Diffusion Recursive Least Square Adaptive Networks with Neighbor-Selection

Wael M. Bazzi

Electrical and Computer Engineering Department
American University in Dubai
Dubai, United Arab Emirates
Email: wbazzi@aud.edu

Vahid Vahidpour, Amir Rastegarnia, and Azam Khalili

Department of Electrical Engineering
Malayer University
Malayer 65719-95863, Iran
Email: {vahidpour,rastegarnia,khalili}@malayeru.ac.ir

***Abstract*—Constrained communication resources and limited communication bandwidth are key issues for any task involving wireless sensor networks. This phenomenon motivated the authors to examine diffusion networks where only a fraction of neighbors participle in the communication process. In this context, we modify the Diffusion Recursive Least Square (DRLS) algorithm by allowing each node to receive intermediate estimates from a subset of its neighbors, called neighbor-selection DRLS. This results in significant reduction in communication overhead at the cost of some possible deterioration in the network performance. We derive a theoretical expression for the steady state Mean Square Deviation (MSD). Both numerical simulations and theoretical findings are used to validate the effectiveness of the proposed algorithm in providing a trade off between communication burden and estimation performance.**

Keywords–Adaptive network; diffusion; neighbor selection; recursive least-squares.

I. INTRODUCTION

Diffusion strategies are well-known techniques that enable real-time learning and collaboration in adaptive networks [1]–[3]. In these methods, information is gathered and processed at all agents in a simultaneous fashion. This results in a live sharing mechanism that ripples frequently through the whole network [4]. Consequently, significant improvements are accrued in estimation performance of each network agent, in comparison to the case in which nodes operate autonomously. Notable properties of such networks are scalability and robustness to node/link failures. Power and bandwidth resources, however, are the major constraints on performing a cooperative task in an adaptive network. Communication is constrained by the limited data transmission through radio links. Therefore, the attained advantages of diffusion strategies in terms of internode communications comes at an additional communication cost [5].

Following on the discussion in the previous paragraph, it is desirable to lower the level of internode communications as much as possible, while maintaining the benefits of cooperation. There are some existing efforts related to reducing the communication overheard, such as decreasing the dimension of the estimates [6]–[8], selecting a subset of the entries of the intermediate estimate vectors [9]–[12], and set membership filtering [13] [14]. In most earlier publications, it is assumed that the degree of each node is fixed and predefined by the network topology and, moreover, that every node senses data that is affected by information diffused by all of its neighbors. To the best of our knowledge, choosing a subset of neighboring nodes was considered in [5], [15]–[19], but only in diffusion least-mean-squares (LMS) networks. In this manuscript, we consider the case where only a subset of agents participate in the communication process. We focus on the scenario in [5], in which every node consults with only a subset of its neighbors and propose a novel reduced communication recursive least square algorithm, called neighbor selection DRLS. In this algorithm, which aims at further releasing the communication density of DRLS, each node updates its estimate and sends the intermediate estimate to only a subset of its neighbors. Moreover, the total amount of internode communication in the network is efficiently decreased with less performance degradation in comparison to the diffusion LMS algorithm. We derive a theoretical expression for the steady state MSD of the Neighbor Selection DRLS algorithm and verify its accuracy through numerical simulations.

The remainder of this paper is organized as follows: In Section II, we recall a conventional DRLS algorithm and formulate the proposed Neighbor Selection DRLS algorithm. The performance analysis is examined in Section III. We provide simulation results in Section IV and draw conclusions in Section V.

Notation: We use plain lowercase letters to denote scalars, lowercase bold letters to denote vectors and boldface uppercase letters for matrices.

II. ALGORITHM DESCRIPTION

A. Conventional Diffusion RLS

We consider a connected network of N nodes which aims to determine an unknown vector, $w^o \in \mathbb{R}^{M\times 1}$, in a distributed manner. At every time instant i and each node k, scalar measurements $d_{k,i} \in \mathbb{R}$ are related to regression vectors, $\mathbf{u}_{k,i} \in \mathbb{R}^{1\times M}$, via the following linear regression model [20]:

$$d_{k,i} = \boldsymbol{u}_{k,i}\mathbf{w}^o + v_{k,i} \tag{1}$$

where $v_{k,i}$ denotes the additive noise process. The vector $\mathbf{w}^o$ denotes the parameter of interest that the agents wish to identify.

