Diffusion Affine Projection Algorithm for Multitask Networks

Vinay Chakravarthi Gogineni[1], Mrityunjoy Chakraborty[2]
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur, INDIA
E.Mail : [1] vinaychakravarthi@ece.iitkgp.ernet.in, [2] mrityun@ece.iitkgp.ernet.in

***Abstract*—Distributed adaptive networks achieve better estimation performance by exploiting temporal as well as spatial diversity. In this paper, we consider the problem of estimating multiple optimal parameter vectors (also termed as tasks) under correlated input, over a sensor network, where the nodes within the same cluster are engaged in estimating a common optimum parameter vector in distributed manner. For this, we present an efficient multitask diffusion affine projection algorithm (APA). The proposed scheme uses a regularized term to promote similarity among the parameter vectors estimated by neighboring clusters. Usage of APA makes the algorithm robust against correlated input. We present important results on the mean and mean square convergence of the proposed strategy. Simulations are carried out to demonstrate the effectiveness of the proposed algorithm. Compared to the non-cooperative APA, the proposed multitask diffusion APA exhibits remarkably improved performance in terms of both convergence rate and steady-state MSD.**

***Index Terms*—Multitask learning, distributed adaptive estimation, cooperative learning, adaptive diffusion networks, affine projection algorithm.**

I. Introduction

Distributed adaptive estimation has emerged as an attractive and challenging research area with the advent of multi-agent networks. Consider a connected network consisting of N agents (often called nodes) observing temporal data arising from different spatial sources with possibly different statistical profiles. The objective is to enable the nodes to estimate the parameter vector of interest in a collaborative manner from the observed data. In adaptive networks, the interconnected nodes continuously learn and adapt, as well as perform the assigned tasks such as parameter estimation from observations collected by the dispersed agents. As the individual nodes share the computational burden in distributed estimation schemes, the complexities are reduced over centralized strategies with comparable estimation accuracy to the centralized solution [1], [2]. The efficiency of the distributed adaptive estimation is subject to the mode of cooperation among the nodes [3]. In incremental mode of cooperation, each node transfers information to its corresponding adjacent node in sequential manner using a cyclic pattern of collaboration. Though this approach reduces communication overhead, it is difficult to establish a cyclic pattern as the number of sensor nodes increases [3]. On the other hand, in diffusion mode of cooperation [3], each node k exchanges information with its neighborhood $\mathcal{N}_k$ (i.e., the set of all neighbors including self) and obtain more information than in incremental mode of cooperation. Moreover, the diffusion mode of cooperation is robust against link failures.

Depending on the number of parameter vectors to be estimated, adaptive networks can be classified into single-task and multitask networks. In a single-task network, all nodes collaboratively estimate a single parameter vector (i.e., each node is assigned the same task). In the context of single-task estimation, several useful distributed strategies such as incremental strategies [4], [5] and diffusion strategies [6]–[10] have been proposed and analyzed in detail.

Beside single-task scenarios, in some applications, multiple parameter vectors need to be estimated simultaneously in collaborative fashion. For example, in distributed active noise control application, agents need to estimate different but related active noise control filters [11]. Similarly, in applications like node-specific cooperative spectrum sensing [12], node-specific speech enhancement and DOA estimation [13], and study of tremor in Parkinson's disease [14], node-specific or multiple optimum parameter vectors need to be estimated simultaneously in collaborative fashion. In a multitask network, the nodes are grouped into clusters and the nodes within the same cluster are engaged in estimating a common parameter vector [15]–[17]. Different clusters generally have different (though related) tasks. The estimation still needs to be performed cooperatively across the network because the data across the clusters may be correlated and, therefore, cooperation across clusters can be beneficial. This concept is relevant to the context of distributed estimation and adaptation over networks. In [18], a least mean square (LMS) based multitask diffusion algorithm has been presented and its performance is analyzed in detail. The performance of the multitask diffusion LMS strategy has been studied in the presence of random link failures and changing topology by extending it to the asynchronous networks [19]. It is well known that in the case of standalone adaptive filter, one major drawback of the lest mean square (LMS) algorithm is its slow convergence rate for colored input signals. For distributed networks, highly correlated inputs thus severely degrade the performance of the multitask diffusion LMS algorithm. The affine projection algorithm (APA) [20] is a better alternative to LMS in such an environment. In this paper, we propose an APA based diffusion multitask estimation scheme to obtain a good compromise between convergence performance and

computational cost. The main contributions of the proposed method include:

- A clustered multitask diffusion affine projection algorithm to estimate the multiple tasks in a distributed manner;
- Important results on mean and mean square convergence of the proposed algorithm;
- Demonstration of the effectiveness of the proposed algorithm through detailed simulations in system identification context.

II. Network Model and Proposed Algorithm

A. Clustered Multitask Network

Consider a network with N nodes which are deployed over a certain geographical area. At every time instant n, each node k has access to time realizations $\{d_k(n), \mathbf{u}_k(n)\}$ with $d_k(n)$ denoting a scalar zero mean reference signal and $\mathbf{u}_k(n) = [u_k(n), u_k(n-1), ..., u_k(n-L+1)]^T$ is a regression vector. The objective is to estimate the $L \times 1$ unknown optimal parameter vector $\mathbf{w}_k^\star$ at each node k in collaborative fashion.

