This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2017.2771731, IEEE Transactions on Signal Processing

Node-Specific Diffusion LMS-Based Distributed Detection Over Adaptive Networks

Sara Al-Sayed, *Member, IEEE,* Jorge Plata-Chaves, *Member, IEEE,* Michael Muma, *Member, IEEE,* Marc Moonen, *Fellow, IEEE* and Abdelhak M. Zoubir, *Fellow, IEEE*

***Abstract*—Diffusion adaptation techniques have shown great promise in addressing the problem of node-specific distributed estimation where the nodes in the network are interested in different, possibly overlapping, sets of parameters. In this work, node-specific distributed detection, which has remained largely unexamined, is considered. In particular, the problem is formulated as one of binary hypothesis testing at each node for each of its parameters of interest. A distributed, online solution for this problem is sought based on diffusion adaptation techniques. In this setting, a signal to be detected by one node constitutes interference that may compromise the ability of the other nodes to detect their signals of interest reliably. Under mild assumptions on the data and network, it is shown that, for sufficiently small adaptation step-sizes, interference can be kept in check. Local detectors are developed where the test-statistics and thresholds adapt to changing conditions in real time. The distributed algorithm is analyzed; and its detection performance characterized and illustrated through numerical simulations.**

***Index Terms*—Distributed detection, diffusion adaptation, adaptive networks, collaborative processing, multi-task learning, mean-square performance analysis.**

I. Introduction

HETEROGENEITY is a staple of modern-day networks, where geographically dispersed nodes have different tasks; computational and communication resources; modes of operation; degrees of readiness for cooperation; privacy concerns; and environmental conditions under which data transmission takes place. For example, in localization and tracking applications, different nodes are interested in localizing and tracking different moving targets; in spectrum sensing, different cognitive radios are interested in detecting and tracking different spectrum holes—more examples can be found in [1], [2].

Typically, these networks operate in harsh wireless environments with scarce computational and communication resources and fluctuating statistical conditions. In order to address these challenges, the nodes engage in collaborative adaptation and learning, where each node exchanges information with its neighbors and updates its estimates online in response to streaming data [3], [4]. Two popular modes of cooperation for parameter estimation are investigated in the literature: consensus strategies [5]–[8] and diffusion strategies [4], [9]–[12]. It was shown in [13] that diffusion strategies enjoy enhanced stability and performance relative to consensus strategies, achieving centralized performance in the slow adaptation regime [4], [10], [14], [15]. Building upon parameter estimation is detection, investigated in both the consensus [16]–[20] and diffusion [21]–[25] contexts.

In distributed node-specific adaptation and learning, the nodes aim to learn different, possibly overlapping, sets of parameters characterizing different phenomena of interest. This problem, sometimes referred to as multi-tasking, is the focus of several recent works [26]–[36]. While the works [31]–[36] address the problem of node-specific distributed estimation in a diffusion setting, the problem of node-specific distributed detection has remained largely unexamined. Hence, we would like to fill this gap by extending the framework of [31] to solve a detection task that is cast as a binary hypothesis testing problem at each node for each of its parameters of interest. That is, we seek a distributed, online strategy such that each node can detect the presence or absence of its parameters of interest. We choose the framework of [31] because, unlike its counterparts in [33], [35], [36], it can be flexibly generalized to accommodate individual, local, and global parameters of interest with potential overlap; and, under fairly general assumptions, asymptotic unbiasedness can be ensured. The developed distributed detection algorithm is further analyzed based on the energy-conservation framework [3], [4], [12] and its performance illustrated through numerical simulations.

The article is organized as follows. In Sec. II, the data model is introduced and the problem is formulated. In Sec. III, we review the node-specific diffusion least-mean-squares (LMS)–based distributed estimation algorithm and scrutinize the asymptotic behavior of the weight estimates. In Sec. IV, the basic form of the node-specific diffusion LMS-based distributed detection algorithm as motivated by asymptotic analysis is described. In Sec. V, we conduct estimation and detection performance analyses, motivating fully distributed online detectors. Simulation results are presented in Sec. VI.

Notation: Lowercase letters are reserved for scalars and vectors, uppercase for matrices; boldface font is reserved for random variables, and normal font for deterministic variables.

This work was done in the frame of the project HANDiCAMS, which acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number: 323944. The work of S. Al-Sayed and M. Muma was also supported by the LOEWE initiative (Hessen, Germany) within the NICER project and by the German Research Foundation (DFG).

S. Al-Sayed is with the Bioinspired Communication Systems Group; M. Muma and A. M. Zoubir are with the Signal Processing Group. Both groups are at Institut für Nachrichtentechnik, Technische Universität Darmstadt, 64283 Darmstadt, Germany (e-mails: sara.al-sayed@bcs.tu-darmstadt.de, {mmuma,zoubir}@spg.tu-darmstadt.de).

J. Plata-Chaves and M. Moonen are with the Department of Electrical Engineering (ESAT-STADIUS), KU Leuven, B-3001 Leuven, Belgium (e-mails: {jorge.plata-chaves,marc.moonen}@esat.kuleuven.be).

1053-587X (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2017.2771731, IEEE Transactions on Signal Processing

The time index appears in parenthesis for scalars, and in the subscript for vectors and matrices. All vectors are column vectors unless mentioned otherwise. Transposition, inversion, and the trace and gradient operators are denoted by $(\cdot)^\top$, $(\cdot)^{-1}$, $\mathrm{Tr}(\cdot)$, and ∇_x, respectively; and the Euclidean norm is denoted by $\|\cdot\|$. Expectation is denoted by $\mathbb{E}$. Set cardinality is denoted by $|\cdot|$. The notation $\mathbb{1}$ and I represents the all-one vector and identity matrix of appropriate sizes, respectively (if not clear from the context, the size is indicated in the subscript). The operator $\mathrm{col}\{\cdot\}$ stacks its arguments vertically; and the operator $\mathrm{diag}\{\cdot\}$ forms a diagonal matrix from its arguments.

II. Data Model and Problem Formulation

A. Data Model

We reappropriate the model in [31] for the problem formulation in this work. We consider a network of N nodes whose topology is described by a graph. The nodes are indexed by $k \in \mathcal{V} \triangleq \{1, \ldots, N\}$. The neighborhood of node k, denoted as $\mathcal{V}_k$, is the set of all nodes, including k, that node k can exchange information with. Information exchange among neighbors is governed by a combination policy, which will be elaborated upon later. There are N_s systems, models, or signals to be detected, indexed by $c \in \mathcal{S} \triangleq \{1, \ldots, N_s\}$, each parametrized by a *known*, real-valued weight vector w_{s_c} of size $M_c \times 1$. Different nodes in the network are interested in detecting different sets of weight vectors, so the network can be visualized as involving clusters of connected nodes, $\mathcal{C}_c \subseteq \mathcal{V}$, $c \in \mathcal{S}$, that are generally not disjoint, such that nodes $k \in \mathcal{C}_c$ are interested in the cth weight vector w_{s_c}. The set of interests of each node k is nonempty and denoted by $\mathcal{I}_k \triangleq \{c | k \in \mathcal{C}_c, c \in \mathcal{S}\}$. It is assumed that each node k knows all nodes $\ell \in \mathcal{V}_k \cap \mathcal{C}_c$ that share its interest in the cth weight vector, which is known, for all $c \in \mathcal{I}_k$. Unlike the formulation in [31], the specification of global, common, and local interests is implicit in the formulation here. The sets $\mathcal{V}$, $\mathcal{V}_k$, $\mathcal{C}_c$, $\mathcal{S}$, and $\mathcal{I}_k$, as well as their subsets, are assumed to be naturally ordered. At each time index $i \geq 0$, every node k has access to real-valued scalar noisy measurements of the form:

$$\begin{aligned} \boldsymbol{d}_k(i) &= \sum_{c \in \mathcal{I}_k} u_{k,c,i} w^o_{s_c} + \boldsymbol{v}_k(i) \\ &= \sum_{c=1}^{N_s} u_{k,c,i} w^o_{s_c} + \boldsymbol{v}_k(i) \end{aligned} \tag{1}$$

in terms of the acting or true value of the weight vectors, $\{w^o_{s_c}\}$, and where

$$u_{k,c,i} \begin{cases} \neq 0, & \text{if} \quad k \in \mathcal{C}_c \\ = 0, & \text{if} \quad k \notin \mathcal{C}_c \end{cases}. \tag{2}$$

The row vectors $\{u_{k,c,i} | c \in \mathcal{I}_k\}$ are real-valued regressors of sizes M_c that are known to node k. It is assumed that the energy of the regressors is bounded from above, i.e., there exists a real-valued positive scalar α such that $\sum_{c \in \mathcal{I}_k} \|u_{k,c,i}\|^2 \leq \alpha$ for all k and i. The noise $\boldsymbol{v}_k(i)$ is a scalar wide-sense stationary zero-mean real Gaussian noise process, spatially independent and temporally white with finite variance $\sigma^2_{v,k} = \mathbb{E}\, \boldsymbol{v}^2_k(i)$.

B. Binary Hypothesis Tests

Each node k in the network aims to establish the absence or presence of each of its weight vectors of interest. This problem can be cast as that of binary hypothesis testing at each node for each of those weight vectors. That is, for every $c \in \mathcal{I}_k$:

$$w^o_{s_c} = \begin{cases} 0 & \text{under} \quad \mathcal{H}^c_0 \\ w_{s_c} \neq 0 & \text{under} \quad \mathcal{H}^c_1 \end{cases} \tag{3}$$

where $\mathcal{H}^c_0$ and $\mathcal{H}^c_1$ denote the null and alternative hypotheses, respectively.

One approach to solving this detection problem is for each node k, given its data $\{\boldsymbol{d}_k(i), \{\boldsymbol{u}_{k,c,i}\}\}$ up to time index i, to calculate batch estimates for $w^o_{s_c}$, $c \in \mathcal{I}_k$, say the least-squares estimates. Local batch processing is riddled with challenges, however, e.g., memory constraints, rank deficiency, computational constraints, missing statistical knowledge, and time-varying scenarios. The approach in this work is to employ diffusion adaptation techniques.

A compact representation of model (1) will come to use later on. Let

$$w^o_s \triangleq \mathrm{col}\Big\{ \big\{ w^o_{s_c} \big\}_{c \in \mathcal{S}} \Big\} \quad (M \times 1) \tag{4}$$

where $M = \sum_{c=1}^{N_s} M_c$, and for each node k,

$$u^\top_{k,i} \triangleq \mathrm{col}\Big\{ \big\{ u^\top_{k,c,i} \big\}_{c \in \mathcal{S}} \Big\} \quad (M \times 1) \tag{5}$$

So the data model can be written more compactly as

$$\boldsymbol{d}_k(i) = u_{k,i} w^o_s + \boldsymbol{v}_k(i), \quad i \geq 0. \tag{6}$$

Arranging the data from all nodes $k = 1, \ldots, N$ into vectors and matrices:

$$\begin{aligned} \boldsymbol{d}_i &\triangleq \mathrm{col}\{\boldsymbol{d}_1(i), \boldsymbol{d}_2(i), \ldots, \boldsymbol{d}_N(i)\} && (N \times 1) \\ U_i &\triangleq \mathrm{col}\{u_{1,i}, u_{2,i}, \ldots, u_{N,i}\} && (N \times M) \\ \boldsymbol{v}_i &\triangleq \mathrm{col}\{\boldsymbol{v}_1(i), \boldsymbol{v}_2(i), \ldots, \boldsymbol{v}_N(i)\} && (N \times 1) \\ R_v &\triangleq \mathrm{diag}\big\{\sigma^2_{v,1}, \sigma^2_{v,2}, \ldots, \sigma^2_{v,N}\big\} && (N \times N) \end{aligned}$$

then time indeces $j = i, i-1, \ldots, 1, 0$:

$$\begin{aligned} \boldsymbol{d}_{0:i} &\triangleq \mathrm{col}\{\boldsymbol{d}_i, \boldsymbol{d}_{i-1}, \ldots, \boldsymbol{d}_0\} && ((i+1)N \times 1) \\ U_{0:i} &\triangleq \mathrm{col}\{U_i, U_{i-1}, \ldots, U_0\} && ((i+1)N \times M) \\ \boldsymbol{v}_{0:i} &\triangleq \mathrm{col}\{\boldsymbol{v}_i, \boldsymbol{v}_{i-1}, \ldots, \boldsymbol{v}_0\} && ((i+1)N \times 1) \\ R_{v,0:i} &\triangleq \mathrm{diag}\{R_v, \ldots, R_v\} && ((i+1)N \times (i+1)N) \end{aligned}$$

leads to the following network data model:

$$\boldsymbol{d}_{0:i} = U_{0:i} w^o_s + \boldsymbol{v}_{0:i}. \tag{7}$$

Additionally, we make the assumption that $U_{0:i}$, as well as U_i, are full-rank for every i, with $M \leq N$. This can be guaranteed by sufficient spatial and temporal diversity among the regressors $\{u_{k,c,i}\}$ without stringent individual linear independence requirements.

1053-587X (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2017.2771731, IEEE Transactions on Signal Processing

III. Node-Specific Diffusion LMS-Based Distributed Estimation

In [31], a distributed stochastic-gradient type algorithm was developed to minimize the following global mean-square-error (MSE) cost:

$$J\Big(\Big\{\{w_{k,c}\}_{c\in\mathcal{I}_k}\Big\}_{k\in\mathcal{V}}\Big) \triangleq \sum_{k\in\mathcal{V}}\mathbb{E}\left|\boldsymbol{d}_k(i)-\sum_{c\in\mathcal{I}_k}\boldsymbol{u}_{k,c,i}w_{k,c}\right|^2. \quad (8)$$

The data appearing in (8), $\{\boldsymbol{d}_k(i),\boldsymbol{u}_{k,i}\}$, follow the data model (1). The regressors $u_{k,c,i}$ for each node k and $c\in\mathcal{I}_k$ are generally random in nature, having been generated from an underlying probability distribution; nevertheless, each node k observes the realizations of its own regressors. For each node k and $c\in\mathcal{I}_k$, the regressors $\{\boldsymbol{u}_{k,c,i}\}$ are zero-mean, spatially independent, temporally uncorrelated, and satisfy a global observability condition: $\sum_{k=1}^{N}R_{u,k}>0$, where $R_{u,k}=\mathbb{E}\,\boldsymbol{u}_{k,i}^{\top}\boldsymbol{u}_{k,i}\geq 0$ is the regressor covariance matrix. The noise and regressors are assumed to be statistically independent. The adapt-then-combine (ATC) version of the node-specific diffusion LMS-based distributed estimation algorithm that attempts to solve (8) is reviewed briefly here. At each time index i, each node k has an estimate $w_{k,c,i-1}$ of the weight vector $w_{s_c}^o$ for all $c\in\mathcal{I}_k$. Each node k updates its current estimates $\{w_{k,c,i-1}\}$ in an LMS fashion with step-size μ_k using its data $\{d_k(i),\{u_{k,c,i}\}\}$ to form intermediate estimates $\{\psi_{k,c,i}\}$. We introduce the $N\times N$ combination matrix $A_c\triangleq[a_{c,\ell k}]$, $\ell,k\in\mathcal{V}$, for fusing the intermediate estimates pertaining to the weight vector w_{s_c}, $c\in\mathcal{S}$, within neighborhoods, forming final estimates $\{w_{k,c,i}\}$. That is, node k scales the intermediate estimate of node ℓ for the weight vector w_{s_c} at each time index i with $a_{c,\ell k}$ for fusion. The entries of each A_c satisfy:

$$\begin{aligned} &a_{c,\ell k}=0,\ \text{if } k\notin\mathcal{C}_c \text{ or } \ell\notin\mathcal{V}_k\cap\mathcal{C}_c; \quad (9)\\ &\text{if } k\in\mathcal{C}_c,\ \sum_{\ell\in\mathcal{V}_k\cap\mathcal{C}_c}a_{c,\ell k}=1 \end{aligned}$$

The algorithm is listed in Table I, using the following definitions:

$$u_{\mathcal{I}_k,i}^{\top}\triangleq\operatorname{col}\Big\{\{u_{k,c,i}^{\top}\}_{c\in\mathcal{I}_k}\Big\} \quad (10)$$

$$w_{\mathcal{I}_k,i}\triangleq\operatorname{col}\Big\{\{w_{k,c,i}\}_{c\in\mathcal{I}_k}\Big\} \quad (11)$$

$$\psi_{\mathcal{I}_k,i}\triangleq\operatorname{col}\Big\{\{\psi_{k,c,i}\}_{c\in\mathcal{I}_k}\Big\} \quad (12)$$

of size $\sum_{c\in\mathcal{I}_k}M_c$ each.

