DISTRIBUTED INCREMENTAL-BASED LMS FOR NODE-SPECIFIC PARAMETER ESTIMATION OVER ADAPTIVE NETWORKS

Nikola Bogdanović, Jorge Plata-Chaves, Kostas Berberidis

Department of Computer Engineering and Informatics
University of Patras & C.T.I RU-8, 26500, Rio - Patra, Greece
E-mails: {bogdanovic, jorge, berberid}@ceid.upatras.gr

ABSTRACT

We introduce an adaptive distributed technique that is suitable for node-specific parameter estimation in an adaptive network where each node is interested in a set of parameters of local interest as well as a set of network global parameters. The estimation of each set of parameters of local interest is undertaken by a local Least Mean Squares (LMS) algorithm at each node. At the same time and coupled with the previous local estimation processes, an incremental mode of cooperation is implemented at all nodes in order to perform an LMS algorithm which estimates the parameters of global interest. In the steady state, the new distributed technique converges to the MMSE solution of a centralized processor that is able to process all the observations. To illustrate the effectiveness of the proposed technique we provide simulation results in the context of cooperative spectrum sensing in cognitive radio networks.

Index Terms— Adaptive distributed networks, incremental algorithm, cooperation, node-specific parameter estimation.

1. INTRODUCTION

With the aim of enabling an energy aware and low-complex distributed implementation of the estimation task, several useful optimization techniques that generally yield linear estimators have been derived. Among them, the techniques following a consensus approach have been extensively applied. In some initial works, for instance [1], the implementation of the consensus strategy is done in two stages. Unfortunately, this kind of implementation is not suitable for real time estimation as required in time-varying environments. Afterwards, motivated by the procedure obtained in [2], alternative implementations of the consensus strategy were presented in the literature (e.g., [3]-[4]) which force agreement among the cooperating nodes in a single time-scale. However, to do so they require conditions that do not let the network undertake a continuous learning and optimization [5]. The latter issue is actually the main one addressed in a single time-scale by distributed estimation algorithms that are based on adaptive filtering techniques. In particular, in this kind of techniques a linear estimator is obtained by distributing a specific stochastic gradient method under an incremental or a diffusion mode of cooperation. In the incremental mode (e.g., [6]-[7]) each node communicates with only one neighbor, and consequently the data are processed in a cyclic manner throughout the network. Better reliability can be achieved at the expense of increased energy consumption in the so-called diffusion mode considered, for instance, in [8]-[10]. Under this strategy, each node can communicate with a subset of neighboring nodes.

Although there are many published techniques addressing different distributed estimation problems, only very few papers consider node-specific settings where the nodes have overlapped but different estimation interests. Possibly, some of the first works explicitly considering the aforementioned setting are [11]-[12]. In these works, for networks with a fully connected and tree topology, Bertrand *et al* proposed distributed algorithms that allow to estimate node-specific desired signals sharing a common latent signal subspace.

There are also a few recent works deriving distributed schemes dealing with problems which can be considered as Node-Specific Parameter Estimation (NSPE) problems. The consensus approach presented in [13] is based on optimization techniques that force different nodes to reach an agreement when estimating parameters of common interest. In the case of schemes based on a distributed implementation of adaptive filtering techniques, the literature is less extensive. In one of these works [14], the authors use diffusion adaption and scalarization techniques to solve the multi-objective cost function that appears in a NSPE problem where there is no prior knowledge regarding the overlapped estimation interests.