In a clustered multitask network, the nodes are grouped into Q clusters with $Q \leq N$ and nodes in each cluster $\mathcal{C}_q$, $q = 1, 2, \cdots, Q$, estimate a particular optimal vector $\mathbf{w}_{C_q}^\star$, implying

$$\mathbf{w}_k^\star = \mathbf{w}_{C_q}^\star, \qquad \text{for} \quad k \in C_q. \tag{1}$$

Similarities exist among the optimal parameter vectors of the neighboring clusters, implying

$$\mathbf{w}_{C_p}^\star \sim \mathbf{w}_{C_q}^\star, \qquad \text{if} \quad C_p, C_q \ \text{ are connected}, \tag{2}$$

where p and q denote two cluster indexes. Note that the two clusters C_p and C_q are connected if there exists at least one edge linking a node from cluster C_p to a node in the cluster C_q. If all the nodes in the network are engaged in estimating a single optimal parameter vector (i.e., $\mathbf{w}_k^\star = \mathbf{w}^\star$, $k = 1, 2, \cdots, N$), then the clustered multitask network reduces to a single-task network. On the other hand, if the cluster size is one, i.e., each node k estimates its own parameter vector, then the clustered multitask network reduces to a fully multitask network.

B. Clustered Multitask Diffusion Affine Projection Algorithm

We consider here a clustered multitask network with N nodes that are grouped into Q clusters. Each node k has access to the input signal $u_k(n)$ and the observable output $d_k(n)$ that are assumed to be related via a linear model

$$d_k(n) = \mathbf{u}_k^T(n)\ \mathbf{w}_k^\star + \vartheta_k(n), \tag{3}$$

where $\mathbf{w}_k^*$ and $\mathbf{u}_k(n)$ are same as defined above. The term $\vartheta_k(n)$ is an observation noise with zero mean and variance $\sigma_{\vartheta,k}^2$ which is taken to be temporally and spatially i.i.d., and independent of input $\mathbf{u}_l(m)$ for all n, m and k, l. The nodes that are grouped in the same cluster C_q, $q = 1, 2, \cdots, Q$, estimate the same $L \times 1$ filter coefficient vector $\mathbf{w}_{C_q}^*$. We use the notation $C(k)$ to denote the cluster to which node k belongs, meaning, $C(k) \in \{C_1, C_2, \cdots, C_Q\}$.

In order to provide independence from fluctuations in input statistics, at each node k, we introduce normalized updates with respect to the input regressor $\mathbf{u}_k(n)$. Assuming the Hessian matrix of the local cost function $J_k(\mathbf{w}_{C(k)})$ which is associated with node k is positive semi-definite, the local cost function $J_k(\mathbf{w}_{C(k)})$ is defined as

$$J_k(\mathbf{w}_{C(k)}) = E\left[\left|\frac{d_k(n) - \mathbf{u}_k^T(n)\ \mathbf{w}_{C(k)}}{\|\mathbf{u}_k(n)\|}\right|^2\right]. \tag{4}$$

Consider two nodes k and l from two different clusters that are mutually connected. Then, similar to clustered multitask diffusion LMS [18], the Euclidean distance based regularizer is enforced at node k to exploit the correlation among their tasks and the squared Euclidean distance regularizer is given by

$$\Delta(\mathbf{w}_{C(k)}, \mathbf{w}_{C(l)}) = \|\mathbf{w}_{C(k)} - \mathbf{w}_{C(l)}\|^2. \tag{5}$$

Combining the local cost (4) and the regularizer (5) at each cluster level, we will have the following regularized problem in terms of Q Nash equilibrium problems [21], where each cluster C_q estimates $\mathbf{w}_{C_q}^*$ by minimizing the regularized cost function $J_{C_q}(\mathbf{w}_{C_q}, \mathbf{w}_{-C_q})$:

$$(\mathcal{P}): \ \min_{\mathbf{w}_{C_q}} J_{C_q}(\mathbf{w}_{C_q}, \mathbf{w}_{-C_q}) \quad \text{for} \quad q = 1, \cdots, Q, \tag{6}$$

where

$$J_{C_q}(\mathbf{w}_{C_q}, \mathbf{w}_{-C_q}) = \sum_{k \in C_q} E\left[\left|\frac{d_k(n) - \mathbf{u}_k^T(n)\mathbf{w}_{C(k)}}{\|\mathbf{u}_k(n)\|}\right|^2\right] + \eta \sum_{k \in C_q} \sum_{l \in \mathcal{N}_k \setminus C_q} \rho_{kl} \|\mathbf{w}_{C(k)} - \mathbf{w}_{C(l)}\|^2, \tag{7}$$