In this work, the regressors $\{u_{k,c,i}\}$ are taken to be deterministic and known. In practice, the regressors may have been estimated, yet it is assumed that the estimation error is negligible—see Sec. VI-B for a practical example. The network nodes run the algorithm in Table I so that at each time index i, each node k has an estimate $w_{k,c,i}$ of the acting or true weight vector $w_{s_c}^o$ for all $c\in\mathcal{I}_k$. In order to design hypothesis tests based on the estimates $\{w_{k,c,i}\}$, it is necessary to derive a relationship to the network data model (7). To this end, it is instructive to rewrite the algorithm in augmented form. In

TABLE I
Diffusion Estimation Algorithm

Initializations: $\mathcal{V}_k$, $\mathcal{I}_k$, μ_k for every node k, $a_{c,\ell k}$ ($\ell,k\in\mathcal{C}_c$, cluster c, $c\in\mathcal{S}$). Start with $w_{k,c,-1}=0$ for every node k and each cluster c, $c\in\mathcal{I}_k$. For every time index $i\geq 0$, repeat

Adaptation step: for every node k, repeat

$$\psi_{\mathcal{I}_k,i}=w_{\mathcal{I}_k,i-1}+\mu_k u_{\mathcal{I}_k,i}^{\top}\big[d_k(i)-u_{\mathcal{I}_k,i}w_{\mathcal{I}_k,i-1}\big] \quad (13)$$

($u_{\mathcal{I}_k,i}$, $w_{\mathcal{I}_k,i}$, and $\psi_{\mathcal{I}_k,i}$ are defined in (10)–(12).)

Combination step: for every node k and each cluster c, $c\in\mathcal{I}_k$, repeat

$$w_{k,c,i}=\sum_{\ell\in\mathcal{V}_k\cap\mathcal{C}_c}a_{c,\ell k}\psi_{\ell,c,i} \quad (14)$$

analogy to the augmented regressors defined according to (5) and (2), let

$$w_{k,i}\triangleq\operatorname{col}\big\{\{w_{k,c,i}\}_{c\in\mathcal{S}}\big\} \quad (M\times 1) \quad (15)$$

$$\psi_{k,i}\triangleq\operatorname{col}\big\{\{\psi_{k,c,i}\}_{c\in\mathcal{S}}\big\} \quad (M\times 1) \quad (16)$$

Moreover, let

$$A_{\ell k}\triangleq\operatorname{diag}\big\{\{a_{c,\ell k}I_{M_c}\}_{c\in\mathcal{S}}\big\} \quad (M\times M) \quad (17)$$

Then, the adaptation and combination steps of the ATC algorithm, (13) and (14), can be rewritten in augmented form as

$$\left\{\begin{array}{rcl}\psi_{k,i} &=& w_{k,i-1}+\mu_k u_{k,i}^{\top}[d_k(i)-u_{k,i}w_{k,i-1}]\\ w_{k,i} &=& \sum\limits_{\ell\in\mathcal{V}_k}A_{\ell k}\psi_{\ell,i}\end{array}\right. \quad (18)$$

Let $Y_{k,i}\triangleq I-\mu_k u_{k,i}^{\top}u_{k,i}$ ($M\times M$) and $E_k\triangleq\operatorname{diag}\{e_k\}$, where e_k is the all-zero vector of length N and kth entry equal to 1. By induction, it can be verified that the following relationship holds between the estimator $\boldsymbol{w}_{k,i}$ for all nodes k and the global measurement vector $\boldsymbol{d}_{0:i}$:

$$\boldsymbol{w}_{k,i}=K_{k,i}\boldsymbol{d}_{0:i} \quad (19)$$

where

$$K_{k,i}=\left[\sum_{\ell\in\mathcal{V}_k}\mu_\ell A_{\ell k}U_i^{\top}E_\ell \quad \sum_{\ell\in\mathcal{V}_k}A_{\ell k}Y_{\ell,i}K_{\ell,i-1}\right] \quad (20)$$

of size $M\times(i+1)N$. The proof follows closely that of [21, Proposition 2]. The cth subvector of node k's estimator $\boldsymbol{w}_{k,i}$, $\boldsymbol{w}_{k,c,i}$, for every $c\in\mathcal{S}$, for all $k\in\mathcal{V}$, can be recovered from (19) using matrices P_c of the form:

$$\begin{aligned}P_c\triangleq\big[&0_{M_c\times M_1},\ldots,0_{M_c\times M_{c-1}},I_{M_c\times M_c},\\ &0_{M_c\times M_{c+1}},\ldots,0_{M_c\times M_{N_s}}\big] \quad (M_c\times M)\end{aligned} \quad (21)$$

where $0_{N_1\times N_2}$ is the all-zero matrix of size $N_1\times N_2$, such that

$$\begin{aligned}\boldsymbol{w}_{k,c,i}&=P_c\boldsymbol{w}_{k,i}\\ &=P_cK_{k,i}U_{0:i}w_s^o+P_cK_{k,i}\boldsymbol{v}_{0:i}\\ &=P_cK_{k,i}U_{c,0:i}w_{s_c}^o+P_cK_{k,i}\sum_{c'\in\mathcal{S}\setminus\{c\}}U_{c',0:i}w_{s_c'}^o\\ &\quad+P_cK_{k,i}\boldsymbol{v}_{0:i}\end{aligned} \quad (22)$$

1053-587X (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2017.2771731, IEEE Transactions on Signal Processing

where

$$U_{c,0:i} = U_{0:i} P_c^\top \quad ((i+1)N \times M_c) \tag{23}$$

The second term in the last equation in (22) constitutes interference to node k's estimate of the cth weight vector w_{s_c} arising from all systems, models, or signals $c' \in \mathcal{S} \setminus \{c\}$. A detector for each weight vector w_{s_c} would need to suppress the respective interference and pre-whiten the noise. But to suppress the interference, the active hypotheses need to be known. Here, we propose a different approach, motivated by the fact that the interference term vanishes at steady-state under certain conditions. This is established in the following lemma.

Lemma 1 (Vanishing Interference). Under the data and network model in Sec. II, if each node k in the network runs the algorithm in Table I using combination policies A_c, $c \in \mathcal{S}$, with entries satisfying (9), then, for sufficiently small step-sizes $\{\mu_k\}$, as $i \to \infty$,

$$K_{k,i} U_{0:i} \to W_k \tag{24}$$

where

$$W_k \triangleq \mathrm{diag}\big\{\{\mathbb{I}_{\mathcal{I}_k}(c) I_{M_c}\}_{c\in\mathcal{S}}\big\} \quad (M \times M) \tag{25}$$

with $\mathbb{I}$ denoting the indicator function: The matrix W_k is an $N_s \times N_s$ block-diagonal matrix whose cth block, of size $M_c \times M_c$, is the identity matrix, I_{M_c}, if $c \in \mathcal{I}_k$.

Proof: Let us define the following matrices:

$$\mathcal{M} \triangleq \mathrm{diag}\big\{\{\mu_k I_M\}_{k\in\mathcal{V}}\big\} \quad (NM \times NM) \tag{26}$$

$$\mathcal{H}_i \triangleq \mathrm{diag}\Big\{\big\{u_{k,i}^\top u_{k,i}\big\}_{k\in\mathcal{V}}\Big\} \quad (NM \times NM) \tag{27}$$

$$\mathcal{D}_i \triangleq \mathrm{col}\Big\{\big\{u_{k,i}^\top u_{k,i}\big\}_{k\in\mathcal{V}}\Big\} = \mathcal{H}_i(\mathbb{1}_N \otimes I_M) \quad (NM \times M) \tag{28}$$

$$K_i \triangleq \mathrm{col}\big\{\{K_{k,i}\}_{k\in\mathcal{V}}\big\} \quad (NM \times (i+1)N) \tag{29}$$

$$\mathcal{A} \triangleq [A_{\ell k}], \quad \ell, k \in \mathcal{V} \quad (NM \times NM) \tag{30}$$

where $X \otimes Z$ is the Kronecker product of the matrices X and Z, and

$$\mathcal{Y}_i \triangleq \mathrm{diag}\big\{\{Y_{k,i}\}_{k\in\mathcal{V}}\big\} = I - \mathcal{M}\mathcal{H}_i \quad (NM \times NM) \tag{31}$$

Given the previous definitions, it is straightforward to establish from (20) that the following recursion holds:

$$(K_i U_{0:i}) = \mathcal{A}^\top \mathcal{M}\mathcal{D}_i + \mathcal{A}^\top \mathcal{Y}_i (K_{i-1} U_{0:i-1}) \tag{32}$$

with initial condition $(K_{-1}U_{0:-1}) = 0$. As $i \to \infty$, assuming convergence, expression (32) becomes

$$\big(I - \mathcal{A}^\top \mathcal{Y}_i\big)(K_i U_{0:i}) = \mathcal{A}^\top \mathcal{M}\mathcal{D}_i. \tag{33}$$

The condition for convergence of (32) coincides with that for mean stability of the augmented algorithm (18)—the latter is derived in Sec. V-B1. To see that $K_{k,i}U_{0:i} \to W_k$ as $i \to \infty$, we first need to show that

$$K_i U_{0:i} = \mathrm{col}\big\{\{W_k\}_{k\in\mathcal{V}}\big\} \triangleq \mathcal{W} \tag{34}$$

solves (33). The left-hand side of (33) with $K_i U_{0:i}$ set to $\mathcal{W}$ and using (31) evaluates to

$$\begin{aligned}\big(I - \mathcal{A}^\top \mathcal{Y}_i\big)\mathcal{W} &= \mathcal{W} - \mathcal{A}^\top \mathcal{W} + \mathcal{A}^\top \mathcal{M}\mathcal{H}_i \mathcal{W} \\ &= \mathcal{A}^\top \mathcal{M}\mathcal{H}_i \mathcal{W}\end{aligned} \tag{35}$$

since $\mathcal{A}^\top \mathcal{W} = \mathcal{W}$ by the properties of the combination policy in (9). The right-hand side of (33) can be re-expressed using (28) as

$$\mathcal{A}^\top \mathcal{M}\mathcal{D}_i = \mathcal{A}^\top \mathcal{M}\mathcal{H}_i(\mathbb{1}_N \otimes I_M) = \mathcal{A}^\top \mathcal{M}\mathcal{H}_i \mathcal{W} \tag{36}$$

given the construction of the regressors according to (5) and (2). Hence, $\mathcal{W}$ solves (33).

Next, we need to show that $\mathcal{W}$ is the unique solution for (33). Equivalently, we show in Appendix A that, under the conditions outlined in Lemma 1, the node-specific diffusion LMS-based distributed estimation algorithm in its augmented form (18) generates iterates $\{\boldsymbol{w}_{k,i}\}$ that converge to $W_k w_s^o$ in the mean-square sense for each node k. The assumption of deterministic regressors $\{u_{k,c,i}\}$ made in this work poses an analysis challenge such that a different route than the one pursued in [31] needs to be taken. ■

Motivated by the above lemma, we consider the following model derived from (22) by ignoring the interference term:

$$\boldsymbol{w}_{k,c,i} \approx P_c K_{k,i} U_{c,0:i} w_{s_c}^o + P_c K_{k,i} \boldsymbol{v}_{0:i}, \quad i \geq 0 \tag{37}$$

for each node k and every $c \in \mathcal{I}_k$.

IV. Node-Specific Diffusion LMS-Based Distributed Detection

Based on model (37), according to the Neyman–Pearson framework [37], the test that maximizes the probability of detection of the weight vector w_{s_c} subject to a given false-alarm rate upper bound is the likelihood-ratio test. To this end, the following local detectors can be set up for each $c \in \mathcal{I}_k$ and every node k at time index $i \geq 0$:

$$T_{k,c,i}(\boldsymbol{w}_{k,c,i}) \underset{\mathcal{H}_1^c}{\overset{\mathcal{H}_0^c}{\lessgtr}} \gamma_{k,c,i}. \tag{38}$$

The test-statistics are given by

$$T_{k,c,i}(\boldsymbol{w}_{k,c,i}) = w_{s_c}^\top Q_{k,c,i} \boldsymbol{w}_{k,c,i} \tag{39}$$

where, for full-rank $P_c K_{k,i}$ (guaranteed by full-rank $U_{0:i}$),

$$Q_{k,c,i} = \big(P_c K_{k,i} U_{0:i} P_c^\top\big)^\top \big(P_c K_{k,i} R_{v,0:i} K_{k,i}^\top P_c^\top\big)^{-1}. \tag{40}$$

The computation of $Q_{k,c,i}$ requires the exchange of large amounts of data across the network. Given communication and computational constraints in networks, an approximation can prove rewarding. Recalling the expression (20) for $K_{k,i}$ and assuming the matrices $\{K_{k,i}\}$ of temporally growing dimensions are not exchanged among neighboring nodes (by setting $A_{\ell k}$ to W_k, if $\ell = k$, and 0, otherwise), and that the step-sizes $\{\mu_k\}$ are sufficiently small, then, a good approximation for $Q_{k,c,i}$ is $Q_{k,c,i} = I_{M_c}$. A detailed derivation is given in Appendix B. The test-statistics are ultimately given by

$$T_{k,c,i}(\boldsymbol{w}_{k,c,i}) = w_{s_c}^\top \boldsymbol{w}_{k,c,i}, \tag{41}$$

relying only on the locally available adaptive estimate $w_{k,c,i}$. It remains to determine the thresholds $\{\gamma_{k,c,i}\}$ based on performance analysis. This is the topic of the next section.

1053-587X (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2017.2771731, IEEE Transactions on Signal Processing

V. Performance Analysis

Throughout this section, substantial use is made of the following error quantity:

$$\widetilde{\boldsymbol{w}}_{k,i} \triangleq w_k^o - \boldsymbol{w}_{k,i}, \qquad w_k^o \triangleq W_k w_s^o, \tag{42}$$

which is referred to as the weight-error vector.