To the authors knowledge, there is no existing literature dealing with incremental-type adaptive filtering for NSPE problems in which the existence of the overlapped estimation interests is known a priori. Motivated by this, as well as by the fact that the aforementioned prior information is available in many applications [13], we state in this paper a new NSPE formulation where all nodes are interested in estimating some parameters of local interest as well as some parameters of global interest. Afterwards, we derive a novel distributed estimation algorithm that converges to the minimum mean-square error (MMSE) solution of the corresponding centralized problem. Similarly to the hierarchical identification principle [15], it undertakes distinct but coupled optimization processes in an adaptive way. All of them but one employ a Least Mean Square (LMS) algorithm that runs locally at each node with the aim of estimating its parameters of local interest. The remaining optimization process, associated with the estimation of the parameters of global interest, relies again on an LMS algorithm implemented at all nodes under an incremental mode of cooperation. Finally, we provide an illustrative application for power spectrum sensing in cognitive radio. In this application, each secondary user (SU) is interested in estimating the power spectrum of the primary users (PU), which is a task of global interest, and the power spectrum of its local interference source(s).

We use boldface letters for random variables and normal fonts for deterministic quantities. Capital letters refer to matrices and small letters refer to both vectors and scalars. The Hermitian transposition and the expectation operator are denoted by $(\cdot)^H$ and $E\{\cdot\}$,

The work was partially supported by the European SmartEN ITN project (Grant No. 238726) under the Marie Curie ITN FP7 program and by the University of Patras.

respectively. Moreover, $R_{\mathbf{A}} = E\{\mathbf{A}^H\mathbf{A}\}$, $R_{\mathbf{A},\mathbf{B}} = E\{\mathbf{A}^H\mathbf{B}\}$ and $r_{\mathbf{A},\mathbf{b}} = E\{\mathbf{A}^H\mathbf{b}\}$ for any random matrices $\mathbf{A}$, $\mathbf{B}$ and any random vector $\mathbf{b}$. Finally, $0_{L\times M}$ denotes a $L \times M$ zero matrix.

2. PROBLEM STATEMENT

Let us consider a network consisting of N nodes randomly deployed over some region (see Fig. 1). Each node k, at discrete time i, has access to data $\{d_{k,i}, U_{k,i}\}$. These data are related to events that take place in the network area and follow the subsequent model

$$\mathbf{d}_{k,i} = \mathbf{U}_{k,i} w_k^o + \mathbf{v}_{k,i} \tag{1}$$

where, for each time instant i,

- w_k^o equals the vector of dimension M_k that gathers all parameters of interest for node k,
- $\mathbf{v}_{k,i}$ denotes measurement and/or model noise with zero mean and covariance matrix $R_{v_k,i}$ of dimensions $L_k \times L_k$,
- $\mathbf{d}_{k,i}$ and $\mathbf{U}_{k,i}$ are zero-mean random variables with dimensions $L_k \times 1$ and $L_k \times M_k$, respectively.

Given the previous observation model, the objective for each node in the network is to use data $\{d_{k,i}, U_{k,i}\}$ in order to estimate its specific unknown vector w_k^o. In particular, we seek the set of linear node-specific estimators $\{w_k\}_{k=1}^N$ that minimize the following global cost function

$$J_{\text{glob}}(\{w_k\}_{k=1}^N) = \sum_{k=1}^{N} E\left\{\|\mathbf{d}_{k,i} - \mathbf{U}_{k,i} w_k\|^2\right\}. \tag{2}$$

In most of the existing papers, e.g., [6]-[8], the derived adaptation strategies minimize (2) when $w_k^o = w^o$ for all $k \in \{1, 2, \ldots, N\}$. However, note that the formulation considered in this paper goes one step further by considering a more general scenario where not all node-specific parameters of interest, i.e. $\{w_k^o\}_{k=1}^N$, are the same. Instead, we allow some of these parameters to differ from one node to another. Specifically, as shown in Fig. 1, each vector $\{w_k^o\}_{k=1}^N$ consists of globally common components as well as of components of local interest for node k. On the one hand, the parameters of global interest in the network may account for an event common to all nodes. On the other hand, the parameters of local interest for each node k may reflect an influence of some local phenomena that is possibly different for each node. In this way, the observation model provided in (1) can be reformulated as