where $\mathbf{w}_{-C_q}$ denotes the collection of weight vectors estimated by the other neighboring clusters, i.e., $\mathbf{w}_{-C_q} = \{\mathbf{w}_{C(l)} | l \in \mathcal{N}_k \setminus C_q, k \in C_q\}$ and $\mathbf{w}_{C(k)} = \mathbf{w}_{C_q}$ for $k \in C_q$. The small positive constant η is the regularization strength parameter and the symbol $\setminus$ is the set difference operator. The non-negative coefficients ρ_{kl} adjust the regularizer strength between node k and l. As in [18], the non-negative coefficients ρ_{kl} are chosen to satisfy the following conditions:

$$\sum_{l=1}^{N} \rho_{kl} = 1, \quad \text{and} \quad \begin{cases} \rho_{kl} > 0, & \text{if } \ l \in \mathcal{N}_k \setminus C(k), \\ \rho_{kl} = 0, & \text{if } \ l \notin \mathcal{N}_k \setminus C(k). \end{cases} \tag{8}$$

We also impose $\rho_{kk} = 1$ if $\mathcal{N}_k \setminus C(k) = \emptyset$. Each cluster C_q estimates $\mathbf{w}_{C_q}$ by minimizing $J_{C_q}(\mathbf{w}_{C_q}, \mathbf{w}_{-C_q})$.

Following the same lines of [18], and extending the argument to apply to the problem (6), we will then have the Adapt-Then-Combine(ATC) clustered multitask diffusion strategy in

three successive stages as follows:

Adaptation:

$$\boldsymbol{\psi}_k^{'}(n+1) = \mathbf{w}_k(n) + \mu \frac{\mathbf{u}_k(n)}{\epsilon + \|\mathbf{u}_k(n)\|^2}\, e_k(n),$$

Combination (inter-cluster):

$$\boldsymbol{\psi}_k(n+1) = \boldsymbol{\psi}_k^{'}(n+1) + \mu\eta \sum_{l \in \mathcal{N}_k \setminus C(k)} \rho_{kl}\big(\mathbf{w}_l(n) - \mathbf{w}_k(n)\big),$$

Combination (intra-cluster):

$$\mathbf{w}_k(n+1) = \sum_{l \in \mathcal{N}_k \cap C(k)} a_{lk}\ \boldsymbol{\psi}_l(n+1), \tag{9}$$

where the combination coefficients a_{lk} are non-negative and are given by

$$\sum_{l=1}^{N} a_{lk} = 1, \quad \text{and} \quad \begin{cases} a_{lk} > 0, & \text{if} \quad l \in \mathcal{N}_k \cap C(k), \\ a_{lk} = 0, & \text{otherwise.} \end{cases} \tag{10}$$

Several methods exist in literature to select the coefficients a_{lk} such as averaging rule, Metropolis rule etc. [1].

One of the major limitations of the NLMS algorithm is its slow convergence rate for highly correlated input signals. In such environments, the Affine Projection Algorithm (APA) [20] is a better alternative to NLMS. As seen earlier, the APA updates the weight vector using the current input regressor vector along with the $P-1$ past input regressor vectors (P: projection order), whereas the NLMS uses only current input regressor vector. The adjustment term $\mu\ \dfrac{\mathbf{u}_k(n)\ e_k(n)}{\epsilon + \|\mathbf{u}_k(n)\|^2}$ of the NLMS weight vector updation is replaced in APA by the more generalized term $\mu\ \mathbf{U}_k(n)\ \big(\epsilon\mathbf{I}_P + \mathbf{U}_k^T(n)\mathbf{U}_k(n)\big)^{-1}\mathbf{e}_k(n)$, where $\mathbf{I}$ is the identity matrix, ϵ is a small positive constant used to avoid the inversion of a rank deficient matrix $\mathbf{U}_k^T(n)\mathbf{U}_k(n)$, $\mathbf{U}_k(n) = \big[\mathbf{u}_k(n), \mathbf{u}_k(n-1), \cdots, \mathbf{u}_k(n-P+1)\big]$ is the input signal matrix, $\mathbf{d}_k(n) = [d_k(n), d_k(n-1), \cdots, d_k(n-P+1)]^T$ is the desired response vector and $\mathbf{e}_k(n) = \mathbf{d}_k(n) - \mathbf{U}_k^T(n)\mathbf{w}_k(n) \equiv [e_k(n), e_k(n-1), \cdots, e_k(n-P+1)]^T$, $k = 1, 2, \cdots, N$ is the error vector.

To make the system more robust against correlated input, we then extend the above clustered multitask diffusion NLMS strategy to APA. From (9), we can obtain the following ATC based clustered multitask diffusion Affine Projection Algorithm:

Adaptation:

$$\boldsymbol{\psi}_k^{'}(n+1) = \mathbf{w}_k(n) + \mu\mathbf{U}_k(n)\big(\epsilon\mathbf{I}_P + \mathbf{U}_k^T(n)\mathbf{U}_k(n)\big)^{-1}\mathbf{e}_k(n),$$

Combination (inter-cluster):

$$\boldsymbol{\psi}_k(n+1) = \boldsymbol{\psi}_k^{'}(n+1) + \mu\eta \sum_{l \in \mathcal{N}_k \setminus C(k)} \rho_{kl}\big(\mathbf{w}_l(n) - \mathbf{w}_k(n)\big),$$

Combination (intra-cluster):

$$\mathbf{w}_k(n+1) = \sum_{l \in \mathcal{N}_k \cap C(k)} a_{lk}\ \boldsymbol{\psi}_l(n+1). \tag{11}$$

It is seen that when $P = 1$, the clustered multitask diffusion APA (11) reduces to the clustered multitask diffusion NLMS (9).