A. Detection Performance

For every node $k \in \mathcal{V}$, $c \in \mathcal{I}_k$, the test-statistics at time index $i \geq 0$, given by (41), are distributed as

$$T_{k,c,i}(\boldsymbol{w}_{k,c,i}) \sim \mathcal{N}\Big(w_{s_c}^\top P_c (w_k^o - \mathbb{E}\,\widetilde{\boldsymbol{w}}_{k,i}), \sigma_{k,c,i}^2\Big) \tag{43}$$

where

$$\sigma_{k,c,i}^2 = w_{s_c}^\top P_c R_{\widetilde{w}_{k,i}} P_c^\top w_{s_c} \tag{44}$$

with $R_{\widetilde{w}_{k,i}} = \mathbb{E}[\widetilde{\boldsymbol{w}}_{k,i} - \mathbb{E}\,\widetilde{\boldsymbol{w}}_{k,i}][\widetilde{\boldsymbol{w}}_{k,i} - \mathbb{E}\,\widetilde{\boldsymbol{w}}_{k,i}]^\top$ denoting the covariance matrix of node k's weight-error vector. It is then clear that the detection performance relates to the estimation performance, which is analyzed in Sec. V-B.

The detection and false-alarm probabilities are given by

$$P_{k,c,i}^d \approx \mathrm{Q}\left(\frac{\gamma_{k,c,i} - w_{s_c}^\top w_{s_c} + w_{s_c}^\top P_c\,\mathbb{E}\,\widetilde{\boldsymbol{w}}_{k,i}}{\sigma_{k,c,i}}\right) \tag{45}$$

$$P_{k,c,i}^f \approx \mathrm{Q}\left(\frac{\gamma_{k,c,i} + w_{s_c}^\top P_c\,\mathbb{E}\,\widetilde{\boldsymbol{w}}_{k,i}}{\sigma_{k,c,i}}\right) \tag{46}$$

where $\mathrm{Q}(\cdot)$ is the right-tail Gaussian probability function:

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{t^2}{2}}\, dt.$$

The misdetection probability is given by $P_{k,c,i}^m = 1 - P_{k,c,i}^d$. A distributed method for the computation of the thresholds $\{\gamma_{k,c,i}\}$ will be presented in Sec. V-C.

B. Estimation Performance

We first introduce the network weight-error vector

$$\widetilde{\boldsymbol{w}}_i \triangleq \mathrm{col}\{\widetilde{\boldsymbol{w}}_{1,i}, \ldots, \widetilde{\boldsymbol{w}}_{N,i}\} \quad (NM \times 1) \tag{47}$$

as well as the following matrix:

$$\mathcal{U}_i = \mathrm{diag}\{u_{1,i}, \ldots, u_{N,i}\} \quad (N \times NM) \tag{48}$$

Recalling the definitions for the matrices $\mathcal{M}$, $\mathcal{H}_i$, and $\mathcal{A}$ in (26), (27), and (30), respectively, the network-wide weight-error recursion is given by[1]

$$\widetilde{\boldsymbol{w}}_i = \mathcal{A}^\top [I - \mathcal{M}\mathcal{H}_i]\widetilde{\boldsymbol{w}}_{i-1} - \mathcal{A}^\top \mathcal{M}\mathcal{U}_i^\top \boldsymbol{v}_i. \tag{49}$$

[1]Apart from the inherent structure of $\mathcal{H}_i$ and $\mathcal{U}_i$ that is brought about by the augmented regressor construction according to (5) and (2), this recursion is of the same form as its counterpart in single-task networks in [21].

1) Mean Performance: Taking the expectation of both sides of the weight-error recursion (49) leads to the following network-wide mean weight-error recursion:

$$\mathbb{E}\,\widetilde{\boldsymbol{w}}_i = \mathcal{A}^\top [I - \mathcal{M}\mathcal{H}_i]\mathbb{E}\,\widetilde{\boldsymbol{w}}_{i-1}. \tag{50}$$

A sufficient condition for asymptotic unbiasedness irrespective of the initial condition w_{-1} according to [21, Lemma 2] is that there exist a time index i_o, a submultiplicative norm $\|\cdot\|$, and a constant $0 < \theta < 1$ such that

$$\big\|\mathcal{A}^\top (I - \mathcal{M}\mathcal{H}_i)\big\| \leq \theta \tag{51}$$

for all $i \geq i_o$. As a matter of fact, sufficiently small step-sizes $\{\mu_k\}$ ensure asymptotic unbiasedness, as the following theorem shows.

Theorem 1 (**Mean Stability).** Under the data and network model in Sec. II, if each node k in the network runs the algorithm in Table I using combination policies A_c, $c \in \mathcal{S}$, with entries satisfying (9), then, for sufficiently small step-sizes $\{\mu_k\}$, $\lim_{i\to\infty} \mathbb{E}\,\widetilde{\boldsymbol{w}}_{k,i} = 0$, for each k. Equivalently, $\lim_{i\to\infty} \mathbb{E}\,\boldsymbol{w}_{k,c,i} = w_{s_c}^o$ for each $c \in \mathcal{I}_k$ and every k.

Proof: See Appendix C. ∎

2) Mean-Square Performance: The network-wide weight-error covariance recursion is given by

$$\begin{aligned}\mathcal{R}_{\widetilde{w}_i} &\triangleq \mathbb{E}[\widetilde{\boldsymbol{w}}_i - \mathbb{E}\,\widetilde{\boldsymbol{w}}_i][\widetilde{\boldsymbol{w}}_i - \mathbb{E}\,\widetilde{\boldsymbol{w}}_i]^\top \\ &= \mathcal{A}^\top [I - \mathcal{M}\mathcal{H}_i]\mathcal{R}_{\widetilde{w}_{i-1}}[I - \mathcal{H}_i\mathcal{M}]\mathcal{A} \\ &\quad + \mathcal{A}^\top \mathcal{M}\mathcal{U}_i^\top R_v \mathcal{U}_i \mathcal{M}\mathcal{A}\end{aligned} \tag{52}$$

with $\mathcal{R}_{\widetilde{w}_{-1}} = 0$, if $w_{k,-1} = 0$ for every node k. The matrix $R_{\widetilde{w}_{k,i}}$ is given by the kth $M \times M$ diagonal block of $\mathcal{R}_{\widetilde{w}_i}$. Convergence analysis of (52) is generally challenging and has not been undertaken in [21]. However, it is clear from (44), (45), and (46) that $\mathcal{R}_{\widetilde{w}_{i-1}}$ must converge to 0 for the detection probability $P_{k,c,i}^d$ to converge to 1 for a given false-alarm probability, as $i \to \infty$, for every node k and each $c \in \mathcal{I}_k$.

C. Distributed Threshold Computation

In the Neyman–Pearson framework, the thresholds $\{\gamma_{k,c,i}\}$ are calculated from the false-alarm rate constraints as

$$\gamma_{k,c,i} \approx \sigma_{k,c,i}\, \mathrm{Q}^{-1}(P_{k,c,i}^f) - w_{s_c}^\top P_c\,\mathbb{E}\,\widetilde{\boldsymbol{w}}_{k,i}. \tag{53}$$

Clearly, the calculation of the thresholds according to (53) at each time index i cannot be performed locally at node k using only the locally available regressors $\{u_{k,c,i} | c \in \mathcal{I}_k\}$, measurements $\{d_k(i)\}$, and intermediate estimates from its neighbors $\{\{\psi_{\ell,c,i}\} | \ell \in \mathcal{V}_k, c \in \mathcal{I}_k\}$ at the respective time index. More information needs to be exchanged for the calculation of the local mean weight-error vector $\mathbb{E}\,\widetilde{\boldsymbol{w}}_{k,i}$ according to (50) and test-statistic variance $\sigma_{k,c,i}^2$ according to (44) and (52). Similar to [21], an alternative distributed method is proposed to approximate the thresholds. First, assuming asymptotic unbiasedness, the thresholds can be approximated as

$$\gamma_{k,c,i} = g_c^{-\frac{1}{2}} \cdot \sigma_{k,c,i}^{\mathrm{loc}}\, \mathrm{Q}^{-1}(P_{k,c,i}^f) \tag{54}$$

where

$$\big(\sigma_{k,c,i}^{\mathrm{loc}}\big)^2 = w_{s_c}^\top P_c R_{\widetilde{w}_{k,i}}^{\mathrm{loc}} P_c^\top w_{s_c} \tag{55}$$

1053-587X (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2017.2771731, IEEE Transactions on Signal Processing

TABLE II
DIFFUSION DETECTION ALGORITHM

Initializations: w_{s_c}, $\mathcal{V}_k$, $\mathcal{I}_k$, μ_k, $a_{c,\ell k}$, $\widehat{R}_{\widetilde{w}_{\mathcal{I}_k},-1}=0$, $P^f_{k,c,i}$, and g_c (nodes $\ell,k\in\mathcal{C}_c$, cluster c, $c\in\mathcal{S}$). Start with $w_{k,c,-1}=0$ for every node k and each cluster c, $c\in\mathcal{I}_k$. For every time index $i\geq 0$, repeat

1) Run one iteration of the **Diffusion Estimation Algorithm** in Table I.
2) **Decision:** for every node k and each cluster c, $c\in\mathcal{I}_k$, repeat
   a) Test-statistic:

$$T_{k,c,i}=w_{s_c}^{\top}w_{k,c,i} \tag{59}$$

   b) Threshold:

$$R^{\rm loc}_{\widetilde{w}_{\mathcal{I}_k},i}=\left[I-\mu_k u^{\top}_{\mathcal{I}_k,i}u_{\mathcal{I}_k,i}\right]R^{\rm loc}_{\widetilde{w}_{\mathcal{I}_k},i}\left[I-\mu_k u^{\top}_{\mathcal{I}_k,i}u_{\mathcal{I}_k,i}\right]+\mu_k^2\sigma^2_{v,k}u^{\top}_{\mathcal{I}_k,i}u_{\mathcal{I}_k,i} \tag{60a}$$

$$\left(\sigma^{\rm loc}_{k,c,i}\right)^2=w_{s_c}^{\top}R^{\rm loc}_{\widetilde{w}_{k,c,i}}w_{s_c} \tag{60b}$$

$$\gamma_{k,c,i}=g_c^{-\frac{1}{2}}\cdot\sigma^{\rm loc}_{k,c,i}\,{\rm Q}^{-1}(P^f_{k,c,i}) \tag{60c}$$

   c) Test:

$$T_{k,c,i}\underset{\mathcal{H}_1^c}{\overset{\mathcal{H}_0^c}{\lessgtr}}\gamma_{k,c,i} \tag{61}$$

and the matrix $R^{\rm loc}_{\widetilde{w}_{k,i}}$ is given by

$$R^{\rm loc}_{\widetilde{w}_{k,i}}=\left[I-\mu_k u^{\top}_{k,i}u_{k,i}\right]R^{\rm loc}_{\widetilde{w}_{k,i}}\left[I-\mu_k u^{\top}_{k,i}u_{k,i}\right]+\mu_k^2\sigma^2_{v,k}u^{\top}_{k,i}u_{k,i},\qquad \widehat{R}^{\rm loc}_{\widetilde{w}_{k,-1}}=0 \tag{56}$$

which is the kth $M\times M$ diagonal block equation recovered from (52) by setting $\mathcal{A}=\mathrm{diag}\{\{W_k\}_{k\in\mathcal{V}}\}$. As a matter of fact, each node k need only run the recursion (56) employing not the augmented regressors $\{u_{k,i}\}$, but $\{u_{\mathcal{I}_k,i}\}$ in (10) instead, absolving the node of the need for global information such as the number of clusters N_s and the lengths of each of the weight vectors w_{s_c}, $c\in\mathcal{S}$:

$$R^{\rm loc}_{\widetilde{w}_{\mathcal{I}_k,i}}=\left[I-\mu_k u^{\top}_{\mathcal{I}_k,i}u_{\mathcal{I}_k,i}\right]R^{\rm loc}_{\widetilde{w}_{\mathcal{I}_k,i}}\left[I-\mu_k u^{\top}_{\mathcal{I}_k,i}u_{\mathcal{I}_k,i}\right]+\mu_k^2\sigma^2_{v,k}u^{\top}_{\mathcal{I}_k,i}u_{\mathcal{I}_k,i},\qquad \widehat{R}^{\rm loc}_{\widetilde{w}_{\mathcal{I}_k,-1}}=0. \tag{57}$$

The factor g_c in (54) accounts for the gain incurred by the diffusion process, and can be estimated offline as

$$g_c\approx\left\langle\frac{1}{|\mathcal{C}_c|}\sum_{k\in\mathcal{C}_c}\frac{\left(\sigma^{\rm loc}_{k,c,i}\right)^2}{\left(\sigma_{k,c,i}\right)^2}\right\rangle \tag{58}$$

where $\langle\cdot\rangle$ is the time-averaging operator and the variances can be computed analytically or using Monte Carlo simulations. The complete detection algorithm is listed in Table II, where $R^{\rm loc}_{\widetilde{w}_{k,c,i}}$ is the appropriate submatrix in $R^{\rm loc}_{\widetilde{w}_{\mathcal{I}_k,i}}$.

Remark 1: The central limit theorem may be invoked to argue for the asymptotic Gaussianity of the test-statistics according to the linearity of the relations (41), (19), and (7). Though this may motivate the same algorithm derivation and analysis under non-Gaussian observation models, it is generally known that both the estimation and detection performance of adaptive algorithms are impacted when the noise model departs from Gaussianity. Non-Gaussian observation models were investigated in [38] in the diffusion-based estimation context, and in [22], [23] in the diffusion-based detection context. In [22], [23], [38], adaptive update rules with error nonlinearities were developed to counteract impulsive noise and were subsequently analyzed.

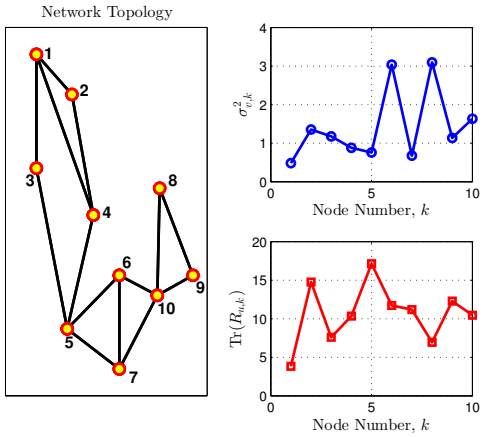


Fig. 1. Network topology, node noise variances $\sigma^2_{v,k}$, and regressor covariance traces $\mathrm{Tr}(R_{u,k})$, for $N=10$ nodes.