$$\begin{aligned}\mathbf{d}_{k,i} &= \begin{bmatrix}\mathbf{U}_{k_g,i} & \mathbf{U}_{k_l,i}\end{bmatrix} \cdot \begin{bmatrix} w^o \\ \xi_k^o \end{bmatrix} + \mathbf{v}_k^{(i)} \\ &= \mathbf{U}_{k_g,i} w^o + \mathbf{U}_{k_l,i} \xi_k^o + \mathbf{v}_k^{(i)}\end{aligned} \tag{3}$$

where, for each $k \in \{1, 2, \ldots, N\}$, and time instant i,

- w^o is a sub-vector of dimension $M_{k_g} \times 1$ that gathers all the parameters of global interest,
- ξ_k^o is a sub-vector of dimension $M_{k_l} \times 1$ that gathers all the parameters of local interest,
- $\mathbf{U}_{k_g,i}$ and $\mathbf{U}_{k_l,i}$ are matrices of dimensions $L_k \times M_{k_g}$ and $L_k \times M_{k_l}$ that might be correlated, and consist of the columns $\mathbf{U}_{k,i}$ associated with w^o and ξ_k^o, respectively.

Thus, according to (2) and (3), our NSPE problem can be restated as

$$\begin{aligned}&\{\hat{w}, \{\hat{\xi}_k\}_{k=1}^N\} \\ &= \underset{w,\{\xi_k\}_{k=1}^N}{\text{argmin}} \left\{\sum_{k=1}^{N} E\left\{\|\mathbf{d}_{k,i} - \mathbf{U}_{k_g,i} w - \mathbf{U}_{k_l,i}\xi_k\|^2\right\}\right\}.\end{aligned} \tag{4}$$

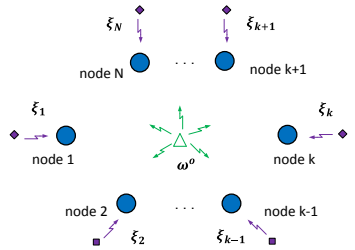

Fig. 1. Network with node-specific parameter estimation interests.

3. A SOLUTION OF THE NEW NSPE PROBLEM

In this section, first we derive a centralized solution of the optimization problem (4), and then we develop a distributed strategy that converges to this centralized solution. For the sake of simplicity and without losing generality, we assume that $M_k = M$, $M_{k_g} = M_g$, $M_{k_l} = M_l$ and $L_k = L$ for all $k \in \{1, 2, \ldots, N\}$.

3.1. Centralized solution

As it can be seen by inspecting (4), in order to solve the considered NSPE problem we have to optimize a scalar real-valued cost function w.r.t. multiple vector variables, i.e., $\{w, \{\xi_k\}_{k=1}^N\}$. After gathering these variables into the following augmented vector

$$\tilde{w} = \begin{bmatrix} w^T & \xi_1^T & \xi_2^T & \cdots & \xi_N^T \end{bmatrix}^T \quad (\widetilde{M} \times 1) \tag{5}$$

where $\widetilde{M} = M_g + N \cdot M_l$, we can easily verify that our optimization problem is equivalent to

$$\widehat{\tilde{w}} = \underset{\tilde{w}}{\text{argmin}} \{J(\tilde{w})\} = \underset{\tilde{w}}{\text{argmin}} \left\{\sum_{k=1}^{N} E\left\{\|\mathbf{d}_{k,i} - \widetilde{\mathbf{U}}_{k,i}\tilde{w}\|^2\right\}\right\} \tag{6}$$

where

$$\widetilde{\mathbf{U}}_{k,i} = \begin{bmatrix}\mathbf{U}_{k_g,i} & 0_{L\times M_a} & \mathbf{U}_{k_l,i} & 0_{L\times M_b}\end{bmatrix} \tag{7}$$

with $M_a = (k-1)M_l$ and $M_b = (N-k)M_l$. It is well-known that the resulting solutions $\widehat{\tilde{w}}$ are given by the normal equations [16]

$$\left(\sum_{k=1}^{N} R_{\widetilde{U}_{k,i}}\right) \cdot \widehat{\tilde{w}} = \sum_{k=1}^{N} r_{\widetilde{U}_{k,i} d_{k,i}}. \tag{8}$$

However, this centralized batch solution requires the inversion of a square matrix whose dimension is actually proportional to the number of nodes N, and hence, a prohitively high computational cost is needed. To alleviate this problem, different iterative procedures can be followed, e.g., an iterative steepest descent method [16].

