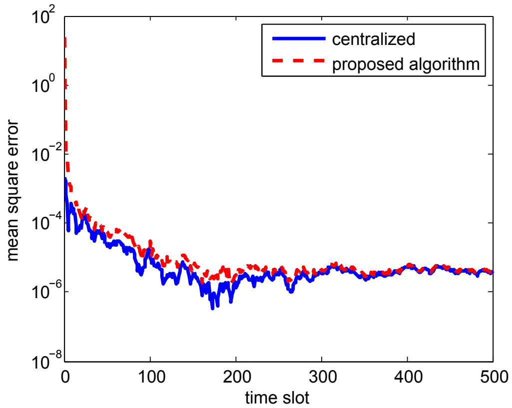

Fig. 3. Performance of a sample run.

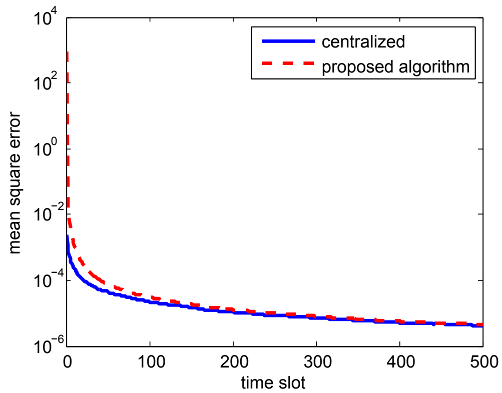

Fig. 4. Averaged performance of 500 runs.

to the centralized algorithm, regardless of the actual values of observation, for as long as they are bounded. Since the centralized estimator (11) is an efficient estimator [15], and thus, the proposed estimator asymptotically approaches the Cramér-Rao bound. The simulations confirm the statement.

Future work includes the explicit quantification of the convergence rate of the algorithm. For example, how fast the distributed estimator converges to the centralized estimator; how it depends on the size or the connectivity of the network. Besides, here we assume that $K_{i,j}$ is known to the nodes i and j. However, a more realistic assumption would be that the nodes only know $C_{i,j}$, and $K_{i,j}$ should be calculated through a distributed matrix inversion. The case for general K is also worth of investigation. In addition to the spatial correlation of noises, the temporal correlation can be incorporated into the model to make the problem more general.

7. REFERENCES

[1] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation*, Prentice Hall Inc., 1989.

[2] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3122–3136, 2008.

[3] N. Takahashi, I. Yamada, and A. H. Sayed, "Diffusion least-mean squares with adaptive combiners: Formulation and performance analysis," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4795–4810, 2010.

[4] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed, "Diffusion recursive least-squares for distributed estimation over adaptive networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 1865–1877, 2008.

[5] S. Kar and J. M. F. Moura, "Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 674–690, 2011.

[6] S. Kar, J. M. F. Moura, and H. V. Poor, "Distributed linear parameter estimation: Asymptotically efficient adaptive strategies," *arXiv preprint arXiv:1109.4960*, 2011.

[7] S.-Y. Tu and A. H. Sayed, "Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6217 –6234, dec. 2012.

[8] A. H. Sayed, "Diffusion adaptation over networks," *arXiv preprint arXiv:1205.4220*, 2012.

[9] I. D. Schizas, G. B. Giannakis, S. I. Roumeliotis, and A. Ribeiro, "Consensus in ad hoc wsns with noisy links - Part II: Distributed estimation and smoothing of random signals," *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1650–1666, 2008.

[10] I. D. Schizas, A. Ribeiro, and G. B. Giannakis, "Consensus in ad hoc WSNs with noisy links - Part I: Distributed estimation of deterministic signals," *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 350–364, 2008.

[11] S. Kar, J. M. F. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3575–3605, 2012.

[12] R. Olfati-Saber, "Distributed Kalman filter with embedded consensus filters," in *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC'05. 44th IEEE Conference on*. IEEE, 2005, pp. 8179–8184.

[13] F. S. Cattivelli and A. H. Sayed, "Diffusion strategies for distributed Kalman filtering and smoothing," *IEEE Transactions on Automatic Control*, vol. 55, no. 9, pp. 2069–2084, 2010.

[14] O. Hlinka, F. Hlawatsch, and P. M Djurić, "Distributed particle filtering in agent networks: A survey, classification, and comparison," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 61–81, 2013.

[15] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory (v. 1)*, Prentice Hall, 1993.

[16] A. Wiesel and A. O. Hero, "Distributed covariance estimation in Gaussian graphical models," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 211–220, 2012.

[17] R. A. Horn and C. R. Johnson, *Matrix analysis*, Cambridge University Press, 1990.