III. Performance Analysis

In this section, we present some important results on the convergence of the proposed clustered multitask diffusion strategy. For this, we define the network level optimal filter coefficient vector $\mathbf{w}^\star$, estimated filter coefficient vector $\mathbf{w}(n)$, input data matrix $\mathbf{U}(n)$ and the observation noise vector $\boldsymbol{\vartheta}(n)$ as follows:

$$\begin{aligned} \mathbf{w}^\star &= \text{col}\big\{\mathbf{w}_1^\star, \mathbf{w}_2^\star, \cdots, \mathbf{w}_N^\star\big\}, \\ \mathbf{w}(n) &= \text{col}\big\{\mathbf{w}_1(n), \mathbf{w}_2(n), \cdots, \mathbf{w}_N(n)\big\}, \\ \boldsymbol{\vartheta}(n) &= \text{col}\big\{\boldsymbol{\vartheta}_1(n), \boldsymbol{\vartheta}_2(n), \cdots, \boldsymbol{\vartheta}_N(n)\big\}, \\ \mathbf{U}(n) &= \begin{bmatrix} \mathbf{U}_1(n) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_2(n) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{U}_N(n) \end{bmatrix}, \end{aligned} \tag{12}$$

where col{.} is used to denote the column wise stacking operator. From these definitions, network level data model is given by

$$\mathbf{d}(n) = \text{col}\big\{\mathbf{d}_1(n), \mathbf{d}_2(n), \cdots, \mathbf{d}_N(n)\big\} = \mathbf{U}^T(n)\mathbf{w}^\star + \boldsymbol{\vartheta}(n), \tag{13}$$

and the global error vector is given by

$$\mathbf{e}(n) = \text{col}\big\{\mathbf{e}_1(n), \mathbf{e}_2(n), \cdots, \mathbf{e}_N(n)\big\} = \mathbf{d}(n) - \mathbf{U}^T(n)\mathbf{w}(n). \tag{14}$$

Using these definitions, at network level, the weight update recursion of the proposed clustered multitask diffusion APA can then be stated as follows:

$$\begin{aligned} \mathbf{w}(n+1) = \boldsymbol{\mathcal{A}}\big(\mathbf{w}(n) + \mu\mathbf{U}(n)\big(\epsilon\mathbf{I}_{PN} + \mathbf{U}^T(n)\mathbf{U}(n)\big)^{-1}\mathbf{e}(n)\big) \\ - \mu\,\eta\,\boldsymbol{\mathcal{A}}\,\boldsymbol{\mathcal{Q}}\,\mathbf{w}(n), \end{aligned} \tag{15}$$

where

$$\begin{aligned} \boldsymbol{\mathcal{A}} &= \mathbf{A}^T \otimes \mathbf{I}_L, \\ \boldsymbol{\mathcal{Q}} &= \mathbf{I}_{LN} - \mathbf{P} \otimes \mathbf{I}_L, \end{aligned} \tag{16}$$

with $\otimes$ denoting the right Kronecker product operator (i.e., for two matrices $\mathbf{X}$ and $\mathbf{Y}$ of size $M \times N$ and $L \times K$ respectively, $\mathbf{X} \otimes \mathbf{Y}$ is a block matrix of size $ML \times KN$, with the $(i,j)^{th}$ block given by $x_{ij}\mathbf{Y}, i = 1, 2, \cdots, M, j = 1, 2, \cdots, N$), $\mathbf{A}$ with $[\mathbf{A}]_{l,k} = a_{lk}$ is a $N \times N$ left stochastic matrix (i.e., each column consists of non-negative real numbers whose sum is unity) that defines the network topology, and $\mathbf{P}$ with $[\mathbf{P}]_{k,l} = \rho_{kl}$ is a $N \times N$ asymmetric right stochastic matrix (i.e., each row consists of non-negative real numbers whose sum is unity) that defines regularizer strength among the nodes.