We are then motivated to consider the following weighted least square problem:

$$\min_{w} \|\mathbf{y}_{k,i} - \mathbf{H}_{k,i}\boldsymbol{\psi}\|^2_{\Lambda_i} \tag{2}$$

where $\mathbf{y}_{k,i}$ and $\mathbf{H}_{k,i}$ are formed by stacking the history of measurement and noise samples of node k up to time i as follows:

$$\mathbf{y}_{k,i} = \operatorname{col}\{d_{k,i}, \ldots, d_{k,1}, d_{k,0}\} \tag{3}$$

 ISBN: 978-1-61208-642-2

$$\mathbf{H}_{k,i} = \mathrm{col}\left\{\mathbf{u}_{k,i}, \ldots, \mathbf{u}_{k,1}, \mathbf{u}_{k,0}\right\} \tag{4}$$

where $\mathrm{col}\{\cdots\}$ denotes a column vector formed by stacking its arguments on top of each other. The solution $\boldsymbol{\psi}_{k,i}$ from (2) is given by [2]:

$$\boldsymbol{\psi}_{k,i} = \left(\mathbf{H}_{k,i}^T \boldsymbol{\Lambda}_i \mathbf{H}_{k,i}\right)^{-1} \left(\mathbf{H}_{k,i}^T \boldsymbol{\Lambda}_i \mathbf{y}_{k,i}\right) \tag{5}$$

where $\boldsymbol{\Lambda}_i \geq 0$ denotes a Hermitian weighting matrix. A common choice for $\boldsymbol{\Lambda}_i$ is

$$\boldsymbol{\Lambda}_i = \mathrm{diag}\left\{1, \lambda, \ldots, \lambda^i\right\} \tag{6}$$

$0 \ll \lambda \leq 1$ denotes an exponential forgetting factor whose value is usually close to unity. In this way, the closer the occurrence time of data is to present, the less heavily scaled it will be. Employing the recursive properties of

$$\mathbf{H}_{k,i}^T \boldsymbol{\Lambda}_i \mathbf{H}_{k,i} = \lambda \mathbf{H}_{k,i-1}^T \boldsymbol{\Lambda}_{i-1} \mathbf{H}_{k,i-1} + \mathbf{u}_{k,i}^T \mathbf{u}_{k,i} \tag{7}$$

$$\mathbf{H}_{k,i}^T \boldsymbol{\Lambda}_i \mathbf{y}_{k,i} = \lambda \mathbf{H}_{k,i-1}^T \boldsymbol{\Lambda}_{i-1} \mathbf{y}_{k,i-1} + \mathbf{u}_{k,i}^T d_{k,i} \tag{8}$$

alongside defining $\mathbf{P}_{k,i} = \left(\mathbf{H}_{k,i}^T \boldsymbol{\Lambda}_i \mathbf{H}_{k,i}\right)^{-1}$, and as well making use of the so-called matrix inversion formula [21], called Sherman-Morrison Formula, to (7), the following recursive equations to assess $\boldsymbol{\psi}_{k,i}$ are given:

$$\mathbf{P}_{k,i} = \lambda^{-1} \left(\mathbf{P}_{k,i-1} - \frac{\lambda^{-1} \mathbf{P}_{k,i-1} \mathbf{u}_{k,i}^T \mathbf{u}_{k,i} \mathbf{P}_{k,i-1}}{1 + \lambda^{-1} \mathbf{u}_{k,i} \mathbf{P}_{k,i-1} \mathbf{u}_{k,i}^T}\right) \tag{9}$$

$$\boldsymbol{\psi}_{k,i} = \boldsymbol{\psi}_{k,i-1} + \mathbf{P}_{k,i} \mathbf{u}_{k,i}^T \left(d_{k,i} - \mathbf{u}_{k,i} \boldsymbol{\psi}_{k,i-1}\right) \tag{10}$$

Due to the fact that the intermediate value $\boldsymbol{w}_{k,i}$ at node k is generally a better estimate for $\mathbf{w}^o$ than $\boldsymbol{\psi}_{k,i}$, we replace $\boldsymbol{\psi}_{k,i-1}$ by $\mathbf{w}_{k,i-1}$ in (10)

$$\boldsymbol{\psi}_{k,i} = \mathbf{w}_{k,i-1} + \mathbf{P}_{k,i} \mathbf{u}_{k,i}^T \left(d_{k,i} - \mathbf{u}_{k,i} \mathbf{w}_{k,i-1}\right) \tag{11}$$

It is common that the local estimates are scattered outside of each node's own neighborhood. Then, the diffusion RLS strategy comprises two stages: adaptation and aggregation.