3.2. Distributed solution

To improve energy efficiency, robustness and scalability of the previously described centralized approach, it is highly desirable to design a distributed and adaptive scheme for the computation of $\widehat{\tilde{w}}$. Toward this goal, our starting point is a fully distributed incremental method that generally has better rate of convergence and steady-state performance than its corresponding steepest descent method [6].

Taking into account that our global cost function $J(\tilde{w})$ is expressed as the sum of N local cost functions $\{J_k(\tilde{w})\}_{k=1}^N$ with

$$J_k(\tilde{w}) = E\left\{\|\mathbf{d}_{k,i} - \widetilde{\mathbf{U}}_{k,i}\tilde{w}\|^2\right\}, \tag{9}$$

at each time instant $i \in \{1,2,\ldots\}$ a distributed incremental-based algorithm for (6) should perform the following steps for all $k \in \{1,2,\ldots,N\}$ and some initialization $\tilde{\psi}_N^{(0)}$

$$\begin{cases} \tilde{\psi}_0^{(i)} \leftarrow \tilde{\psi}_N^{(i-1)} \\ \tilde{\psi}_k^{(i)} = \tilde{\psi}_{k-1}^{(i)} - \mu_k \left[\nabla J_k(\tilde{\psi}_{k-1}^{(i)})\right]^H \end{cases} \tag{10}$$

with

$$\left[\nabla J_k(\tilde{\psi}_{k-1}^{(i)})\right]^H = R_{\widetilde{U}_{k,i}}\tilde{\psi}_{k-1}^{(i)} - r_{\widetilde{U}_{k,i}d_{k,i}} \tag{11}$$

where $\tilde{\psi}_k^{(i)}$ denotes a local estimate of $\tilde{w}^o$ at node k and time i according to (6), $\tilde{\psi}_N^{(0)}$ equals an initial guess about $\tilde{w}^o$, and $\mu_k > 0$ is a suitable chosen positive step-size parameter. For sufficiently small step-sizes $\{\mu_k\}_{k=1}^N$, and relying on locally available information, this strategy ensures that $\lim_{i\to\infty}\tilde{\psi}_k^{(i)} = \widehat{\tilde{w}}$ with $k \in \{1,2,\ldots,N\}$ and $\widetilde{\tilde{w}}$ satisfying (8) (see [17] and [18]). The aforementioned convergence is achieved by a cyclic cooperation where each node of the network transmits its local estimate $\tilde{\psi}_k^{(i)}$ to only one immediate neighbor. Nonetheless, since the dimension of $\tilde{\psi}_k^{(i)}$ depends on the number of nodes, this kind of iterative solutions is still non-scalable w.r.t. both communication resources and computational power, an issue that will be addressed in the following.

Due to the structure of the augmented regressors $\widetilde{\mathbf{U}}_{k,i}$ defined in (7), the involved correlation quantities are written as

$$R_{\widetilde{U}_{k,i}} = \begin{bmatrix} R_{U_{k_g,i}} & 0_{M_g\times M_a} & R_{U_{k_g,i},U_{k_l,i}} & 0_{M_g\times M_b} \\ 0_{M_a\times M_g} & 0_{M_a\times M_a} & 0_{M_a\times M_l} & 0_{M_a\times M_b} \\ R^H_{U_{k_g,i},U_{k_l,i}} & 0_{M_l\times M_a} & R_{U_{k_l,i}} & 0_{M_l\times M_b} \\ 0_{M_b\times M_g} & 0_{M_b\times M_a} & 0_{M_b\times M_l} & 0_{M_b\times M_b} \end{bmatrix} \tag{12}$$

and

$$r_{\widetilde{U}_{k,i}d_{k,i}} = \begin{bmatrix} r^H_{U_{k_g,i}d_{k,i}} & 0^H_{M_a\times 1} & r^H_{U_{k_l,i}d_{k,i}} & 0^H_{M_b\times 1} \end{bmatrix}^H. \tag{13}$$