Denoting the global weight deviation vector of the proposed clustered multitask diffusion APA at n^{th} index as $\widetilde{\mathbf{w}}(n) =$

$\mathbf{w}^\star - \mathbf{w}(n)$, recalling that $\boldsymbol{\mathcal{A}}\mathbf{w}^\star = \mathbf{w}^\star$, from (15), the recursion for $\widetilde{\mathbf{w}}(n)$ can then be written as

$$\widetilde{\mathbf{w}}(n+1) = \boldsymbol{\mathcal{B}}(n)\widetilde{\mathbf{w}}(n) - \mu\boldsymbol{\mathcal{A}}\mathbf{U}(n)\big(\epsilon\mathbf{I}_{PN} + \mathbf{U}^T(n)\mathbf{U}(n)\big)^{-1}\boldsymbol{\vartheta}(n) + \mathbf{r}, \tag{17}$$

where $\boldsymbol{\mathcal{B}}(n) = \boldsymbol{\mathcal{A}}\big(\mathbf{I}_{LN} - \mu\mathbf{Z}(n) - \mu\eta\boldsymbol{\mathcal{Q}}\big)$ with $\mathbf{Z}(n) = \mathbf{U}(n)(\epsilon\mathbf{I}_{PN} + \mathbf{U}^T(n)\mathbf{U}(n))^{-1}\mathbf{U}^T(n)$ and $\mathbf{r} = \mu\eta\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\mathbf{w}^\star$.

To obtain the results on convergence of the proposed clustered multitask diffusion algorithm, we make the following assumptions :

Assumption 1: The data signal $u_k(n)$ arises from a random process that is temporally stationary with correlation matrix $\mathbf{R}_{u,k} = E[\mathbf{u}_k(n)\mathbf{u}_k^T(n)]$, and also the data matrices $\mathbf{U}_k(n)$, $k = 1, 2, \cdots N$ are spatially independent.

Assumption 2: The observation noise $\vartheta_k(n)$ is taken to be spatially and temporally i.i.d. Gaussian with mean zero and variance $\sigma^2_{\vartheta,k}$.

Assumption 3: The network topology is assumed to be static, meaning the combiner coefficients are constant throughout the process.

Assumption 4: The step size μ is sufficiently small so that the terms involving higher order moments of μ in in the matrix $\boldsymbol{\mathcal{F}} = E\big[\boldsymbol{\mathcal{B}}(n) \otimes_b \boldsymbol{\mathcal{B}}(n)\big]$ ($\otimes_b$ denotes the right block Kronecker product operator [22]) can be ignored.

The above assumptions are commonly used in the analysis of diffusion adaptive strategies to simplify derivations and they do not alter the operation of the algorithm.

Theorem 1. ***Convergence in the mean****: Assuming the data model* (13) *and the Assumptions* 1-3 *to hold, a sufficient condition for the proposed clustered multitask diffusion APA to converge in mean is*

$$0 < \mu < \frac{2}{\max\limits_{1\le k\le N}\big\{\max\limits_{1\le i\le L}\{\lambda_i(\overline{\mathbf{Z}}_k)\}\big\} + 2\,\eta}, \tag{18}$$

where $\overline{\mathbf{Z}}_k = E\Big[\mathbf{U}_k(n)\big(\epsilon\boldsymbol{I}_P + \mathbf{U}_k^T(n)\mathbf{U}_k(n)\big)^{-1}\ \mathbf{U}_k^T(n)\Big]$ *with* $\lambda_i(\cdot)$ *denoting the* i*-th eigenvalue of its argument matrix.* $E[\mathbf{Z}_k(n)]$ *is independent of* n *due to stationarity of* $\mathbf{u}_k(n)$*. As a result, we have dropped the index* n *from* $\overline{\mathbf{Z}}_k$*.*

Proof: Skipped due to page limitation. ■

Theorem 2. ***Convergence in the mean square****: Assuming the data model* (13) *and the Assumptions* 1-4 *to hold, the proposed clustered multitask diffusion APA converges in mean square sense if the matrix* $\boldsymbol{\mathcal{F}} = E\big[\boldsymbol{\mathcal{B}}(n) \otimes_b \boldsymbol{\mathcal{B}}(n)\big]$ *is stable, which is guaranteed under*

$$0 < \mu < \frac{1}{\max\limits_{1\le k\le N}\big\{\max\limits_{1\le i\le L}\{\lambda_i(\overline{\mathbf{Z}}_k)\}\big\} + 2\,\eta}. \tag{19}$$

Proof: Skipped due to page limitation. ■

IV. Simulation Studies and Discussion

In this section, we demonstrate performance of the proposed clustered multitask diffusion algorithm via simulation studies. For simulation, we considered a clustered multitask network consisting of $N = 21$ nodes with the topology shown in Fig. 1. The nodes in the network are grouped in 4 clusters: $C_1 = \{1, 2, 18, 19, 20, 21\}$, $C_2 = \{12, 13, 14, 15, 16, 17\}$, $C_3 = \{8, 9, 10, 11\}$ and $C_4 = \{3, 4, 5, 6, 7\}$. These clusters aim to estimate their respective 256 tap optimal vectors in collaborative fashion which are chosen as, $\mathbf{w}^*_{C_q} = \mathbf{w}_0 + \delta_{C_q}\mathbf{w}_0$ for $q = 1, 2, 3, 4$, with $\delta_{C_1} = 0, \delta_{C_2} = -0.025$, $\delta_{C_3} = 0.05$ and $\delta_{C_4} = -0.05$. The coefficient vector $\mathbf{w}_0$ is generated from zero mean, unity variance Gaussian distribution.