1) *Adaptation:* Each node estimator is updated utilizing observed data $\{d_{k,i}, \mathbf{u}_{k,i}\}$ in (9) and (11). The resulting pre-estimates are called $\boldsymbol{\psi}_{k,i}$ as in (12).
2) *Aggregation:* Each node diffuses its local pre-estimate with its neighbors, collects the estimators from its neighbors and performs a weighted average as

$$\mathbf{w}_{k,i} = \sum_{l \in \mathcal{N}_k'} c_{lk} \boldsymbol{\psi}_{l,i} \tag{12}$$

to obtain the estimate $\mathbf{w}_{k,i}$ (via so-called spatial update). Where $\mathcal{N}_k'$ denotes the close neighborhood of node k, i.e., it consists of a set of all nodes communicating to node k, including k itself. The coefficients c_{lk} are designed to satisfy the following condition:

$$C^T \mathbb{1}_N = \mathbb{1}_N \tag{13}$$

where the notation $\mathbb{1}$ denotes an $N \times 1$ column vector with all one entries. To minimize the communication density, here, we cover the diffusion RLS strategy, which does not involve any information exchange.

B. Neighbor-Seletion Diffusion RLS

The aggregation step (12) improves the estimation performance. However, this is compromised by the communication density. In order to reduce the amount of communication, we consider the case in which each node is allowed to diffuse the update estimate with only a subset of its neighborhood $\mathcal{N}_k$ [5]. Generally speaking, because a subset of the information is available at node k to perform (12), the aggregation phase would be updated so that it could be performed with current available information. Doing so, we can decrease the internode communications that is being accomplished among nodes and establish a trade-off between estimation performance and communication cost.

Let $\delta_k = |\mathcal{N}_k|$ be the degree or valency of node k, where $|\cdot|$ is the cardinality operator. To achieve this, assume node k communicates at each time instant i to receive the intermediate estimate, $\boldsymbol{\psi}_{l,i}$, from $0 < n_k \leq \delta_k$. To-be-selected neighboring nodes of node k at iteration i are characterized by a neighborhood-selection variable as $a_{kl,i}$. This variable determines the status of the link, being active or inactive, between node k and l at time instant i.

The neighbor-selection variable is defined as follows:

$$a_{lk,i} = \begin{cases} 1 & \text{if } l \in \mathcal{N}_{k,i}^D \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

where $\mathcal{N}_{k,i}^D$is the neighborhood of node k at time instant i and consists of all the nodes, which transmit their intermediate estimates to node k. Adjusting $a_{kl,i} = 1$ means that node k communicates with its neighboring node l at iteration i and receives its intermediate estimate to employ at the aggregation step. Having, $a_{lk,i} = 0$ means that node k does not receive the intermediate estimate of its neighbor l at iteration i.

With regards to the proposed neighbor selection scheme, the following remark is made [5]:

Remark 1. *The neighbor-selection variable $\{a_{lk,i}\}$ is mutually independent of each element of set $\{\mathbf{u}_{k,i}, d_k(i), v_k(i)\}$. Moreover, the neighbor-selection probability, denoted by ρ_k is shift-invariant and identical for all the neighbors. This probability is expressed as:*

$$\rho_k = \mathbb{E}\left[a_{lk,i}\right] = \frac{n_k}{d_k}$$

When the intermediate estimates of only n_k neighbors are received at node k, we instead propose a new aggregation method that uses the node's own intermediate estimate as a proxy [19] for missing data and changes (3):

$$\mathbf{w}_{k,i} = c_{kk} \boldsymbol{\psi}_{k,i} + \sum_{l \in \mathcal{N}_k} c_{lk} [a_{lk,i} \boldsymbol{\psi}_{l,i} + (1 - a_{lk,i}) \boldsymbol{\psi}_{k,i}] \tag{15}$$

Accordingly, the considered neighbor-selection diffusion RLS algorithm utilizes (9) and (11) in the adaptation phase and (15) for the aggregation phase. It is noteworthy to say that (3) and (15) have the same computational complexity. Consequently, the considered algorithm, i.e., (9), (11), and (15), requires the same number of arithmetic operations as the diffusion RLS algorithm.