VI. SIMULATION RESULTS

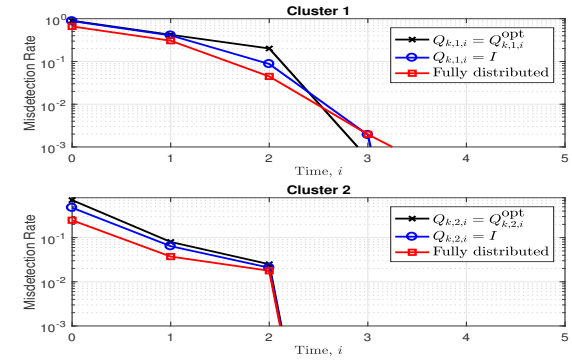
A. Algorithm Assessment

We consider $N=10$ nodes, and two node clusters $\mathcal{C}_1=\{1,2,3,4,5,6,7\}$, $\mathcal{C}_2=\{4,5,6,7,8,9,10\}$. The nodes in cluster $\mathcal{C}_c$ are interested in detecting the presence or absence of the randomly initiated unit-norm weight vectors w_{s_c}, $c=1$, 2, of dimensions $M_1=M_2=2$. The regressors $\{u_{k,c,i}\}$ and noise samples $\{v_k(i)\}$ are drawn independently of one another, independently across time and space, and identically distributed across time: For each node k, the regressors $\{\boldsymbol{u}_{k,c,i}\}$ are additionally uncorrelated for $c\in\mathcal{I}_k$, and the samples are drawn from a multivariate zero-mean Gaussian distribution with covariances $\{R_{u,k}\}$, where $R_{u,k}=\mathbb{E}\,\boldsymbol{u}^{\top}_{k,i}\boldsymbol{u}_{k,i}$; and the noise samples are drawn according to a zero-mean Gaussian distribution with variances $\{\sigma^2_{v,k}\}$. The same set of regressors is maintained throughout the experiments. The network topology, regressor covariance traces, and noise variances are shown in Fig. 1. The nodes employ the uniform combination rule for both weight vectors, i.e., the nonzero combination weights are of the form $a_{c,\ell k}=1/|\mathcal{V}_k\cap\mathcal{C}_c|$, $c=1$, 2. The step-size parameter μ_k is set to 0.01 for all k. The target false-alarm rate $P^f_{k,c,i}$ is set to 10^{-2} for all k, $c\in\mathcal{I}_k$, and i. The factor g_c was estimated offline and found to be 6, $c=1$, 2. All results are obtained over $10,000$ Monte Carlo runs.

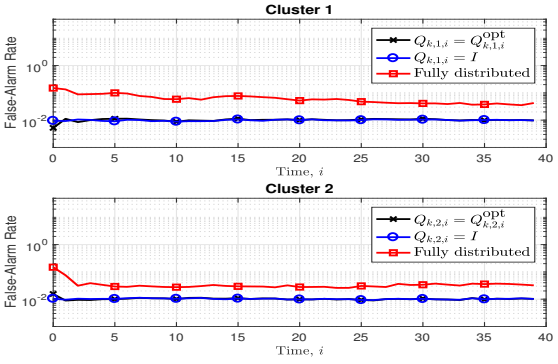
First, we numerically assess the approximations involved in deriving the diffusion detection algorithm in Table II, particularly the test-statistics (41) and thresholds (54). Specifically, we compare the detection and false-alarm performance, when the weight vectors w_{s_1} and w_{s_2} are both present or both absent, respectively, of three schemes with increasing level of approximation for the test-statistics and thresholds. The last scheme corresponds to the choices (41) and (54), respectively. The schemes are listed in Table III. The resulting average detection and false-alarm performance across the two clusters

1053-587X (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2017.2771731, IEEE Transactions on Signal Processing



(a)



(b)

Fig. 2. For a target false-alarm rate 10^{-2}, a) average detection performance across cluster 1 and cluster 2 (top and bottom plots, respectively) when both weight vectors are present, and b) average false-alarm performance when both weight vectors are absent of three schemes with increasing level of approximation for the test-statistics and thresholds. The schemes are listed in Table III.

$\mathcal{C}_1$ and $\mathcal{C}_2$ is shown in Figs. 2a and 2b, respectively. The results show that $Q_{k,c,i} = I$ is a reasonable approximation. Moreover, the fully distributed threshold computation is shown to lead to nearly the same detection performance at the expense of worse false-alarm performance. Similar results were reported for single-task networks in [21].

We consider now two scenarios to assess the detection and false-alarm performance of the diffusion detection algorithm, respectively, as well as the tracking performance. For comparison, we additionally plot throughout the results assuming no cooperation among the respective cluster nodes, as well as full cooperation—in the latter case, the uniform combination rule is employed. The factor g_c, $c = 1, 2$, is obviously 1 in the case of no cooperation, and was found to be 7 in the case of full cooperation.

1) Scenario 1—Detection Performance: The scenario is described as follows. For $i_{\text{switch}} = 250$,

$$w_s^o = \begin{cases} \begin{bmatrix} w_{s_1}^T & w_{s_2}^T \end{bmatrix}^T, & \text{if} \quad i < i_{\text{switch}} \\ \begin{bmatrix} w_{s_1}^T & 0_{1\times M} \end{bmatrix}^T, & \text{if} \quad i \geq i_{\text{switch}} \end{cases}. \tag{62}$$

See Figs. 3–4. In Fig. 3a, the simulated transient network estimation and tracking performance for the aforementioned three modes of cooperation is depicted in terms of the network mean weight-error norm, $\frac{1}{N}\|\mathbb{E}\,\widetilde{\boldsymbol{w}}_i\|$, (top), and network mean-square deviation (MSD), $\frac{1}{N}\,\mathbb{E}\|\widetilde{\boldsymbol{w}}_i\|^2$, (bottom). The average detection performance across $\mathcal{C}_1$ and $\mathcal{C}_2$ is plotted in Fig. 3b (top and bottom plots, respectively). It can be seen that cooperation among the cluster nodes boosts performance. The simulated test-statistic means for weight vectors w_{s_c}, $c = 1$, 2, of three exemplary nodes are plotted in Figs. 4a and 4b, respectively: nodes 1 (in cluster 1 only), 4 (in clusters 1 and 2), and 8 (in cluster 2 only). Even though according to the diffusion estimation algorithm in Table I neither do nodes 1, 2, or 3 compute test-statistics for w_{s_2} nor nodes 8, 9, or 10 for w_{s_1}, the results are plotted as if the nodes were running the augmented algorithm (18) instead to emphasize that node k's estimates for weight vectors c when $c \notin \mathcal{I}_k$ are zero. Finally, it is worth noting that a blip occurs in the simulated test-statistic means for cluster 1 nodes (1 through 7) at time index $i = i_{\text{switch}}$ due to transient interference from w_{s_2} as it changes state.

2) Scenario 2—False-Alarm Performance: The scenario is described as follows. For $i_{\text{switch}} = 250$,

$$w_s^o = \begin{cases} \begin{bmatrix} 0_{1\times M} & 0_{1\times M} \end{bmatrix}^T & \text{if} \quad i < i_{\text{switch}} \\ \begin{bmatrix} w_{s_1}^T & 0_{1\times M} \end{bmatrix}^T & \text{if} \quad i \geq i_{\text{switch}} \end{cases}. \tag{63}$$

See Figs. 5–6. Analogous observations to those in *Scenario 1* can be made. However, it is noteworthy that the target false-alarm rate is met only in the case of no intra-cluster cooperation, while the false-alarm performance is slightly compromised in the case of cooperation. This is due to the approximation $Q_{k,c,i} = I$, $c = 1, 2$, which is exact in the case of no cooperation.

B. Collaborative Spectrum Sensing Application

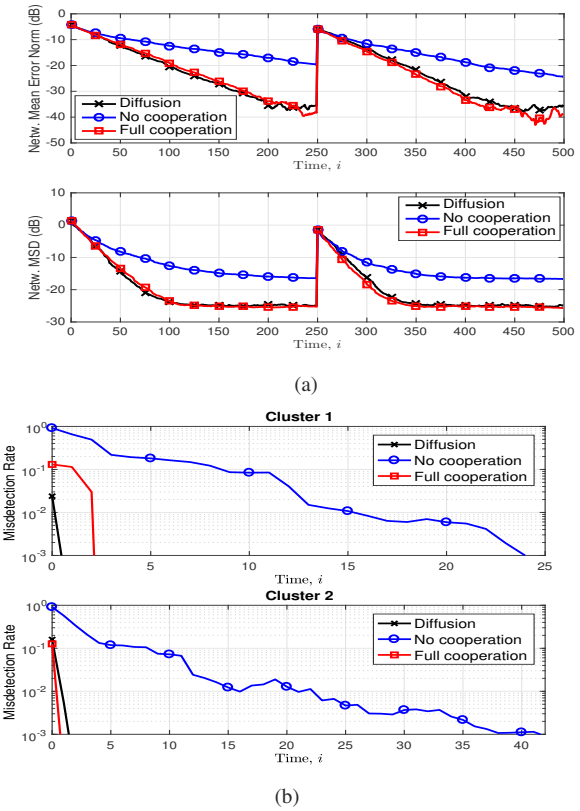
In collaborative spectrum sensing, secondary users (SUs) cooperate to detect transmissions of primary users (PUs) of licensed radio spectrum bands. SUs can access these bands in the absence of transmissions from PUs and cease their own transmissions when PUs resume theirs. In [21], the problem was formulated assuming one PU. We extend the model here

TABLE III
APPROXIMATION ASSESSMENT - ALTERNATIVE SCHEMES

Scheme Name	Test-statistic	Threshold
$Q_{k,c,i} = Q^{\text{opt}}_{k,c,i}$	$T_{k,c,i} = w_{s_c}^\top Q^{\text{opt}}_{k,c,i} w_{k,c,i}$, where $Q^{\text{opt}}_{k,c,i}$ is computed as in (40).	$\gamma_{k,c,i} = \sigma^{\text{opt}}_{k,c,i}\, \mathrm{Q}^{-1}(P^f_{k,c,i})$, where $\left(\sigma^{\text{opt}}_{k,c,i}\right)^2 = w_{s_c}^\top Q^{\text{opt}}_{k,c,i} P_c R_{\widetilde{w}_{k,i}} P_c^\top Q^{\text{opt}}_{k,c,i} w_{s_c}$ and $R_{\widetilde{w}_{k,i}}$ is given by (52).
$Q_{k,c,i} = I$	$T_{k,c,i} = w_{s_c}^\top w_{k,c,i}$	$\gamma_{k,c,i} = \sigma_{k,c,i}\, \mathrm{Q}^{-1}(P^f_{k,c,i})$, where $\sigma^2_{k,c,i} = w_{s_c}^\top P_c R_{\widetilde{w}_{k,i}} P_c^\top w_{s_c}$.
Fully distributed	$T_{k,c,i} = w_{s_c}^\top w_{k,c,i}$	See Table II.

1053-587X (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

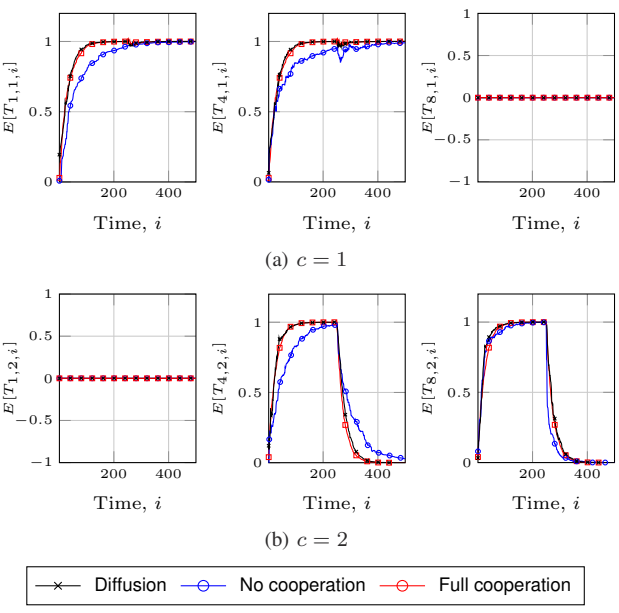
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2017.2771731, IEEE Transactions on Signal Processing



(a)

(b)

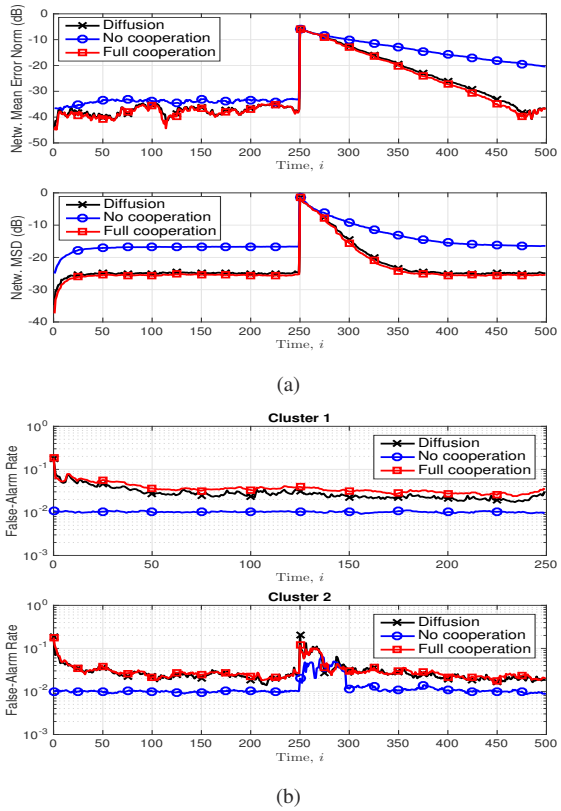
Fig. 3. *Scenario 1*: a) Transient network estimation and tracking performance in terms of network mean weight-error norm and network mean-square deviation (MSD) (top and bottom plots, respectively), approximated via Monte Carlo simulations, for three modes of cooperation. One of the weight vectors to be detected changes state at time index $i = 250$—see Eq. (62). b) For a target false-alarm rate 10^{-2}, average detection performance across cluster 1 and cluster 2 (top and bottom plots, respectively).



(a) $c = 1$

(b) $c = 2$

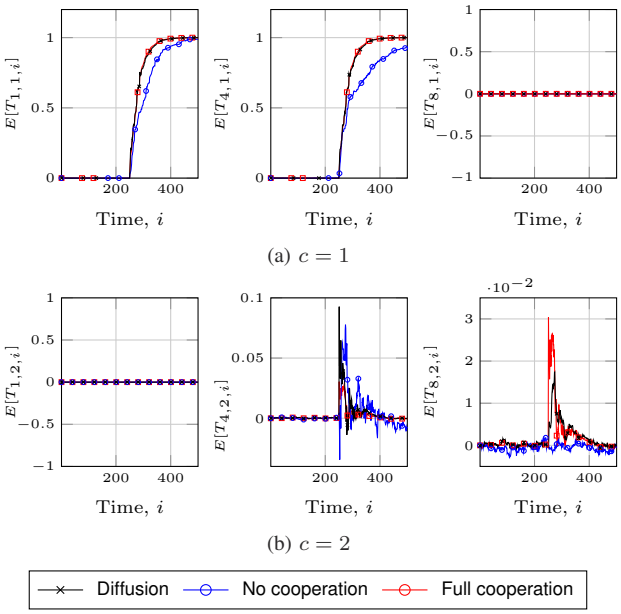
Fig. 4. *Scenario 1*: Mean of test-statistics of nodes 1 (in cluster 1 only), 4 (in clusters 1 and 2), and 8 (in cluster 2 only), approximated via Monte Carlo simulations, for weight vector w_{s_c}: a) $c = 1$, b) $c = 2$.