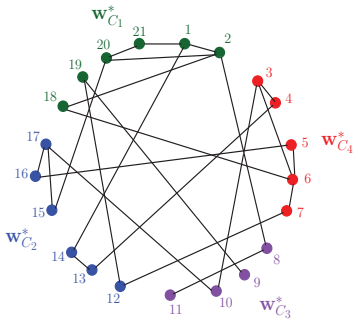


Fig. 1. Network Topology

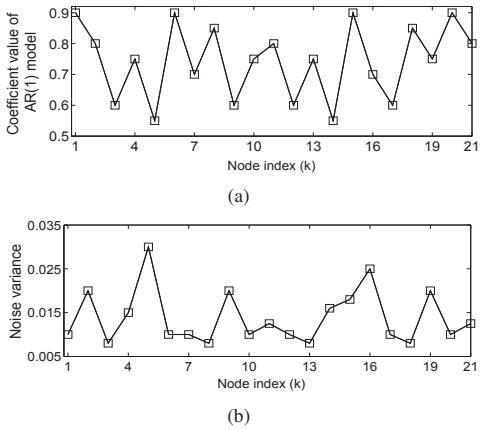


Fig. 2. Statistical settings of the network: a). coefficient value of AR(1) model θ_k; b). noise variance $\sigma^2_{\vartheta,k}$.

Simulations are conducted for colored Gaussian input of unit variance, where at each node k, a unity variance colored, Gaussian input $u_k(n)$ is generated by driving the following first order auto regressive (AR) model: $u_k(n) = \theta_k\, u_k(n-1) + \sqrt{1-\theta_k^2}\, z_k(n)$, $|\theta_k| < 1$, with a unity variance, white Gaussian input $z_k(n)$. The coefficient θ_k varies from node to node and its distribution against the node index k is shown in Fig. 2(a). The observation noise $\vartheta_k(n)$ is taken to be zero

mean, i.i.d. Gaussian with variance $\sigma^2_{\vartheta,k}$, which is plotted against k in Fig. 2(b).

At each node, the projection order is fixed at $P = 4$ and the initial values of the taps are taken to be zero. The step size μ is set at 0.5 for all the nodes. Similar to [18], the regularization coefficients ρ_{kl} are set to $\rho_{kl} = |\mathcal{N}_k \setminus C(k)|^{-1}$ (where $|\cdot|$ denotes the cardinality of the set) for $l \in \mathcal{N}_k \setminus C(k)$ and $\rho_{kl} = 0$ for any other l ($\rho_{kk} = 1$ if $\mathcal{N}_k \setminus C(k) = \emptyset$). The combining coefficients a_{lk} are obtained using the Metropolis rule [1]. Detailed simulations are carried out to study the performance of several learning scenarios such as 1). Clustered multitask diffusion APA; 2). Multitask diffusion APA (which is obtained by assigning a cluster to each node and setting $\eta \neq 0$ and the combiner matrix $\mathbf{A} = \mathbf{I}$ in the clustered multitask diffusion APA); 3). Non-cooperative APA (which is obtained by setting $\eta = 0$ and the combiner matrix $\mathbf{A} = \mathbf{I}$ in the clustered multitask diffusion APA).

The simulation results are displayed by plotting the network level normalized MSD (in dB) against the iteration index n, obtained by averaging over 100 independent experiments. The resulting plots, which are popularly known as learning curves are shown in Fig. 3.

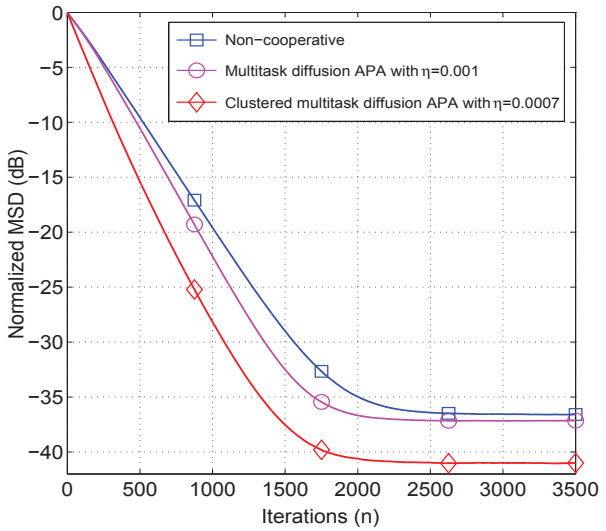
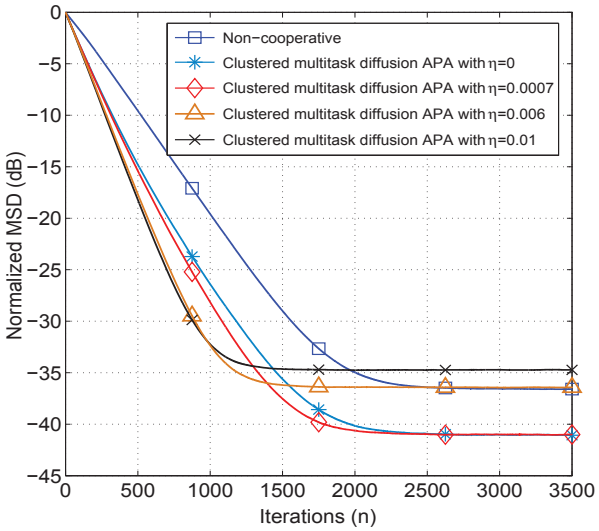


Fig. 3. Learning curves of the proposed clustered multitask diffusion APA strategy at network level.