From (11), (12) and (13), we can easily see that, only two sub-vectors of $\tilde{\psi}_k^{(i)}$ are updated at each time instant i, when a specific node k performs the update step of (10). In particular, according to (5) and (10), only the sub-vectors associated with the local estimates of w and ξ_k at node k and time instant i, denoted as $\psi_k^{(i)}$ and $\xi_k^{(i)}$, respectively, are updated based on a linear combination of $\psi_{k-1}^{(i)}$ and $\xi_k^{(i-1)}$. Therefore, without any loss of optimality w.r.t. the incremental algorithm, the previous fact allows to properly modify (10) and (11) to obtain the following incremental-based NSPE algorithm,

$$\begin{cases} \psi_0^{(i)} \leftarrow \psi_N^{(i-1)} \\ \begin{bmatrix}\psi_k^{(i)} \\ \xi_k^{(i)}\end{bmatrix} = \begin{bmatrix}\psi_{k-1}^{(i)} \\ \xi_k^{(i-1)}\end{bmatrix} + \mu_k \left[r_{U_{k,i}d_{k,i}} - R_{U_{k,i}}\begin{bmatrix}\psi_{k-1}^{(i)} \\ \xi_k^{(i-1)}\end{bmatrix}\right] \end{cases} \tag{14}$$

for $k \in \{1,2,\ldots,N\}$ and some initializations $\psi_N^{(0)}$ and $\{\xi_k^{(0)}\}_{k=1}^N$.

Unlike the incremental algorithm of (10) and (11), the above NSPE algorithm is scalable in terms of computational burden and energy resources. On the one hand, regarding the computational complexity, at each time instant, each node only needs to update two vectors whose dimensions are independent of the number of nodes. In particular, according to (14) these two vectors are $\psi_k^{(i)}$ and $\xi_k^{(i)}$ of dimensions M_g and M_l, respectively. On the other hand, decreasing the power consumption, at each time instant i the NSPE strategy requires a cyclic mode of cooperation where all nodes are involved, and where each node transmits its local estimate of w^o to its immediate neighbor (see Fig. 2).

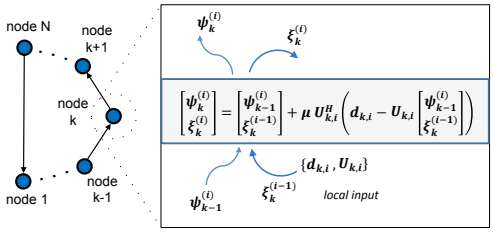

Fig. 2. Structure of the distributed I-NSPE.

To proceed further, let us derive a suitable adaptive mechanism that will enable the network to respond to time-variations in the underlying signal statistics. To do so, several approaches may be followed. Among them, in this work we adopt the instantaneous approximations $r_{U_{k,i}d_{k,i}} \approx U_{k,i}^H d_{k,i}$ and $R_{U_{k,i}} \approx U_{k,i}^H U_{k,i}$ in (14). This leads to a distributed incremental-based LMS type algorithm depicted in Fig. 2 and summarized below

Incremental-Based NSPE (I-NSPE)

- Start with some initial guess $\psi_N^{(0)}$ and $\{\xi_k^{(0)}\}_{k=1}^N$.
- At each time i, for each $k \in \{1,2,\ldots,N\}$ execute

$$\begin{aligned} &\psi_0^{(i)} \leftarrow \psi_N^{(i-1)} \\ &\begin{bmatrix}\psi_k^{(i)} \\ \xi_k^{(i)}\end{bmatrix} = \begin{bmatrix}\psi_{k-1}^{(i)} \\ \xi_k^{(i-1)}\end{bmatrix} + \mu_k\, U_{k,i}^H \left[d_{k,i} - U_{k,i}\begin{bmatrix}\psi_{k-1}^{(i)} \\ \xi_k^{(i-1)}\end{bmatrix}\right] \end{aligned} \tag{15}$$