(a)

(b)

Fig. 5. *Scenario 2*: a) Transient network estimation and tracking performance in terms of network mean weight-error norm and network mean-square deviation (MSD) (top and bottom plots, respectively), approximated via Monte Carlo simulations, for three modes of cooperation. One of the weight vectors to be detected changes state at time index $i = 250$—see Eq. (63). b) For a target false-alarm rate 10^{-2}, average false-alarm performance across cluster 1 and cluster 2 (top and bottom plots, respectively).



(a) $c = 1$

(b) $c = 2$

Fig. 6. *Scenario 2*: Mean of test-statistics of nodes 1 (in cluster 1 only), 4 (in clusters 1 and 2), and 8 (in cluster 2 only), approximated via Monte Carlo simulations, for weight vector w_{s_c}: a) $c = 1$, b) $c = 2$.

1053-587X (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2017.2771731, IEEE Transactions on Signal Processing

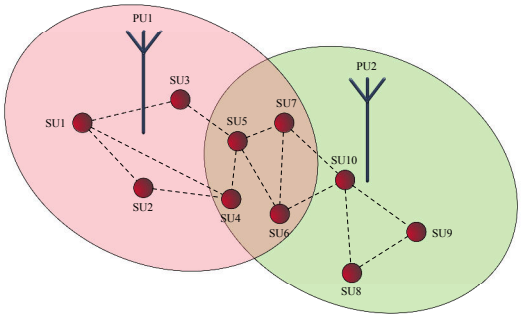


Fig. 7. Collaborative sensing application network setup.

to several PUs. A network can be used to model cooperation within and among possibly overlapping clusters of SUs, each interested in detecting the presence or absence of one PU. The SUs are indexed by k, $k \in \mathcal{V} = \{1, \ldots, N_{\text{SU}}\}$, and the PUs by c, $c \in \mathcal{S} = \{1, \ldots, N_{\text{PU}}\}$. The SUs in cluster $\mathcal{C}_c \subseteq \mathcal{V}$ are interested in tracking the cth PU. For simplicity, consider the setup in Fig. 7, with $N_{\text{SU}} = 10$ and $N_{\text{PU}} = 2$. The clusters $\mathcal{C}_1$ and $\mathcal{C}_2$ in this setup coincide with the communication ranges of the two PUs: $\mathcal{C}_1 = \{1, 2, 3, 4, 5, 6, 7\}$, $\mathcal{C}_2 = \{4, 5, 6, 7, 8, 9, 10\}$, with the nodes 4, 5, 6, and 7 lying within both communication ranges. The cth PU's continuous-time real-valued signal is denoted as $s_c^o(t)$, $t \in \mathbb{R}$, where, over some time interval, $s_c^o(t) = 0$ under $\mathcal{H}_0^c$, and $s_c^o(t) = s_c(t) \neq 0$ under $\mathcal{H}_1^c$. The channels from the PUs to the SUs are assumed to be band-limited. Sampling at a frequency $\frac{1}{T_s}$ satisfying the Nyquist sampling theorem, the received signal at the kth SU at time $t = nT_s$, $n \in \mathbb{Z}$, is given by $\boldsymbol{y}_k(nT_s) = r_{1k}^o(nT_s) + r_{2k}^o(nT_s) + \boldsymbol{\nu}_k(nT_s)$, where, for $c = 1, 2$,

$$r_{ck}^o(nT_s) = \sum_{j=0}^{L_{ck}-1} h_{ck}(j) s_c^o\Big((n-j)T_s\Big).$$

The channel $h_{ck}(j)$ from the cth PU to the kth SU is modeled as a time-invariant finite-impulse response filter with L_{ck} taps, and it is nonzero only if $k \in \mathcal{C}_c$, i.e., when the kth SU is within the cth PU's communication range. The signal $s_c(nT_s)$, $c = 1, 2$, is assumed to be known to all nodes $k \in \mathcal{C}_c$, e.g., $s_c(t)$ is a pilot or frame preamble signal whose samples may be estimated by the respective nodes during instances when they know the PU to be active. The channel $h_{ck}(j)$ is also assumed to be known to the kth SU or to have been estimated. The noise $\boldsymbol{\nu}_k(i)$ is a scalar wide-sense stationary zero-mean real Gaussian noise process, spatially independent and temporally white with variance $\sigma_{\nu,k}^2 = \mathbb{E}\,\boldsymbol{\nu}_k^2(t)$. The noise variance $\sigma_{\nu,k}^2$ is assumed as well to be known to the kth SU or to have been estimated during instances of PU inactivity.

As in [21], it is assumed that the kth SU observes a block of $L = \frac{T}{T_s}$ samples of $\boldsymbol{y}_k(t)$, then it processes them for a diffusion iteration at time $t = iT$, $i \geq 0$, where L is the block length and T is the block duration. It is also assumed that the signal $s_c(t)$, $c = 1, 2$, repeats itself every L samples, i.e., $s_c(nTs) = s_c\big((n+L)T_s\big)$ for all n. At time $t = iT$, the kth SU collects the last L samples of $\boldsymbol{y}_k(t)$ from time instants $t = (i-1)T + T_s$ to $t = iT$ into a vector, leading to the data model $\boldsymbol{y}_{k,i} = H_k s^o + \boldsymbol{\nu}_{k,i}$ following the definitions

$$\begin{aligned}
\boldsymbol{y}_{k,i} &= \Big[\boldsymbol{y}_k(iT)\, \boldsymbol{y}_k(iT - T_s) \ldots \boldsymbol{y}_k\Big(iT - (L-1)T_s\Big)\Big]^\top \\
H_k &= [H_{1k}\ H_{2k}] \\
H_{ck} &= \begin{bmatrix} h_{ck}(0) & \ldots & h_{ck}(L_{ck}-1) & \ldots & 0 \\ \vdots & \ddots & & \ddots & \vdots \\ 0 & \ldots & h_{ck}(0) & \ldots & h_{ck}(L_{ck}-1) \end{bmatrix} \\
s &= [s_1\ s_2]^\top \\
s_c &= \Big[s_c(iT)\, s_c(iT - T_s) \ldots s_c\Big(iT - (M_c-1)T_s\Big)\Big]^\top \\
s^o &= [s_1^o\ s_2^o]^\top \\
s_c^o &= \Big[s_c^o(iT)\, s_c^o(iT - T_s) \ldots s_c^o\Big(iT - (M_c-1)T_s\Big)\Big]^\top \\
\boldsymbol{\nu}_{k,i} &= \Big[\boldsymbol{\nu}_k(iT)\, \boldsymbol{\nu}_k(iT - T_s) \ldots \boldsymbol{\nu}_k\Big(iT - (L-1)T_s\Big)\Big]^\top
\end{aligned}$$

of dimensions $L \times 1$, $L \times M$, $L \times M_c$, $M \times 1$, $M_c \times 1$, $M \times 1$, $M_c \times 1$, and $L \times 1$, respectively, where $M_c = L + L_{ck} - 1$, $c = 1, 2$. The noise vectors $\boldsymbol{\nu}_{k,i}$ and $\boldsymbol{\nu}_{\ell,j}$ are independent for $k \neq \ell$ or $i \neq j$. The binary hypothesis tests are given by $s_c^o = 0$ under $\mathcal{H}_0^c$, and $s_c^o = s_c \neq 0$ under $\mathcal{H}_1^c$. Using the following definitions:

$$\begin{aligned}
\boldsymbol{d}_k(i) &= \frac{1}{\sigma_{\nu,k}\|H_k s\|} s^\top H_k^\top \boldsymbol{y}_{k,i} \\
u_{k,i} &= [u_{k,1,i}\ u_{k,2,i}] = \frac{1}{\sigma_{\nu,k}\|H_k s\|} s^\top H_k^\top H_k \qquad (1 \times M) \\
w_s &= [w_{s_1}\ w_{s_2}] = s \qquad (M \times 1) \\
w_s^o &= [w_{s_1}^o\ w_{s_2}^o] = s^o \qquad (M \times 1) \\
\boldsymbol{v}_k(i) &= \frac{1}{\sigma_{\nu,k}\|H_k s\|} s^\top H_k^\top \boldsymbol{\nu}_{k,i}
\end{aligned}$$

and noting that with the proposed construction for H_k it holds that $u_{k,c,i} \neq 0$, if $k \in \mathcal{C}_c$, and $u_{k,c,i} = 0$, if $k \notin \mathcal{C}_c$, $c = 1, 2$, we thus obtain the data model $\boldsymbol{d}_k(i) = u_{k,i} w_s^o + \boldsymbol{v}_k(i)$, which is identical to the data model in Sec. II with time-invariant regressors and noise variance $\sigma_{v,k}^2 = 1$ for all k.

For the simulation, $L = 4$ samples per observation block, $L_{1k} = L_{2k} = 2$ channel taps for all k, leading to lengths $M_1 = M_2 = 5$ for the weight vectors w_{s_1} and w_{s_2}. The channel taps $h_{ck}(0), \ldots, h_{ck}(L_{ck}-1)$ for the kth SU are drawn independently according to zero-mean Gaussian distributions with variance $\sigma_{h_{c,k}}^2$, $c = 1, 2$, as plotted in the bottom plots of Fig. 8, while the noise variances $\{\sigma_{\nu,k}^2\}$ are plotted in the top plot. The factors g_1 and g_2 were estimated offline and found to be 3 and 1, respectively. All other parameters are the same as in Sec. VI-A. For the case when both PUs are present, the results showed that the average misdetection rate across the two clusters is zero for a target false-alarm rate of 10^{-2}, while the resulting average false-alarm performance for the case when both PUs are absent is 0.02 for cluster 1 and 0.09 for cluster 2.

VII. Conclusion

We proposed a distributed online detection algorithm based on node-specific diffusion LMS estimation to solve the prob-

1053-587X (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2017.2771731, IEEE Transactions on Signal Processing

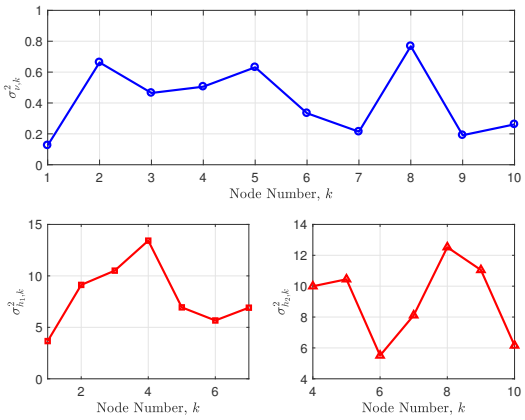


Fig. 8. SU noise variances $\sigma^2_{\nu,k}$ (top) and channel tap variances $\sigma^2_{h_c,k}$, $c = 1, 2$, for the SUs of clusters 1 and 2, respectively (bottom).

lem where each node in a network would like to establish the presence or absence of each of its parameters of interest, with possible overlap among the node interests. The developed detectors are motivated by the dynamics of the diffusion estimation algorithm, are fully distributed, and are able to track changing conditions in real time. The detection performance of the algorithm is intertwined with its estimation performance, which is analyzed using the energy-conservation framework, showing that the diffusion LMS iterates converge in the mean and mean-square sense to the true weight vectors given sufficiently small step-sizes. The approximations undertaken in deriving the detectors as well as the resulting performance were assessed in simulations. Finally, we demonstrated the performance of the distributed detection algorithm in a practical collaborative sensing application.

Appendix A
Mean-Square Stability of Diffusion Estimation Algorithm

Let us define the update vector

$$\hat{\boldsymbol{s}}_{k,i}(\boldsymbol{w}_{k,i-1}) = -u_{k,i}^{\top}[\boldsymbol{d}_k(i) - u_{k,i}\boldsymbol{w}_{k,i-1}]. \tag{64}$$

The ATC diffusion algorithm in augmented form, written in terms of the update vector (64), is given by the following update equations:

$$\left\{ \begin{array}{lcl} \boldsymbol{\psi}_{k,i} & = & \boldsymbol{w}_{k,i} - \mu_k \hat{\boldsymbol{s}}_{k,i}(\boldsymbol{w}_{k,i-1}) \\ \boldsymbol{w}_{k,i} & = & \sum_{\ell\in\mathcal{V}_k} A_{\ell k}\boldsymbol{\psi}_{\ell,i} \end{array} \right. \tag{65}$$

where $A_{\ell k}$ is of the form (9) and (17). We show in the following lemma that $\hat{\boldsymbol{s}}_{k,i}(\boldsymbol{w}_{k,i-1})$ satisfies certain properties that will be exploited in the analysis.

Lemma 2 (Update Vector Properties). Under the data model assumptions in Sec. II, the update vector $\hat{\boldsymbol{s}}_{k,i}$ (64) for each k and i satisfies the following properties:

Randomness: There exists an $M \times 1$ deterministic vector function $s_{k,i}(\boldsymbol{w})$ such that, for all $M \times 1$ vectors $\boldsymbol{w}$ in the filtration $\boldsymbol{\mathcal{F}}_{i-1}$ generated by the past history of iterates $\{\boldsymbol{w}_{k,j}\}$ for $j \le i-1$, it holds that

$$\mathbb{E}\{\hat{\boldsymbol{s}}_{k,i}(\boldsymbol{w})|\boldsymbol{\mathcal{F}}_{i-1}\} = s_{k,i}(\boldsymbol{w}) \tag{66}$$

where

$$s_{k,i}(\boldsymbol{w}) \triangleq -u_{k,i}^{\top}u_{k,i}(w^o - \boldsymbol{w}). \tag{67}$$

Moreover, for each k, there exists $\sigma^2_{s,k} \ge 0$ such that for each i and all $\boldsymbol{w} \in \boldsymbol{\mathcal{F}}_{i-1}$, it holds that

$$\mathbb{E}\Big\{\|\hat{\boldsymbol{s}}_{k,i}(\boldsymbol{w}) - s_k(\boldsymbol{w})\|^2 \big| \boldsymbol{\mathcal{F}}_{i-1}\Big\} \le \sigma^2_{s,k} \tag{68}$$

where $\sigma^2_{s,k} \triangleq \sigma^2_{v,k}\max_i \|u_{k,i}\|^2$.

Lipschitz: There exists $\lambda_U > 0$ such that for all x, $y \in \mathbb{R}^M$, it holds that

$$\|s_{k,i}(x) - s_{k,i}(y)\| \le \lambda_U\|x - y\| \tag{69}$$

where the subscript "U" in λ_U refers to the latter determining the upper bound with $\lambda_U \triangleq \max_{k,i}\|u_{k,i}\|^2$. This property is guaranteed by the bounded energy assumption on the regressors in Sec. II, where in this appendix λ_U is used instead of α for notational convenience.