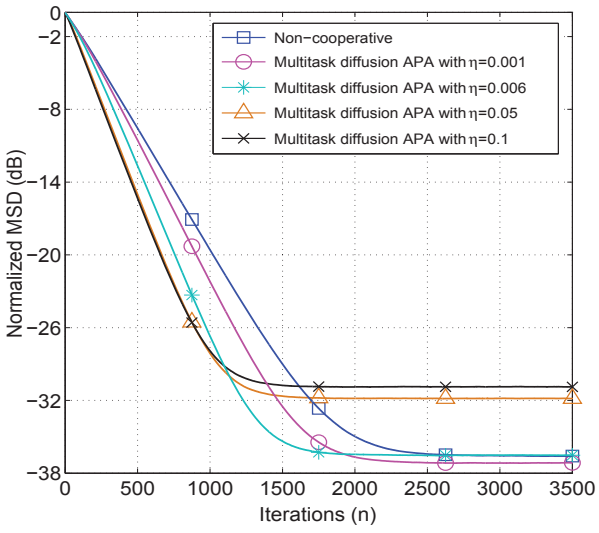
From Fig. 3, several observations can be made as described below:

1) As the nodes do not collaborate for additional benefit, the non-cooperative APA exhibits poor performance in terms of convergence rate and steady-state MSD.
2) In multitask diffusion APA, the regularization term in adaptation stage enables inter-cluster cooperation among nodes, leading to improved performance over non-cooperative APA.
3) In clustered multitask diffusion APA, the presence of both inter-cluster cooperation and intra-cluster cooperation help to achieve improved performance over the aforementioned strategies such as multitask and non-cooperative APA.

To investigate the effects of η on the performance of clustered multitask diffusion APA and multitask diffusion APA (i.e, $\mathbf{A} = \mathbf{I}$), the above simulation exercise was carried out for different values of η. The corresponding learning curves are shown in Figs. 4(a) and 4(b) for clustered multitask diffusion APA and multitask diffusion APA, respectively.



(a)



(b)

Fig. 4. Learning curves of the proposed strategy at network level for different values of η: (a). clustered multitask diffusion APA; (b). multitask diffusion APA

From Fig. 4(a), it can be observed that for $\eta = 0$ (i.e., absence of inter-cluster cooperation), the clustered multitask diffusion APA shows superior performance over non-cooperative strategy in terms of both convergence rate and

steady-state MSD by exploiting the intra-cluster collaboration alone. As η increases, say to $\eta = 0.0007$, the presence of inter cluster collaboration along with the intra-cluster collaboration results in further improvement in performance. However, as η increases still further, say to $\eta = 0.006$, the convergence rate of course increases, but with considerable degradation in steady-state MSD. It is also seen that for $\eta = 0.006$, the steady-state MSD of the clustered multitask diffusion APA is at par with that for the non-cooperative strategy, though the convergence rate is greatly improved. Beyond this point, e.g., for $\eta = 0.01$, it is seen that the steady-state MSD degrades further with no improvement in convergence rate.

Similarly, from Fig. 4(b), it can be observed that for $\eta = 0.001$, the multitask diffusion APA exhibits superior performance over non-cooperative strategy in terms of both convergence rate and steady-state MSD. However, as η increases, say to $\eta = 0.006$, the convergence rate increases further with slight degradation in steady-state MSD. Infact, for $\eta = 0.006$, the steady-state MSD of the multitask diffusion APA is at par with that for the non-cooperative strategy, though the convergence rate is greatly improved. Beyond this point, as η increases further, we observe very little improvement in convergence but significant degradation in steady-state MSD performance. Beyond the value of $\eta = 0.05$, it is seen that the steady-state MSD degrades further with no improvement in convergence rate.

From the above results, it can be seen that the performance of clustered multitask diffusion APA and multitask diffusion APA is greatly dependent on the value of regularization strength η. Since the neighboring cluster tasks are only having some kind of similarity relationship, but are not exactly same, the inter-cluster cooperation has to be terminated near convergence. However, the clustered multitask diffusion strategy continues the inter-cluster cooperation even in the steady-state and the degree of cooperation is in proportion to the value of η. Due to this lack of control on inter-cluster cooperation, the steady-state MSD performance of clustered multitask diffusion APA deteriorates.

V. Conclusion

In this paper, we presented a clustered multitask diffusion strategy for simultaneously estimating the multiple tasks over distributed adaptive networks. The proposed clustered multitask diffusion strategy is robust against the correlated input conditions. By exploiting both inter-cluster cooperation and intra-cluster cooperation, the proposed clustered multitask diffusion APA achieves improved performance over non-cooperative strategy in terms of convergence rate and steady-state MSD. The performance of the proposed strategy is demonstrated through detailed simulations.