 ISBN: 978-1-61208-642-2

III. Performance Analysis

To proceed with the analysis, we make the following assumptions:

Assumption:

1) The regression data $\mathbf{u}_{k,i}$ are temporally white and spatially independent random variables with zero mean and covariance matrix $\mathbf{R}_{u,k} \triangleq \mathbb{E}\left[\mathbf{u}_{k,i}^T \mathbf{u}_{k,i}\right] \geq 0$.
2) The noise signal $v_{k,i}$ is temporally white and spatially independent random variable with zero mean and variance $\sigma_{v,k}^2$.
3) The regression data $\{\mathbf{u}_{m,i_1}\}$, and the model noise signals v_{n,i_2}, are mutually independent random variables for all indexes $\{i_1, i_2, m, n\}$.
4) For sufficiently large i, at any node k, we can replace $\mathbf{P}_{k,i}$ and $\mathbf{P}_{k,i}^{-1}$ with their expected values, $\mathbb{E}\left[\mathbf{P}_{k,i}\right]$ and $\mathbb{E}\left[\mathbf{P}_{k,i}^{-1}\right]$, respectively.
5) For a sufficiently large i, at any node k, we have $\mathbb{E}\left[\mathbf{P}_{k,i}\right] = \mathbb{E}\left[\mathbf{P}_{k,i}^{-1}\right]^{-1}$

A. Network update equation

Define $M \times 1$ error vector as follows:

$$\tilde{\boldsymbol{\psi}}_{k,i} \triangleq \mathbf{w}^o - \boldsymbol{\psi}_{k,i} \tag{16}$$

$$\tilde{\mathbf{w}}_{k,i} \triangleq \mathbf{w}^o - \mathbf{w}_{k,i} \tag{17}$$

$$\tilde{\mathbf{w}}_i \triangleq \operatorname{col}\{\tilde{\mathbf{w}}_{1,i}, \ldots, \tilde{\mathbf{w}}_{N,i}\} \tag{18}$$

Using the data model in (1) and subtracting $\mathbf{w}^o$ from both sides of the relation in (11), we get

$$\tilde{\boldsymbol{\psi}}_{k,i} = \tilde{\mathbf{w}}_{k,i-1} - \mathbf{P}_{k,i}\mathbf{u}_{k,i}^T\left[\mathbf{u}_{k,i}\tilde{\mathbf{w}}_{k,i-1} + v_{k,i}\right] \tag{19}$$

Using the same procedure as stated in [12], the equation above can be written in the following form:

$$\tilde{\boldsymbol{\psi}}_{k,i} = \lambda\tilde{\mathbf{w}}_{k,i} - (1-\lambda)\,\mathbf{R}_{u,k}^{-1}\mathbf{u}_{k,i}^T v_{k,i} \tag{20}$$

Moreover, subtracting both sides of (20) from w^o gives

$$\tilde{\mathbf{w}}_{k,i} = \left(1 - \sum_{l\in\mathcal{N}_k} a_{lk,i}c_{lk}\right)\tilde{\boldsymbol{\psi}}_{k,i} + \sum_{l\in\mathcal{N}_k} a_{lk,i}c_{lk}\tilde{\boldsymbol{\psi}}_{l,i} \tag{21}$$

which leads to

$$\tilde{\mathbf{w}}_i = \lambda\mathfrak{B}_i\tilde{\mathbf{w}}_{i-1} - \mathfrak{B}_i\boldsymbol{\Pi}\mathbf{s}_i \tag{22}$$