Strong monotonicity: There exists $\lambda_L > 0$, $\lambda_L \le \lambda_U$, such that for all x, $y \in \mathbb{R}^M$, it holds that

$$(x-y)^{\top}\sum_{k\in\mathcal{V}}[s_{k,i}(x) - s_{k,i}(y)] \ge \lambda_L\|x - y\|^2 \tag{70}$$

where the subscript "L" in λ_L refers to the latter determining the lower bound with $\lambda_L \triangleq \min_i \lambda_{\min}\big(\sum_{k\in\mathcal{V}} u_{k,i}^{\top}u_{k,i}\big)$. Here, $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue of its square matrix argument. This property is guaranteed by the full-rank assumption on U_i in Sec. II. □

Properties equivalent to the abovementioned Lipschitz and strong monotonicity properties can be established as follows. First, let

$$H_{k,i} \triangleq \nabla_{\boldsymbol{w}} s_{k,i}(\boldsymbol{w}) = u_{k,i}^{\top}u_{k,i} \tag{71}$$

for any $\boldsymbol{w} \in \boldsymbol{\mathcal{F}}_{i-1}$. Then, the equivalent properties are

$$H_{k,i} \ge 0, \quad \|H_{k,i}\| \le \lambda_U, \quad \sum_{k\in\mathcal{V}} H_{k,i} \ge \lambda_L I_M. \tag{72}$$

Extended error dynamics: The objective is to examine the evolution of the offset between the estimate $\boldsymbol{w}_{k,i}$ for each k and the limit point $w^o_k \triangleq W_k w^o_s$, where

$$W_k \triangleq \mathrm{diag}\big\{\{\mathbb{I}_{\mathcal{I}_k}(c)I_{M_c}\}_{c\in\mathcal{S}}\big\}. \tag{73}$$

We first define the error quantities

$$\widetilde{\boldsymbol{\psi}}_{k,i} \triangleq w^o_k - \boldsymbol{\psi}_{k,i}, \qquad \widetilde{\boldsymbol{w}}_{k,i} \triangleq w^o_k - \boldsymbol{w}_{k,i} \tag{74}$$

Due to the structure of the regressors according to (5) and (2), the expected value of the update vector can be written for each k, i, and $\boldsymbol{w}_{k,i-1} \in \boldsymbol{\mathcal{F}}_{i-1}$ as

$$\begin{aligned} s_{k,i}(\boldsymbol{w}_{k,i-1}) &= -u_{k,i}^{\top}u_{k,i}(w^o - \boldsymbol{w}_{k,i-1}) \\ &= -u_{k,i}^{\top}u_{k,i}W_k(w^o - \boldsymbol{w}_{k,i-1}) \\ &= -u_{k,i}^{\top}u_{k,i}(w^o_k - \boldsymbol{w}_{k,i-1}) \\ &= -H_{k,i}\widetilde{\boldsymbol{w}}_{k,i-1} \end{aligned} \tag{75}$$

where we used (74) and (71).

For node k, time index i, and any $\boldsymbol{w} \in \boldsymbol{\mathcal{F}}_{i-1}$, let

$$\boldsymbol{v}^s_{k,i}(\boldsymbol{w}) \triangleq \hat{\boldsymbol{s}}_{k,i}(\boldsymbol{w}) - s_{k,i}(\boldsymbol{w}) = -u_{k,i}^{\top}\boldsymbol{v}_k(i) \equiv \boldsymbol{v}^s_{k,i} \tag{76}$$

1053-587X (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2017.2771731, IEEE Transactions on Signal Processing

represent the noise incurred by stochastic approximation. We refer to it as the update noise vector. Then, the update equations (65) using (76) and (75) become

$$\left\{\begin{array}{rcl}\boldsymbol{\psi}_{k,i} &=& \boldsymbol{w}_{k,i-1}+\mu_k H_{k,i}\widetilde{\boldsymbol{w}}_{k,i-1}-\mu_k \boldsymbol{v}^s_{k,i}\\ \boldsymbol{w}_{k,i} &=& \sum_{\ell\in\mathcal{V}_k} A_{\ell k}\boldsymbol{\psi}_{\ell,i}\end{array}\right. \tag{77}$$

Subtracting each of the update equations in (77) from w^o_k and using the fact that $\sum_{\ell\in\mathcal{V}_k} A_{\ell k} w^o_k = w^o_k$,

$$\left\{\begin{array}{rcl}\widetilde{\boldsymbol{\psi}}_{k,i} &=& [I-\mu_k H_{k,i}]\widetilde{\boldsymbol{w}}_{k,i-1}+\mu_k \boldsymbol{v}^s_{k,i}\\ \widetilde{\boldsymbol{w}}_{k,i} &=& \sum_{\ell\in\mathcal{V}_k} A_{\ell k}\widetilde{\boldsymbol{\psi}}_{\ell,i}\end{array}\right. \tag{78}$$

We introduce the following network variables:

$$\widetilde{\boldsymbol{\psi}}_i \triangleq \left[\begin{array}{c}\widetilde{\boldsymbol{\psi}}_{1,i}\\ \vdots\\ \widetilde{\boldsymbol{\psi}}_{N,i}\end{array}\right],\ \widetilde{\boldsymbol{w}}_i \triangleq \left[\begin{array}{c}\widetilde{\boldsymbol{w}}_{1,i}\\ \vdots\\ \widetilde{\boldsymbol{w}}_{N,i}\end{array}\right],\ \boldsymbol{v}^s_i \triangleq \left[\begin{array}{c}\boldsymbol{v}^s_{1,i}\\ \vdots\\ \boldsymbol{v}^s_{N,i}\end{array}\right] \tag{79}$$

Recalling that $\mathcal{A}=[A_{\ell k}]$, $\ell,\ k\in\mathcal{V}$, the network relations become

$$\left\{\begin{array}{rcl}\widetilde{\boldsymbol{\psi}}_i &=& [I-\mathcal{M}\mathcal{H}_i]\widetilde{\boldsymbol{w}}_{i-1}+\mathcal{M}\boldsymbol{v}^s_i\\ \widetilde{\boldsymbol{w}}_i &=& \mathcal{A}^\top\widetilde{\boldsymbol{\psi}}_i\end{array}\right. \tag{80}$$

where $\mathcal{M}$ and $\mathcal{H}_i$ were defined in (26) and (27), respectively. This leads to the following network weight-error recursion:

$$\widetilde{\boldsymbol{w}}_i = \mathcal{B}_i\widetilde{\boldsymbol{w}}_{i-1}+\mathcal{A}^\top\mathcal{M}\boldsymbol{v}^s_i,\quad i\geq 0 \tag{81}$$

with feedback matrix

$$\mathcal{B}_i \triangleq \mathcal{A}^\top(I-\mathcal{M}\mathcal{H}_i). \tag{82}$$

Properties of the matrix A_c: Consider the $|\mathcal{C}_c|\times|\mathcal{C}_c|$ combination matrix $[A_c]_{\mathcal{C}_c}$, formed from A_c by keeping only the rows and columns with indices in $\mathcal{C}_c$, where $c\in\mathcal{S}$. Under the assumptions from Sec. II that each of the clusters is connected and that each node is its own neighbor, it follows that the matrices $[A_c]_{\mathcal{C}_c}$ are primitive. The following properties then hold for each matrix A_c:

a) The matrix A_c has a single eigenvalue at 1 corresponding to the left eigenvector $\bar{\mathbb{1}}_c$ with entries

$$\bar{\mathbb{1}}_{c,k} = \begin{cases}1, & \text{if } k\in\mathcal{C}_c\\ 0, & \text{otherwise}\end{cases}. \tag{83}$$

b) All other eigenvalues of A_c lie strictly inside the unit circle.

c) The right eigenvector corresponding to the eigenvalue at 1 of A_c is given by p_c such that $A_c p_c = p_c$, with entries $p_{c,k}$ that with proper scaling and normalization sum up to one: $\mathbb{1}^\top p_c = 1$, and are positive, if $k\in\mathcal{C}_c$, and 0 otherwise.

Jordan canonical decomposition of $\mathcal{A}$: The matrix $\mathcal{A}$ is now to be decomposed in terms of the Jordan canonical decomposition of the matrices A_c, $c\in\mathcal{S}$, of the form [39, P. 128], for $\epsilon>0$:

$$A_c = V_{\epsilon,c}J_c V^{-1}_{\epsilon,c} \tag{84}$$

where

$$J_c = \left[\begin{array}{cc}1 & \\ & J_{\epsilon,c}\end{array}\right],\ V_{\epsilon,c}=[p_c|V_{R,c}],\ V^{-1}_{\epsilon,c}=\left[\begin{array}{c}\bar{\mathbb{1}}_c^\top\\ V_{L,c}^\top\end{array}\right] \tag{85}$$

Let

$$E_c \triangleq \mathrm{diag}\Big\{\big\{\mathbb{I}_{c'=c}I_{M_{c'}}\big\}_{c'\in\mathcal{S}}\Big\} = E_c^\top, \tag{86}$$

which is a block diagonal matrix whose cth diagonal block is the identity matrix of size M_c, while the other diagonal blocks c', of sizes $M_{c'}$, are all-zero matrices. Then, the Jordan canonical decomposition of $\mathcal{A}$ can be expressed as[2]

$$\mathcal{A} = \mathcal{V}_\epsilon\mathcal{J}\mathcal{V}_\epsilon^{-1} \tag{87}$$

where

$$\mathcal{J} = \mathrm{diag}\left\{I_M,\ \underbrace{\overbrace{\sum_{c\in\mathcal{S}} J_{\epsilon,c}\otimes E_c}^{\triangleq\mathcal{J}_\epsilon}}_{(N-1)M\times(N-1)M}\right\} \tag{88}$$

$$\mathcal{V}_\epsilon = \left[\underbrace{\sum_{c\in\mathcal{S}} p_c\otimes E_c}_{NM\times M}\ \Bigg|\ \underbrace{\sum_{c\in\mathcal{S}} V_{R,c}\otimes E_c}_{NM\times(N-1)M}\right] \tag{89}$$

$$\mathcal{V}_\epsilon^{-1} = \left[\begin{array}{c}\sum_{c\in\mathcal{S}}\bar{\mathbb{1}}_c^\top\otimes E_c^\top\\ \hline \sum_{c\in\mathcal{S}} V_{L,c}^\top\otimes E_c^\top\end{array}\right]\begin{array}{l}M\times NM\\ (N-1)M\times NM\end{array} \tag{90}$$

Now, the feedback matrix $\mathcal{B}_i$ defined in (82) can be expressed in terms of the decomposition as

$$\mathcal{B}_i = \big(\mathcal{V}_\epsilon^{-1}\big)^\top\big(\mathcal{J}^\top-\mathcal{D}_i^\top\big)\mathcal{V}_\epsilon^\top \tag{91}$$

with $\mathcal{D}_i^\top \triangleq \mathcal{V}_\epsilon^\top\mathcal{A}^\top\mathcal{M}\mathcal{H}_i\big(\mathcal{V}_\epsilon^{-1}\big)^\top$. Let us consider the following partitioning of the matrix $\mathcal{D}_i^\top$:

$$\mathcal{D}_i^\top = \begin{bmatrix}D_{i,11}^\top & D_{i,21}^\top\\ D_{i,12}^\top & D_{i,22}^\top\end{bmatrix} \tag{92}$$

where $D_{i,11}^\top$ is an $M\times M$ matrix. We will proceed to evaluate each of the partitions. As for $D_{i,11}^\top$,

$$D_{i,11}^\top = \left(\sum_{c\in\mathcal{S}} p_c^\top\otimes E_c\right)\mathcal{A}^\top\mathcal{M}\mathcal{H}_i\left(\sum_{c\in\mathcal{S}}\bar{\mathbb{1}}_c\otimes E_c\right). \tag{93}$$

The term pre-multiplying $\mathcal{H}_i$ can be simplified:

$$\begin{aligned}&\mathcal{M}\mathcal{A}\left(\sum_{c\in\mathcal{S}} p_c\otimes E_c\right)\\ &=\mathcal{M}\left(\sum_{c\in\mathcal{S}} A_c\otimes E_c\right)\left(\sum_{c\in\mathcal{S}} p_c\otimes E_c\right)\\ &=\mathcal{M}\left(\sum_{c\in\mathcal{S}} A_c p_c\otimes E_c\right)=\mathcal{M}\left(\sum_{c\in\mathcal{S}} p_c\otimes E_c\right)\\ &=\sum_{c\in\mathcal{S}}(\mathrm{diag}\{\mu_1,\ldots,\mu_N\}\otimes I_M)(p_c\otimes E_c)\\ &=\sum_{c\in\mathcal{S}}(\mathrm{diag}\{\mu_1,\ldots,\mu_N\}p_c)\otimes E_c\\ &=\sum_{c\in\mathcal{S}} q_c\otimes E_c\end{aligned} \tag{94}$$

where

$$q_c \triangleq \mathrm{diag}\{\mu_1,\ldots,\mu_N\}p_c \tag{95}$$

[2]Equivalently, $\mathcal{A}=\mathcal{V}\mathcal{J}'\mathcal{V}^{-1}$, where $\mathcal{J}'$ has ones instead of ϵ on the lower diagonal. Let $x\triangleq[1,\epsilon,\ldots,\epsilon^{N-1}]^\top$, $x_\epsilon\triangleq[\epsilon,\epsilon^2,\ldots,\epsilon^{N-1}]^\top$, so $x=\begin{bmatrix}1 & x_\epsilon^\top\end{bmatrix}^\top$, and $X_\epsilon\triangleq\mathrm{diag}\{x_\epsilon\}$. Let $\mathcal{X}\triangleq\mathrm{diag}\big\{I_M,\sum_{c\in\mathcal{S}}X_\epsilon\otimes E_c\big\}$. Then, $\mathcal{A}=\mathcal{V}\mathcal{X}^{-1}\mathcal{X}\mathcal{J}'\mathcal{X}^{-1}\mathcal{X}\mathcal{V}^{-1}=\mathcal{V}_\epsilon\mathcal{J}\mathcal{V}_\epsilon^{-1}$, where $\mathcal{J}=\mathcal{X}\mathcal{J}'\mathcal{X}^{-1}$, $\mathcal{V}_\epsilon\triangleq\mathcal{V}\mathcal{X}^{-1}$, and $\mathcal{V}_\epsilon^{-1}\triangleq\mathcal{X}\mathcal{V}^{-1}$.

1053-587X (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2017.2771731, IEEE Transactions on Signal Processing

with entries $q_{c,k} = \mu_k p_{c,k}$. Substituting (94) into (93),

$$\begin{aligned} D_{i,11}^{\top} &= \left(\sum_{c\in\mathcal{S}} q_c^{\top}\otimes E_c\right)\mathcal{H}_i\left(\sum_{c\in\mathcal{S}} \bar{1}_c\otimes E_c\right) \\ &= \sum_{k=1}^{N}\left(\sum_{c\in\mathcal{S}} q_{c,k}E_c\right)H_{k,i}\left(\sum_{c\in\mathcal{S}} \bar{1}_{c,k}E_c\right) \\ &= \sum_{k=1}^{N} Q_k H_{k,i} W_k \qquad \text{(Eqs. (83), (73))} \\ &= \sum_{k=1}^{N} Q_k H_{k,i} \qquad \text{(Eqs. (71), (5), (2))} \end{aligned} \tag{96}$$

where

$$Q_k \triangleq \operatorname{diag}\big\{\{q_{c,k}I_{M_c}\}_{c\in\mathcal{S}}\big\}. \tag{97}$$

It follows that

$$D_{i,11} = \sum_{k=1}^{N} H_{k,i}Q_k. \tag{98}$$

The other three partitions of the matrix $\mathcal{D}_i$ can be simplified similarly.