A careful inspection of (15) reveals that the I-NSPE algorithm is equivalent to the solution of (4) by means of the so-called Hierarchical Least Squares Identification Principle (HLSI) [15]. In particular, similar to the methods derived under the aforementioned principle, the new algorithm couples, in a fully distributed fashion, a set of $N+1$ optimization processes. One of them consists in the estimation of w^o by means of a global LMS algorithm that is implemented at all nodes under an incremental mode of cooperation. On the contrary, each one of the N remaining processes is solved by an LMS algorithm that is locally performed at each node in order to estimate its parameters of local interest.

From the theory related to stochastic gradient optimization [16], it can be shown that the convergence of the algorithm provided in (15) is ensured when, for all $k \in \{1,\ldots,N\}$, $0 < \mu_k < 2/\lambda_{\max}$, with $\lambda_{\max}$ denoting the largest eigenvalue of $\sum_{k=1}^N R_{\widetilde{U}_{k,i}}$. Additionally, in a scenario where the local statistics of $\mathbf{d}_{k,i}$ and $\mathbf{U}_{k,i}$ are stationary at all nodes, if all the matrices $R_{U_{k_l,i}}$ are positive definite and at least one of the autocorrelation matrices $R_{U_{k_g,i}}$ is positive definite, it can be proved that the I-NSPE algorithm asymptotically converges to the unique centralized solution of (6) as $\mu_k \to 0$. Consequently, the proposed strategy leverages cooperation to alleviate the absence of local observability of the global parameters, w^o, at a specific subset of nodes.

4. SIMULATIONS

To illustrate the performance of the proposed algorithm, we provide numerical results in the context of cooperative spectrum sensing in cognitive radio networks based on [19] and [20, Section 2.4]. In brief, there are Q primary users (PU) transmitting and N secondary users (SU) sensing the power spectrum. In addition to PUs, for each SU we also assume a different low-power interference source. The aim for each SU is to estimate aggregated spectrum transmitted by all the PUs as well as the spectrum of its own local interferer (LI). In this setting, we use $\Phi_q^t(f)$ to denote the power spectral density (PSD) of the signal transmitted by the q-th PU. $\Phi_q^t(f)$ can be represented as

$$\Phi_q^t(f) = \sum_{j=1}^{J} b_j(f) w_{qj}^o = b_0^T(f) w_q^o \tag{16}$$

where $b_0(f) = [b_1(f), \ldots, b_J(f)]^T \in \mathbb{R}^J$ is a vector of basis functions evaluated at frequency f and $w_q^o = [w_{q1}^o, \ldots, w_{qJ}^o]^T \in \mathbb{R}^J$ is a vector of weighting coefficients representing the power transmitted by the q-th PU over each basis. Thus, if $p_{qk,i}$ denotes the path loss coefficients between the q-th PU and the k-th SU and $p_{Ik,i}$ refers the path loss coefficients between the local interferer and k-th SU for each instant i, then the signal received by the k-th SU at time i can be expressed as

$$\Phi_{k,i}^r(f) = b_{k,i}^T(f) \tilde{w}_k^o + z_{k,i} \tag{17}$$

where $\tilde{w}_k^o = [w_1^o, \ldots, w_Q^o, \xi_k^o]^T \in \mathbb{R}^{(Q+1)J}$, $z_{k,i}$ is the measurement and/or model noise and $b_{k,i}(f) = p_{k,i} \otimes b_0(f) \in \mathbb{R}^{(Q+1)J}$ with $\otimes$ standing for the Kronecker product, ξ_k^o equal to the vector of weighting coefficients representing the power transmitted by the LI associated with the k-th SU and $p_{k,i} = [p_{1k,i}, \ldots, p_{qk,i}, p_{Ik,i}]^T$.