where

$$\boldsymbol{\Pi} \triangleq (1-\lambda)\operatorname{diag}\left\{\boldsymbol{R}_{u,1}^{-1}, \ldots, \boldsymbol{R}_{u,N}^{-1}\right\} \tag{23}$$

$$\boldsymbol{s}_i \triangleq \left\{\mathbf{u}_{1,i}^T v_{1,i}, \ldots, \mathbf{u}_{N,i}^T v_{N,i}\right\} \tag{24}$$

$$\mathfrak{B}_i = \boldsymbol{\mathcal{B}}_i \otimes \mathbf{I}_M \tag{25}$$

$$\boldsymbol{\mathcal{B}}_i = \begin{bmatrix} \mathbf{B}_{11,i} & \cdots & \mathbf{B}_{1N,i} \\ \vdots & \ddots & \vdots \\ \mathbf{B}_{N1,i} & \cdots & \mathbf{B}_{NN,i} \end{bmatrix} \tag{26}$$

where

$$\mathbf{B}_{p,q,i} = \begin{cases} 1 - \sum_{l\in\mathcal{N}_p} a_{pl,i}c_{pl} & \text{if } q = p \\ a_{pq,i}c_{pq} & \text{if } q \in \mathcal{N}_p \\ 0 & \text{otherwise} \end{cases} \tag{27}$$

and $\otimes$ denotes the Kronecker product.

For any arbitrary symmetric nonnegative-definite matrix $\boldsymbol{\Sigma}$, using the alternative notation $\|\mathbf{x}\|_{\boldsymbol{\sigma}}^2$, where $\boldsymbol{\sigma} = \operatorname{vec}\{\boldsymbol{\Sigma}\}$, to refer to the weighted square quantity $\mathbf{x}^T\Sigma\mathbf{x}$ and following similar arguments to those in [3], we arrive at the following variance relation:

$$\mathbb{E}\left[\|\tilde{\mathbf{w}}_i\|_{\boldsymbol{\sigma}}^2\right] = \mathbb{E}\left[\|\tilde{\mathbf{w}}_{i-1}\|_{\lambda^2\boldsymbol{\Phi}\boldsymbol{\sigma}}^2\right] + \operatorname{vec}^T\{\boldsymbol{\mathcal{G}}\}\,\boldsymbol{\Phi}\boldsymbol{\sigma} \tag{28}$$

where

$$\boldsymbol{\mathcal{G}} = \boldsymbol{\Pi}\mathbb{E}\left[\mathbf{s}_i\mathbf{s}_i^T\right]\boldsymbol{\Pi} \tag{29}$$

which, in view of the *Assumptions*, can be expressed as

$$\boldsymbol{\mathcal{G}} = (1-\lambda)^2\left\{\sigma_{v,1}^2\mathbf{R}_{u,1}^{-1}, \ldots, \sigma_{v,N}^2\mathbf{R}_{u,N}^{-1}\right\} \tag{30}$$

and

$$\boldsymbol{\Phi} = \mathbb{E}\left[\mathfrak{B}_i^T \otimes \mathfrak{B}_i^T\right] \tag{31}$$

We arrive at the following expression for the network MSD (η)

$$\eta = \frac{1}{N}\operatorname{vec}^T\{\boldsymbol{\mathcal{G}}\}\,\boldsymbol{\Phi}\left(\mathbf{I}_{M^2N^2} - \lambda^2\boldsymbol{\Phi}\right)^{-1}\operatorname{vec}\{I_{MN}\} \tag{32}$$

We consider an adaptive network with $N = 20$ nodes. Each node has between one and seven neighbors excluding itself. We assume that each node, on average, is connected to four other agents. The regressors, $\mathbf{u}_{k,i}$, were chosen Gaussian i.i.d with randomly generated different diagonal covariance matrices, $\mathbf{R}_{u,k}$. The additive noise signals at nodes are zero mean Gaussian with variances $\sigma_{v,k}^2$ and independent of the regression data. The traces of the covariance matrix regressors and the noise variances at all nodes, $\operatorname{tr}\{\mathbf{R}_{u,k}\}$ and $\sigma_{v,k}^2$, are shown in Fig. 1. It is noteworthy that we obtain the network MSD learning curves of all figures by averaging over 50 experiments and the unknown parameter $\mathbf{w}^o$ of length $M = 8$ is randomly generated. In the proposed algorithm, we determine the number of neighbors with which each node communicates at each iteration to receive their intermediate estimates via $n_k = \min(K, \delta_k)$ where $0 \leq K \in \mathbb{N} < \delta_k$ specifies the maximum number of consulted neighbors of every node at each iteration. We use the relative-degree weights for $\{c_{lk}\}$ in the combination phase. In Fig. 2, we simulate the learning curves of instantaneous network MSD for different values of K. Fig. 3 also demonstrates the steady-state MSD of all the nodes for different values of K when $\lambda = 0.99$.