Bounding the matrix $\mathcal{D}_i$: Let $\mu_{\max} \triangleq \max_k \mu_k$. Then, the step-sizes can be expressed as

$$\mu_k = \tau_k \mu_{\max} \tag{99}$$

where $0 < \tau_k \leq 1$, $k \in \mathcal{V}$. Note that following (95) and (99), the entries of the vector q_c, $q_{c,k}$, $k \in \mathcal{V}$, $c \in \mathcal{S}$, can be expressed as

$$q_{c,k} = \mu_k p_{c,k} = \mu_{\max}\tau_k p_{c,k}. \tag{100}$$

It is clear from this relation that, like $\{p_{c,k}\}$, $q_{c,k} > 0$ if $k \in \mathcal{C}_c$ and 0 otherwise.

We will proceed to show that the partitions of the matrix $\mathcal{D}_i$ are bounded. For example, $D_{i,11} = \sum_{k=1}^{N} H_{k,i}Q_k$ is bounded from above and below for all i as follows. Consider the transposed matrix, $D_{i,11}^{\top}$:

$$\begin{aligned} \sum_{k\in\mathcal{V}} Q_k H_{k,i} &= \sum_{k\in\mathcal{V}}\sum_{c\in\mathcal{S}} q_{c,k}E_c H_{k,i} \\ &= \sum_{c\in\mathcal{S}}\sum_{k\in\mathcal{V}} q_{c,k}E_c H_{k,i} \\ &\geq \sum_{c\in\mathcal{S}}\left[\left(\min_{k',\, q_{c,k'}>0} q_{c,k'}\right)\sum_{k\in\mathcal{V}} E_c H_{k,i}\right] \\ &\geq q_{\min}\sum_{c\in\mathcal{S}}\sum_{k\in\mathcal{V}} E_c H_{k,i} \\ &= q_{\min}\sum_{k\in\mathcal{V}}\sum_{c\in\mathcal{S}} E_c H_{k,i} = q_{\min}\sum_{k\in\mathcal{V}} H_{k,i} \\ &\geq q_{\min}\lambda_L I_M \qquad \text{(Eq. (72))} \\ &= \mathcal{O}(\mu_{\max}) \end{aligned} \tag{101}$$

where $q_{\min} \triangleq \min_{c,\,k,\,q_{c,k}>0} q_{c,k}$. Following the same reasoning, exploiting the Lipschitz property (72), it can be shown that

$$\sum_{k\in\mathcal{V}} Q_k H_{k,i} \leq q_{\max} N \lambda_U I_M = \mathcal{O}(\mu_{\max}) \tag{102}$$

where $q_{\max} \triangleq \max_{c,\,k} q_{c,k}$. Hence, there exist positive constants, θ_1 and θ_2, where $\theta_1 \leq \theta_2$, independent of $\mu_{\max}$ such that the following relation holds for all i:

$$\theta_1 \mu_{\max} I_M \leq D_{i,11} \leq \theta_2 \mu_{\max} I_M. \tag{103}$$

Since the eigenvalues of $D_{i,11}$ are real-valued and positive, they are bounded as

$$\theta_1 \mu_{\max} \leq \lambda(D_{i,11}) \leq \theta_2 \mu_{\max}. \tag{104}$$

This implies that the eigenvalues of $I_M - D_{i,11}^{\top}$ are $1-\mathcal{O}(\mu_{\max})$ for sufficiently small $\mu_{\max}$ such that

$$\|I_M - D_{i,11}\| \leq 1 - \sigma_{11}\mu_{\max} = 1 - \mathcal{O}(\mu_{\max}) \tag{105}$$

for some positive constant σ_{11} independent of $\mu_{\max}$ and i.

Similarly, it can be shown that $D_{i,12} = \mathcal{O}(\mu_{\max})$, $D_{i,21} = \mathcal{O}(\mu_{\max})$, and $D_{i,22} = \mathcal{O}(\mu_{\max})$ for all i. For example,

$$\begin{aligned} \|D_{i,21}\| &\leq \left\|\sum_{c\in\mathcal{S}} V_{L,c}^{\top}\otimes E_c\right\|\left\|\sum_{c\in\mathcal{S}} q_c\otimes E_c\right\|\|\mathcal{H}_i\| \\ &\leq \left\|\sum_{c\in\mathcal{S}} V_{L,c}^{\top}\otimes E_c\right\|\left\|\sum_{c\in\mathcal{S}} q_c\otimes E_c\right\|\left(\max_{k\in\mathcal{V}}\|H_{k,i}\|\right) \\ &\leq \left\|\sum_{c\in\mathcal{S}} V_{L,c}^{\top}\otimes E_c\right\|\left\|\sum_{c\in\mathcal{S}} q_c\otimes E_c\right\|\lambda_U \\ &= \left\|\sum_{c\in\mathcal{S}} V_{L,c}^{\top}\otimes E_c\right\|\max_{c\in\mathcal{S}}\|q_c\|\lambda_U \qquad (106a) \\ &\leq \left\|\sum_{c\in\mathcal{S}} V_{L,c}^{\top}\otimes E_c\right\|\sqrt{N q_{c_o,k_o}^2}\lambda_U \\ &= \left\|\sum_{c\in\mathcal{S}} V_{L,c}^{\top}\otimes E_c\right\|\sqrt{N}\mu_{\max}\tau_{k_o}p_{c_o,k_o}\lambda_U \qquad (106b) \end{aligned}$$

where $c_o \triangleq \arg\max_{c\in\mathcal{S}}\|q_c\|$, $k_o \triangleq \arg\max_{k\in\mathcal{V}} q_{c_o,k}$, and (100) was used in (106b). The upper bound in (106a) can be motivated as follows:

$$\left\|\sum_{c\in\mathcal{S}} q_c\otimes E_c\right\| = \max_{c\in\mathcal{S}}\|q_c\otimes E_c\| = \max_{c\in\mathcal{S}}\|q_c\|. \tag{107}$$

It follows that $\|D_{i,21}\| \leq \sigma_{21}\mu_{\max} = \mathcal{O}(\mu_{\max})$, where $\sigma_{21} > 0$ is a constant that is independent of $\mu_{\max}$ and i. Similarly, $\|D_{i,12}\| \leq \sigma_{12}\mu_{\max} = \mathcal{O}(\mu_{\max})$, $\|D_{i,22}\| \leq \sigma_{22}\mu_{\max} = \mathcal{O}(\mu_{\max})$, for some constants $\sigma_{12}, \sigma_{22} > 0$ that are independent of $\mu_{\max}$ and i.

Transformed network weight-error recursion: The feedback matrix $\mathcal{B}_i$ in (91) can be expressed in terms of the partitions of the matrix $\mathcal{D}_i$ as

$$\mathcal{B}_i = \big(\mathcal{V}_{\epsilon}^{-1}\big)^{\top}\begin{bmatrix} I_M - D_{i,11}^{\top} & -D_{i,21}^{\top} \\ -D_{i,12}^{\top} & \mathcal{J}_{\epsilon}^{\top} - D_{i,22}^{\top}\end{bmatrix}\mathcal{V}_{\epsilon}^{\top}. \tag{108}$$

Transforming the network weight-error recursion (81) by pre-multiplying by $\mathcal{V}_{\epsilon}^{\top}$, then, it holds for $i \geq 0$ that

$$\mathcal{V}_{\epsilon}^{\top}\widetilde{\boldsymbol{w}}_i = \mathcal{V}_{\epsilon}^{\top}\mathcal{B}_i\big(\mathcal{V}_{\epsilon}^{-1}\big)^{\top}\mathcal{V}_{\epsilon}^{\top}\widetilde{\boldsymbol{w}}_{i-1} + \mathcal{V}_{\epsilon}^{\top}\mathcal{A}^{\top}\mathcal{M}\boldsymbol{v}_i^s. \tag{109}$$

Let

$$\mathcal{V}_{\epsilon}^{\top}\widetilde{\boldsymbol{w}}_i = \begin{bmatrix}\big(\sum_{c\in\mathcal{S}} p_c^{\top}\otimes E_c\big)\widetilde{\boldsymbol{w}}_i \\ \big(\sum_{c\in\mathcal{S}} V_{R,c}^{\top}\otimes E_c\big)\widetilde{\boldsymbol{w}}_i\end{bmatrix} \triangleq \begin{bmatrix}\bar{\boldsymbol{w}}_i \\ \check{\boldsymbol{w}}_i\end{bmatrix} \tag{110}$$

1053-587X (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2017.2771731, IEEE Transactions on Signal Processing

and

$$\mathcal{V}_\epsilon^\top \mathcal{A}^\top \mathcal{M} \boldsymbol{v}_i^s \triangleq \begin{bmatrix} \bar{\boldsymbol{v}}_i^s \\ \check{\boldsymbol{v}}_i^s \end{bmatrix}. \tag{111}$$

Therefore, in expanded form, the transformed network weight-error recursion (109) becomes

$$\bar{\boldsymbol{w}}_i = \big(I_M - D_{i,11}^\top\big)\bar{\boldsymbol{w}}_{i-1} - D_{i,21}^\top \check{\boldsymbol{w}}_{i-1} + \bar{\boldsymbol{v}}_i^s \tag{112a}$$

$$\check{\boldsymbol{w}}_i = \big(\mathcal{J}_\epsilon^\top - D_{i,22}^\top\big)\check{\boldsymbol{w}}_{i-1} - D_{i,12}^\top \bar{\boldsymbol{w}}_{i-1} + \check{\boldsymbol{v}}_i^s \tag{112b}$$

Second-order moment stability: Evaluating the second-order moments of $\bar{\boldsymbol{w}}_i$ and $\check{\boldsymbol{w}}_i$ in (112a)–(112b), conditioned on $\boldsymbol{\mathcal{F}}_{i-1}$:

$$\mathbb{E}\Big[\|\bar{\boldsymbol{w}}_i\|^2 \Big| \boldsymbol{\mathcal{F}}_{i-1}\Big] = \big\|\big(I_M - D_{i,11}^\top\big)\bar{\boldsymbol{w}}_{i-1} - D_{i,21}^\top \check{\boldsymbol{w}}_{i-1}\big\|^2 + E\Big[\|\bar{\boldsymbol{v}}_i^s\|^2 \Big| \boldsymbol{\mathcal{F}}_{i-1}\Big] \tag{113a}$$

$$\mathbb{E}\Big[\|\check{\boldsymbol{w}}_i\|^2 \Big| \boldsymbol{\mathcal{F}}_{i-1}\Big] = \big\|\big(\mathcal{J}_\epsilon^\top - D_{i,22}^\top\big)\check{\boldsymbol{w}}_{i-1} - D_{i,12}^\top \bar{\boldsymbol{w}}_{i-1}\big\|^2 + E\Big[\|\check{\boldsymbol{v}}_i^s\|^2 \Big| \boldsymbol{\mathcal{F}}_{i-1}\Big] \tag{113b}$$

where we used (111), (79), and (76), along with the fact that the noise samples $\{\boldsymbol{v}_k(i)\}$ are zero-mean. Taking the expectation again,

$$\mathbb{E}\|\bar{\boldsymbol{w}}_i\|^2 = \mathbb{E}\big\|\big(I_M - D_{i,11}^\top\big)\bar{\boldsymbol{w}}_{i-1} - D_{i,21}^\top \check{\boldsymbol{w}}_{i-1}\big\|^2 + \mathbb{E}\|\bar{\boldsymbol{v}}_i^s\|^2 \tag{114a}$$

$$\mathbb{E}\|\check{\boldsymbol{w}}_i\|^2 = \mathbb{E}\big\|\big(\mathcal{J}_\epsilon^\top - D_{i,22}^\top\big)\check{\boldsymbol{w}}_{i-1} - D_{i,12}^\top \bar{\boldsymbol{w}}_{i-1}\big\|^2 + \mathbb{E}\|\check{\boldsymbol{v}}_i^s\|^2 \tag{114b}$$

Bounding (114a) by considering an arbitrary constant $t \in (0,1)$, invoking Jensen's inequality [40], and recalling the bounds on $\big(I_M - D_{i,11}^\top\big)$ and $D_{i,21}$,

$$\begin{aligned}\mathbb{E}\|\bar{\boldsymbol{w}}_i\|^2 &= \mathbb{E}\left\|\frac{1-t}{1-t}\big(I_M - D_{i,11}^\top\big)\bar{\boldsymbol{w}}_{i-1} - \frac{t}{t}D_{i,21}^\top \check{\boldsymbol{w}}_{i-1}\right\|^2 + \mathbb{E}\|\bar{\boldsymbol{v}}_i^s\|^2 \\ &\le \frac{\big(1-\sigma_{11}\mu_{\max}\big)^2}{1-t}\mathbb{E}\|\bar{\boldsymbol{w}}_{i-1}\|^2 + \frac{\sigma_{21}^2\mu_{\max}^2}{t}\mathbb{E}\|\check{\boldsymbol{w}}_{i-1}\|^2 + \mathbb{E}\|\bar{\boldsymbol{v}}_i^s\|^2.\end{aligned} \tag{115}$$

Choosing $t = \sigma_{11}\mu_{\max} < 1$,

$$\mathbb{E}\|\bar{\boldsymbol{w}}_i\|^2 \le (1-\sigma_{11}\mu_{\max})E\|\bar{\boldsymbol{w}}_{i-1}\|^2 + \frac{\sigma_{21}^2\mu_{\max}}{\sigma_{11}}\mathbb{E}\|\check{\boldsymbol{w}}_{i-1}\|^2 + \mathbb{E}\|\bar{\boldsymbol{v}}_i^s\|^2. \tag{116}$$

Bounding (114b) using similar reasoning,

$$\mathbb{E}\|\check{\boldsymbol{w}}_i\|^2 \le \frac{1}{t}\mathbb{E}\big\|\mathcal{J}_\epsilon^\top \check{\boldsymbol{w}}_{i-1}\big\|^2 + \mathbb{E}\|\check{\boldsymbol{v}}_i^s\|^2 + \frac{1}{1-t}\mathbb{E}\big\|D_{i,22}^\top \check{\boldsymbol{w}}_{i-1} + D_{i,12}^\top \bar{\boldsymbol{w}}_{i-1}\big\|^2. \tag{117}$$

As for the first term in the upper bound in (117), invoking the Rayleigh–Ritz eigenvalue characterization [39]: $\mathbb{E}\big\|\mathcal{J}_\epsilon^\top \check{\boldsymbol{w}}_{i-1}\big\|^2 \le \rho(\mathcal{J}_\epsilon\mathcal{J}_\epsilon^*)\mathbb{E}\|\check{\boldsymbol{w}}_{i-1}\|^2$, where $\rho(\cdot)$ denotes the spectral radius of its Hermitian matrix argument. Note that the product $\mathcal{J}_\epsilon\mathcal{J}_\epsilon^*$ can be decomposed as

$$\begin{aligned}\mathcal{J}_\epsilon\mathcal{J}_\epsilon^* &= \left(\sum_{c\in\mathcal{S}} J_{\epsilon,c}\otimes E_c\right)\left(\sum_{c\in\mathcal{S}} J_{\epsilon,c}^*\otimes E_c\right) \\ &= \sum_{c\in\mathcal{S}} J_{\epsilon,c}J_{\epsilon,c}^*\otimes E_c\end{aligned} \tag{118}$$

and that

$$\rho(\mathcal{J}_\epsilon\mathcal{J}_\epsilon^*) = \max_{c\in\mathcal{S}}\rho\big(J_{\epsilon,c}J_{\epsilon,c}^*\big) \le \max_{c\in\mathcal{S}}\big\|J_{\epsilon,c}J_{\epsilon,c}^*\big\|_1 \tag{119}$$

where $\|\cdot\|_1$ is the maximum absolute column sum. Let $c_1 \triangleq \arg\max_{c\in\mathcal{S}}\big\|J_{\epsilon,c}J_{\epsilon,c}^*\big\|_1$. Say J_{ϵ,c_1} has L_{c_1} Jordan blocks each with eigenvalue $\lambda_{c_1,\ell}$, $1 \le \ell \le L_{c_1}$. Then,

$$\rho\big(J_{\epsilon,c_1}J_{\epsilon,c_1}^*\big) \le \max_{1\le\ell\le L_{c_1}}\big(|\lambda_{c_1,\ell}| + \epsilon\big)^2 = \big(\rho(J_{\epsilon,c_1}) + \epsilon\big)^2 \tag{120}$$