References

[1] A. H . Sayed, "Diffusion adaptation over networks," in *Academic Press Library in Signal Processing,* R. Chellappa and S. Theodoridis, Eds., pp. 322-454, Elsevier, 2013. Also available as arXiv:1205.4220 [cs.MA], May 2012.

[2] A. H. Sayed, *Adaptation, Learning, and Optimization over Networks, Foundations and Trends in Machine Learning,* vol. 7, issue 4-5, NOW Publishers, Boston-Delft, 2014.

[3] A. H. Sayed, "Adaptive Networks," in *Proc. of IEEE*, vol. 102, no. 4, pp. 460-497, Apr. 2014.

[4] C. G. Lopes and A. H. Sayed, "Incremental adaptive strategies over distributed networks," *IEEE Trans. Signal Process.,* vol. 55, no. 8, pp. 4064-4077, Aug. 2007.

[5] L. Li, J. A. Chambers, C. G. Lopes, and A. H. Sayed, "Distributed estimation over an adaptive incremental network based on the Affine Projection Algorithm," *IEEE Trans. Signal Process.,* vol. 58, no. 1, pp. 151-164, 2010.

[6] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: formulation and performance analysis," *IEEE Trans. Signal Process.,* vol. 56, no. 7, pp. 3122-3136, Jul. 2008.

[7] F. S. Cattivelli and A. H. Sayed, "Diffusion LMS stretagies for distributed estimation," *IEEE Trans. Signal Process.,* vol. 58, no. 3, pp. 1035-1048, May 2010.

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

[9] L. Li and J. A. Chambers, "Distributed adaptive estimation based on the APA algorithm over diffusion networks with changing topology," in *Proc. IEEE/SP Workshop on Statistical Signal Process.,* Cardiff, 2009, pp. 757-760.

[10] A. H . Sayed, S.-Y. Tu, J. Chen, X. Zhao and Z. J. Towfic, "Diffusion strategies for adaptation and learning over networks," *IEEE Signal Process. Mag.,* vol. 30, no. 3, pp. 155-171, May 2013.

[11] J. Plata-Chaves, A. Bertrand and M. Moonen, "Incremental multiple error filtered-X LMS for node-specific active noise control over wireless acoustic sensor networks," in *Proc. IEEE Sensor Array and Multichannel Signal Process. Workshop (SAM)*, Rio de Janerio, 2016, pp. 1-5.

[12] J. Plata-Chaves, A. Bertrand, M. Moonen, S. Theodoridis and A. M. Zoubir, "Heterogeneous and multitask wireless sensor networks—algorithms, applications, and challenges,"*IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 3, pp. 450-465, Apr. 2017.

[13] A. Hassani, J. Plata-Chaves, M. H. Bahari, M. Moonen and A. Bertrand, "Multi-task wireless sensor network for joint distributed node-specific signal enhancement, LCMV beamforming and DOA estimation," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 3, pp. 518-533, Apr. 2017.

[14] S. Monajemi, K. Eftaxias, S. Sanei and S. H. Ong, "An informed multitask diffusion adaptation approach to study tremor in Parkinson's disease," *IEEE J. Sel. Topics Signal Process.,* vol. 10, no. 7, pp. 1306-1314, Oct. 2016.

[15] J. Chen, C. Richard and A. H. Sayed, "Diffusion LMS for clustered multitask networks," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, Florence, 2014, pp. 5487-5491.

[16] N. Bogdanovic, J. Platta-Chaves and K. Berberidis, "Distributed incremental-based LMS for node specific parameter estimation over adaptive networks," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, Vancouver, 2013, pp. 5425-5429.

[17] J. Plata-Chaves, N. Bogdanovic and K. Berberidis, "Distributed diffusion-based LMS for node-specific adaptive parameter estimation," *IEEE Trans. Signal Process.,*, vol. 63, no. 13, pp. 3448-3460, Jul. 2015.

[18] J. Chen, C. Richard and A. H. Sayed, "Multitask diffusion adaptation over networks," in *IEEE Trans. Signal Process.,* vol. 62, no. 16, pp. 4129-4144, Aug. 2014.

[19] R. Nassif, C. Richard, A. Ferrari and A. H. Sayed, "Multitask diffusion adaptation over asynchronous networks," *IEEE Trans. on Signal Process.*, vol. 64, no. 11, pp. 2835-2850, Jun. 2016.

[20] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Electron. Commun. Japan*, vol. 67-A, no. 5, pp. 19-27, 1984.

[21] T. Basar and G. Olsder, *Dynamic noncooperative game theory*, Society for Industrial and Applied Mathematics, 1998.

[22] D. S. Tracy and R. P. Singh, "A new matrix product and its applications in partitioned matrix differentiation," *Statistica Neerlandica.,* vol. 26, no. NR.4, pp. 143-157, 1972.