IV. Conclusion and Future Work

We introduced a neighbor-selection DRLS for distributed adaptive estimation. This algorithm provides reduced internode communication and bandwidth usage by allowing each node to receive intermediate estimates from a fraction of its neighbors. We derive an expression for the network MSD. The simulation results conform with the theoretical derivations. They illustrate that a trade off between communication cost and estimation performance can be obtained. In our future work we plan to derive the optimum combination weights that minimize the steady-state MSD at every node.

 ISBN: 978-1-61208-642-2

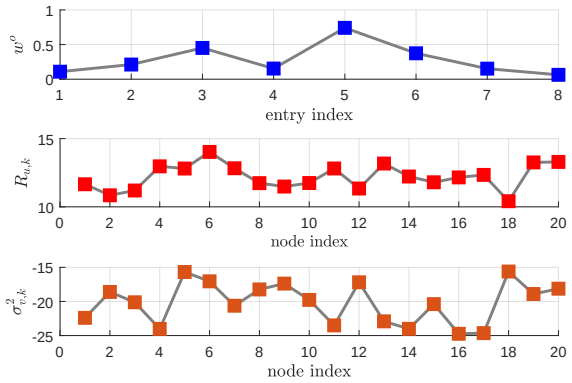

Figure 1. Entries of $\mathbf{w}^o$, $\mathrm{tr}\left\{\mathbf{R}_{u,k}\right\}$, and $\left\{\sigma^2_{v,k}\right\}$ used in simulation.

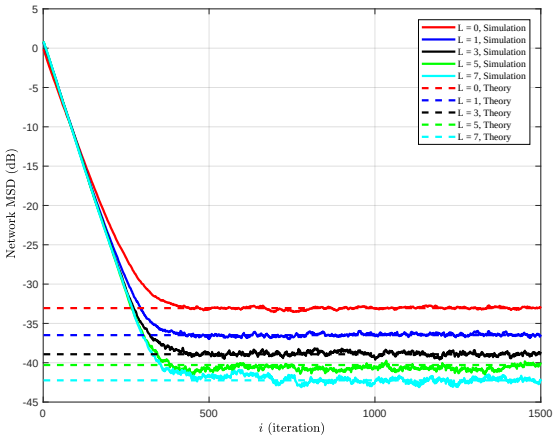

Figure 2. Experimental and thoretical network MSD curves of the neighbor-selection DRLS algorithm with different values K of when $\lambda = 0.95$.

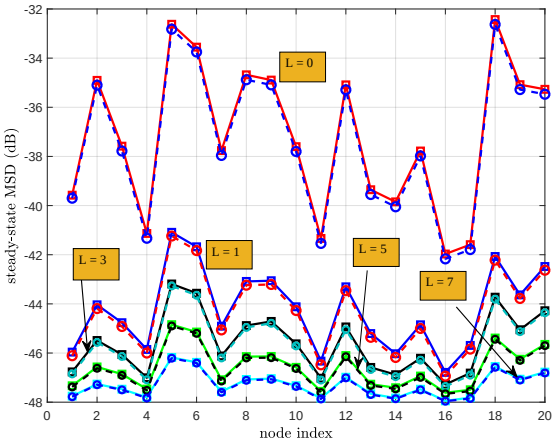

Figure 3. Theoretical and experimental steady-state MSDs of the neighbor-selection DRLS algorithm at each node for different values of K when $\lambda = 0.99$.

REFERENCES

[1] F. S. Cattivelli and A. H. Sayed, "Diffusion LMS strategies for distributed estimation," Signal Processing, IEEE Transactions on, vol. 58, no. 3, 2010, pp. 1035–1048.