Considering that, at discrete time i, each node k observes the received PSD in (17) over L frequency samples $\{f_m\}_{m=1}^L$, the subsequent vector lineal model is obtained

$$\mathbf{d}_{k,i} = \mathbf{U}_{k,i} \tilde{w}_k^o + \mathbf{v}_{k,i} \tag{18}$$

where $\mathbf{v}_{k,i}$ denotes noise with zero mean and covariance matrix $R_{v_{k,i}}$ of dimension $L \times L$ and where $\mathbf{U}_{k,i} = [b_{k,i}(f_1) \ldots b_{k,i}(f_m)]^T$ is of dimension $L \times (Q+1)J$ with $L > (Q+1)J$. Note that the temporal index i in the regressors $\mathbf{U}_{k,i}$ allows to account for possible variations in the channel conditions over time.

For the computer simulations presented here, we compare the I-NSPE scheme with an LMS-based non-cooperative strategy. It should be emphasized at this point that we do not compare the I-NSPE algorithm with the incremental strategy designed in [6] since the latter was developed for a scenario where $w_k^o = w^o$ for all $k \in \{1, 2, \ldots, N\}$. Thus, that comparison would not be fair since for [6] the aforementioned strategy would experience the term $\mathbf{U}_{k_l,i}\, \xi_k^o$ as additional noise at each node. Furthermore, since the I-NSPE scheme undertakes N updates of the estimate of w^o and one update of the estimate of each ξ_k^o per time step, we assume that

$$\mu_{\xi_k}^{I-NSPE} = \mu_{\xi_k}^{nc} = \mu_w^{nc} = N \mu_w^{I-NSPE}$$

where μ_a^{I-NSPE} and μ_a^{nc} stands for the step-size used by the I-NSPE and the non-cooperative schemes to estimate the vector a, respectively. This way, we get a fair comparison between both strategies.

Figure 3 depicts the learning behavior of the two schemes in terms of the network mean-square deviation (MSD) associated with

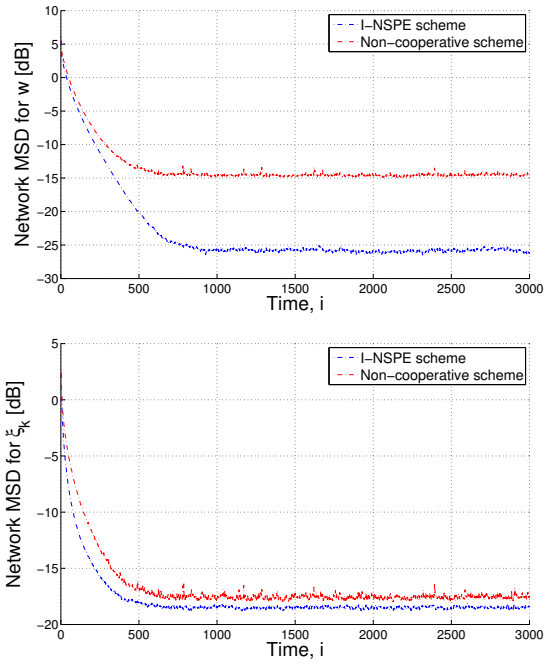

Fig. 3. Learning behavior of network MSD.

the estimation of w and ξ_k. Each network MSD is the result of averaging the local MSDs associated with the estimation of w and ξ_k at each node. To generate each plot, we have averaged the results over 50 independent experiments where we assumed $Q = 2$ PUs, $N = 10$ SUs and $J = 8$ Gaussian basis functions. Furthermore, we have considered that each SU scans $L = 40$ channels between 30 MHz and 45 MHz and that each path-loss coefficient follows $p_{tk,i}(d_{tk,i}) = (d_{tk,i}/d_0)^{-2} + n_{tk}$, where n_{tk} denotes model noise, d_0 is a reference distance and $d_{tk,i}$ accounts for the distance between the t-th transmitter and the k-th SU at time i. Under this setting and for SNRs varying from 7 dB to 15 dB across the network, due to the cooperation between the nodes, we observe that the proposed scheme outperforms the non-cooperative one, especially when estimating w^o. Although there is no exchange of estimates of ξ_k^o throughout the network, the I-NSPE scheme has enhanced rate of convergence in comparison with the non-cooperative strategy. This is a consequence of the coupling between the two estimation tasks undertaken by I-NSPE.

5. CONCLUSION

We have addressed a novel NSPE problem where the estimation interests of the nodes consists of a set of local parameters and a set of network global parameters. To do so, we have proposed a distributed adaptive scheme where a local LMS is run at each node in order to estimate each set of local parameters. Coupled with all these local estimation processes, the parameters of global interest are estimated by a global LMS implemented at all nodes under an incremental mode of cooperation. After showing that the proposed scheme converges to the MMSE solution of the centralized problem, computer simulations have been provided to illustrate the performance in the context of cooperative spectrum sensing in cognitive radio networks.

6. REFERENCES

[1] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1520–1533, 2004.

[2] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computation: numerical methods*, Athena Scientific, Singapore, 1997.

[3] I. D. Schizas, G. Mateos, and G. B. Giannakis, "Distributed lms for consensus-based in-network adaptive processing," *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2365–2382, 2009.

[4] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1847–1864, 2010.

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

[6] C. G. Lopes and A. H. Sayed, "Incremental adaptive strategies over distributed networks," *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 4064–4077, 2007.

[7] 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 Transactions on Signal Processing*, vol. 58, no. 1, pp. 151–164, 2010.

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

[9] S. Chouvardas, K. Slavakis, and S. Theodoridis, "Adaptive robust distributed learning in diffusion sensor networks," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4692–4707, 2011.

[10] S. Tu and A. Sayed, "Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks," *Submitted to IEEE Transactions on Signal Processing*, 2012 [Online]. Available: http://arxiv.org/abs/1205.3993.

[11] A. Bertrand and M. Moonen, "Distributed adaptive node-specific signal estimation in fully connected sensor networks - part I: Sequential node updating," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5277–5291, 2010.

[12] A. Bertrand and M. Moonen, "Distributed adaptive node-specific signal estimation in fully connected sensor networks - part II: Simultaneous and asynchronous node updating," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5292–5306, 2010.

[13] V. Kekatos and G. B. Giannakis, "Distributed robust power system state estimation," *Submitted to IEEE Transactions on Power Systems*, 2012 [Online]. Available: http://arxiv.org/abs/1204.0991.

[14] J. Chen and A. H. Sayed, "Distributed Pareto-optimal solutions via diffusion adaptation," in *IEEE Statistical Signal Processing Workshop, 2012. SSP 2012.*, 2012, pp. 648–651.

[15] F. Ding and T. Chen, "Hierarchical least squares identification methods for multivariable systems," *IEEE Transactions on Automatic Control*, vol. 50, no. 3, pp. 397–402, 2005.

[16] A. H. Sayed, *Adaptive filters*, Wiley-IEEE Press, 2011.

[17] D. P. Bertsekas, "A new class of incremental gradient methods for least squares problems," *SIAM Journal on Optimization*, vol. 7, no. 4, pp. 913–926, 1997.

[18] M. G. Rabbat and R. D. Nowak, "Quantized incremental algorithms for distributed optimization," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 798–808, 2005.

[19] P. Di Lorenzo, S. Barbarossa, and A. H. Sayed, "Bio-inspired swarming for dynamic radio access based on diffusion adaptation," in *IEEE 19th European Signal Conference, 2011. EUSIPCO 2011*, 2012, pp. 1–6.

[20] A. H. Sayed, "Diffusion adaptation over networks," *To appear in E-Reference Signal Processing, R. Chellapa and S. Theodoridis, Eds., Elsevier, 2013*, 2012 [Online]. Available: http://arxiv.org/abs/1205.4220.