[2] 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, May 2008, pp. 1865–1877.

[3] A. H. Sayed, "Diffusion adaptation over networks," Academic Press Library in Signal Processing, vol. 3, 2013, pp. 323–454.

[4] J. Chen and A. H. Sayed, "Diffusion adaptation strategies for distributed optimization and learning over networks," Signal Processing, IEEE Transactions on, vol. 60, no. 8, 2012, pp. 4289–4305.

[5] R. Arablouei, S. Werner, K. Dogancay, and Y.-F. Huang, "Analysis of a reduced-communication diffusion LMS algorithm," Signal Processing, vol. 117, 2015, pp. 355–361.

[6] M. O. Sayin and S. S. Kozat, "Single bit and reduced dimension diffusion strategies over distributed networks," Signal Processing Letters, IEEE, vol. 20, no. 10, 2013, pp. 976–979.

[7] ——, "Compressive diffusion strategies over distributed networks for reduced communication load," Signal Processing, IEEE Transactions on, vol. 62, no. 20, 2014, pp. 5308–5323.

[8] S. Chouvardas, K. Slavakis, and S. Theodoridis, "Trading off complexity with communication costs in distributed adaptive learning via Krylov subspaces for dimensionality reduction," Selected Topics in Signal Processing, IEEE Journal of, vol. 7, no. 2, 2013, pp. 257–273.

[9] R. Arablouei, S. Werner, Y.-F. Huang, and K. Dogancay, "Distributed least mean-square estimation with partial diffusion," Signal Processing, IEEE Transactions on, vol. 62, no. 2, 2014, pp. 472–484.

[10] R. Arablouei, K. Dogancay, S. Werner, and Y.-F. Huang, "Adaptive distributed estimation based on recursive least-squares and partial diffusion," Signal Processing, IEEE Transactions on, vol. 62, no. 14, 2014, pp. 3510–3522.

[11] V. Vahidpour, A. Rastegarnia, A. Khalili, W. M. Bazzi, and S. Sanei, "Analysis of Partial Diffusion LMS for Adaptive Estimation Over Networks with Noisy Links," IEEE Transactions on Network Science and Engineering, 2017.

[12] V. Vahidpour, A. Rastegarnia, A. Khalili, and S. Sanei, "Analysis of partial diffusion recursive least squares adaptation over noisy links," IET Signal Processing, 2017.

[13] S. Gollamudi, S. Nagaraj, S. Kapoor, and Y.-F. Huang, "Set-membership filtering and a set-membership normalized LMS algorithm with an adaptive step size," Signal Processing Letters, IEEE, vol. 5, no. 5, 1998, pp. 111–114.

[14] J. R. Deller Jr, M. Nayeri, and S. F. Odeh, "Least-square identification with error bounds for real-time signal processing and control," Proceedings of the IEEE, vol. 81, no. 6, 1993, pp. 815–849.

[15] O. L. Rørtveit, J. H. Husøy, and A. H. Sayed, "Diffusion LMS with communication constraints," in Signals, Systems and Computers (ASILOMAR), 2010 Conference Record of the Forty Fourth Asilomar Conference on. IEEE, 2010, pp. 1645–1649.

[16] C. G. Lopes and A. H. Sayed, "Diffusion adaptive networks with changing topologies," in 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, 2008.

[17] N. Takahashi and I. Yamada, "Link probability control for probabilistic diffusion least-mean squares over resource-constrained networks," in Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on. IEEE, 2010, pp. 3518–3521.

[18] X. Zhao and A. H. Sayed, "Single-link diffusion strategies over adaptive networks," in Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on. IEEE, 2012, pp. 3749–3752.

[19] J.-W. Lee, S.-E. Kim, and W.-J. Song, "Data-selective diffusion LMS for reducing communication overhead," Signal Processing, vol. 113, 2015, pp. 211–217.

[20] A. H. Sayed, Fundamentals of Adaptive Filtering. Wiley-IEEE Press, Jun. 2003.

[21] C. D. Meyer, Matrix analysis and applied linear algebra. Siam, 2000, vol. 2.

 ISBN: 978-1-61208-642-2