So, $\rho(\mathcal{J}_\epsilon\mathcal{J}_\epsilon^*) \le \big(\rho(J_{\epsilon,c_1}) + \epsilon\big)^2$. Note that $\rho(J_{\epsilon,c}) < 1$ for all $c \in \mathcal{S}$. Hence, ϵ can be selected such that $\big(\rho(J_{\epsilon,c_1}) + \epsilon\big) \in (0,1)$. Substituting the upper bound (120) into (117),

$$\mathbb{E}\|\check{\boldsymbol{w}}_i\|^2 \le \frac{1}{t}\big(\rho(J_{\epsilon,c_1}) + \epsilon\big)^2\mathbb{E}\|\check{\boldsymbol{w}}_{i-1}\|^2 + \mathbb{E}\|\check{\boldsymbol{v}}_i^s\|^2 + \frac{1}{1-t}\mathbb{E}\big\|D_{i,22}^\top \check{\boldsymbol{w}}_{i-1} + D_{i,12}^\top \bar{\boldsymbol{w}}_{i-1}\big\|^2. \tag{121}$$

Choosing $t = \big(\rho(J_{\epsilon,c_1}) + \epsilon\big) \in (0,1)$ leads to

$$\mathbb{E}\|\check{\boldsymbol{w}}_i\|^2 \le \big(\rho(J_{\epsilon,c_1}) + \epsilon\big)\mathbb{E}\|\check{\boldsymbol{w}}_{i-1}\|^2 + \mathbb{E}\|\check{\boldsymbol{v}}_i^s\|^2 + \frac{1}{1-\rho(J_{\epsilon,c_1}) - \epsilon}\mathbb{E}\big\|D_{i,22}^\top \check{\boldsymbol{w}}_{i-1} + D_{i,12}^\top \bar{\boldsymbol{w}}_{i-1}\big\|^2. \tag{122}$$

As for the third term in the upper bound in (122),

$$\begin{aligned}&\mathbb{E}\big\|D_{i,22}^\top \check{\boldsymbol{w}}_{i-1} + D_{i,12}^\top \bar{\boldsymbol{w}}_{i-1}\big\|^2 \\ &\quad\le 2\mathbb{E}\big\|D_{i,22}^\top \check{\boldsymbol{w}}_{i-1}\big\|^2 + 2\mathbb{E}\big\|D_{i,12}^\top \bar{\boldsymbol{w}}_{i-1}\big\|^2 \\ &\quad\le 2\sigma_{22}^2\mu_{\max}^2\mathbb{E}\|\check{\boldsymbol{w}}_{i-1}\|^2 + 2\sigma_{12}^2\mu_{\max}^2\mathbb{E}\|\bar{\boldsymbol{w}}_{i-1}\|^2.\end{aligned} \tag{123}$$

Substituting the upper bound in (123) into (122),

$$\mathbb{E}\|\check{\boldsymbol{w}}_i\|^2 \le \left(\rho(J_{\epsilon,c_1}) + \epsilon + \frac{2\sigma_{22}^2\mu_{\max}^2}{1-\rho(J_{\epsilon,c_1}) - \epsilon}\right)\mathbb{E}\|\check{\boldsymbol{w}}_{i-1}\|^2 + \frac{2\sigma_{12}^2\mu_{\max}^2}{1-\rho(J_{\epsilon,c_1}) - \epsilon}\mathbb{E}\|\bar{\boldsymbol{w}}_{i-1}\|^2 + \mathbb{E}\|\check{\boldsymbol{v}}_i^s\|^2. \tag{124}$$

It remains to bound the update noise terms in (116) and (124), $\mathbb{E}\|\bar{\boldsymbol{v}}_i^s\|^2$ and $\mathbb{E}\|\check{\boldsymbol{v}}_i^s\|^2$, respectively:

$$\begin{aligned}\mathbb{E}\|\bar{\boldsymbol{v}}_i^s\|^2 + \mathbb{E}\|\check{\boldsymbol{v}}_i^s\|^2 &= \mathbb{E}\big\|\mathcal{V}_\epsilon^\top \mathcal{A}^\top \mathcal{M}\boldsymbol{v}_i^s\big\|^2 \\ &\le \big\|\mathcal{V}_\epsilon^\top \mathcal{A}^\top\big\|^2\|\mathcal{M}\|^2\mathbb{E}\|\boldsymbol{v}_i^s\|^2 \\ &\le v_1^2\mu_{\max}^2\mathbb{E}\|\boldsymbol{v}_i^s\|^2\end{aligned} \tag{125}$$

where $v_1 \triangleq \big\|\mathcal{V}_\epsilon^\top \mathcal{A}^\top\big\| > 0$, independent of $\mu_{\max}$. Note that, for all i,

$$\mathbb{E}\|\boldsymbol{v}_i^s\|^2 = \sum_{k=1}^N \mathbb{E}\big\|\boldsymbol{v}_{k,i}^s\big\|^2 \le \sum_{k=1}^N \sigma_{s,k}^2 = \sigma_s^2 \tag{126}$$

where we used (76) and (68), and $\sigma_s^2 \triangleq \sum_{k=1}^N \sigma_{s,k}^2$. Hence, referring back to (125),

$$\mathbb{E}\|\bar{\boldsymbol{v}}_i^s\|^2 + \mathbb{E}\|\check{\boldsymbol{v}}_i^s\|^2 \le v_1^2\mu_{\max}^2\sigma_s^2. \tag{127}$$

1053-587X (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2017.2771731, IEEE Transactions on Signal Processing

Substituting into the upper bounds for $E\|\bar{\boldsymbol{w}}_i\|^2$ and $E\|\check{\boldsymbol{w}}_i\|^2$ in (116) and (124), respectively,

$$\mathbb{E}\|\bar{\boldsymbol{w}}_i\|^2 \le (1-\sigma_{11}\mu_{\max})\mathbb{E}\|\bar{\boldsymbol{w}}_{i-1}\|^2 + \tfrac{\sigma_{21}^2\mu_{\max}}{\sigma_{11}}\mathbb{E}\|\check{\boldsymbol{w}}_{i-1}\|^2 + v_1^2\mu_{\max}^2\sigma_s^2 \quad (128a)$$

$$\mathbb{E}\|\check{\boldsymbol{w}}_i\|^2 \le \Big(\rho(J_{\epsilon,c_1}) + \epsilon + \tfrac{2\sigma_{22}^2\mu_{\max}^2}{1-\rho\left(J_{\epsilon,c_1}\right)-\epsilon}\Big)\mathbb{E}\|\check{\boldsymbol{w}}_{i-1}\|^2 + \tfrac{2\sigma_{12}^2\mu_{\max}^2}{1-\rho\left(J_{\epsilon,c_1}\right)-\epsilon}\mathbb{E}\|\bar{\boldsymbol{w}}_{i-1}\|^2 + v_1^2\mu_{\max}^2\sigma_s^2 \quad (128b)$$

Introducing the scalar coefficients $a = 1-\sigma_{11}\mu_{\max} = 1-\mathcal{O}(\mu_{\max})$, $b = \frac{\sigma_{21}^2\mu_{\max}}{\sigma_{11}} = \mathcal{O}(\mu_{\max})$,

$$c = \tfrac{2\sigma_{12}^2\mu_{\max}^2}{1-\rho\left(J_{\epsilon,c_1}\right)-\epsilon} = \mathcal{O}\big(\mu_{\max}^2\big)$$

$$d = \rho(J_{\epsilon,c_1}) + \epsilon + \tfrac{2\sigma_{22}^2\mu_{\max}^2}{1-\rho\left(J_{\epsilon,c_1}\right)-\epsilon} = \rho(J_{\epsilon,c_1}) + \epsilon + \mathcal{O}\big(\mu_{\max}^2\big)$$

and $e = v_1^2\mu_{\max}^2\sigma_s^2$, we can write the inequalities (128a)–(128b) more compactly as

$$\begin{bmatrix}\mathbb{E}\|\bar{\boldsymbol{w}}_i\|^2\\ \mathbb{E}\|\check{\boldsymbol{w}}_i\|^2\end{bmatrix} \preceq \underbrace{\begin{bmatrix}a & b\\ c & d\end{bmatrix}}_{\Gamma}\begin{bmatrix}\mathbb{E}\|\bar{\boldsymbol{w}}_{i-1}\|^2\\ \mathbb{E}\|\check{\boldsymbol{w}}_{i-1}\|^2\end{bmatrix} + e\mathbb{1}_2 \quad (129)$$

where $\preceq$ denotes entry-wise less-than-or-equal-to comparison; and the entries of the matrix Γ indicated above are of the form

$$\Gamma = \begin{bmatrix}1-\mathcal{O}(\mu_{\max}) & \mathcal{O}(\mu_{\max})\\ \mathcal{O}\big(\mu_{\max}^2\big) & \rho(J_{\epsilon,c_1})+\epsilon+\mathcal{O}\big(\mu_{\max}^2\big)\end{bmatrix}. \quad (130)$$

Since the spectral radius of a matrix is upper bounded by any of its norms, then, invoking this property in terms of the maximum absolute column sum, it holds that

$$\rho(\Gamma) \le \max\Big\{1-\mathcal{O}(\mu_{\max})+\mathcal{O}\big(\mu_{\max}^2\big), \rho(J_{\epsilon,c_1})+\epsilon+\mathcal{O}(\mu_{\max})+\mathcal{O}\big(\mu_{\max}^2\big)\Big\}. \quad (131)$$

Since $J_{\epsilon,c_1} < 1$ and is independent of $\mu_{\max}$, and since the latter and ϵ are small positive numbers that can be chosen arbitrarily small and independently of each other, it is clear that the upper bound above can be made strictly smaller than one for sufficiently small $\mu_{\max}$ and ϵ. If this is the case, $\rho(\Gamma) < 1$ so that Γ is a stable matrix. Given stability, if (129) is iterated, it can be concluded that

$$\limsup_{i\to\infty}\begin{bmatrix}\mathbb{E}\|\bar{\boldsymbol{w}}_i\|^2\\ \mathbb{E}\|\check{\boldsymbol{w}}_i\|^2\end{bmatrix} \preceq (I-\Gamma)^{-1}e\mathbb{1}_2 = \begin{bmatrix}\mathcal{O}(\mu_{\max})\\ \mathcal{O}\big(\mu_{\max}^2\big)\end{bmatrix}, \quad (132)$$

i.e.,

$$\limsup_{i\to\infty}\mathbb{E}\|\bar{\boldsymbol{w}}_i\|^2 = \mathcal{O}(\mu_{\max}) \quad (133)$$

$$\limsup_{i\to\infty}\mathbb{E}\|\check{\boldsymbol{w}}_i\|^2 = \mathcal{O}\big(\mu_{\max}^2\big) \quad (134)$$

This leads to

$$\limsup_{i\to\infty}\mathbb{E}\|\widetilde{\boldsymbol{w}}_i\|^2 = \limsup_{i\to\infty}\mathbb{E}\left\|\big(\mathcal{V}_\epsilon^{-1}\big)^\top\begin{bmatrix}\bar{\boldsymbol{w}}_{i-1}\\ \check{\boldsymbol{w}}_{i-1}\end{bmatrix}\right\| \le \limsup_{i\to\infty} v_2^2\Big[\mathbb{E}\|\bar{\boldsymbol{w}}_i\|^2 + \mathbb{E}\|\check{\boldsymbol{w}}_i\|^2\Big] = \mathcal{O}(\mu_{\max}) \quad (135)$$

with $v_2 \triangleq \big\|\mathcal{V}_\epsilon^{-1}\big\|$, implying that, for each k, $\limsup_{i\to\infty}\mathbb{E}\|\widetilde{\boldsymbol{w}}_{k,i}\|^2 = \mathcal{O}(\mu_{\max})$, thus concluding the proof. ■

APPENDIX B
APPROXIMATION FOR $Q_{k,c,i}$

Setting $A_{\ell k}$ to W_k, if $\ell = k$, and 0, otherwise, $K_{k,i}$ in (20) becomes

$$K_{k,i} = \begin{bmatrix}\mu_k U_i^\top E_k & Y_{k,i}K_{k,i-1}\end{bmatrix} = \begin{bmatrix}\mu_k U_i^\top E_k & \mu_k Y_{k,i}U_{i-1}^\top E_k & \dots & \mu_k(Y_{k,i}\dots Y_{k,1})U_0^\top E_k\end{bmatrix} \quad (136)$$

Note that for sufficiently small step-sizes $\{\mu_k\}$,

$$Y_{k,i}\dots Y_{k,j+1} \approx I - \mu_k\sum_{m=j+1}^{i}u_{k,m}^\top u_{k,m}. \quad (137)$$

Evaluating the first parenthesized expression in (40) and ignoring $\mathcal{O}(\mu_k^2)$ terms,

$$P_cK_{k,i}U_{0:i}P_c^\top = \mu_kP_c\left(\sum_{j=0}^{i}(Y_{k,i}\dots Y_{k,j+1})u_{k,j}^\top u_{k,j}\right)P_c^\top \approx \mu_k\sum_{j=0}^{i}u_{k,c,j}^\top u_{k,c,j}. \quad (138)$$

Ignoring $\mathcal{O}(\mu_k^2)$ terms can also be motivated by asymptotically vanishing cross-interest interference, established in Lemma 1. Evaluating the second parenthesized expression in (40) and keeping terms only up to $\mathcal{O}(\mu_k^2)$,

$$P_cK_{k,i}R_{v,0:i}K_{k,i}^\top P_c^\top = \mu_k^2\sigma_{v,k}^2P_c\Big(\sum_{j=0}^{i}(Y_{k,i}\dots Y_{k,j+1})u_{k,j}^\top u_{k,j}(Y_{k,i}\dots Y_{k,j+1})^\top\Big)P_c^\top \approx \mu_k^2\sigma_{v,k}^2\sum_{j=0}^{i}u_{k,c,j}^\top u_{k,c,j} \approx \mu_k\sigma_{v,k}^2P_cK_{k,i}U_{0:i}P_c^\top \quad (139)$$

Since constants multiplying test-statistics have no bearing on the detection performance, it can be concluded that a good approximation for $Q_{k,c,i}$ under small step-sizes $\{\mu_k\}$ is $Q_{k,c,i} = I_{M_c}$.

APPENDIX C
MEAN STABILITY OF DIFFUSION ESTIMATION ALGORITHM

Starting from (50), and with $B_i = \mathcal{A}^\top[I-\mathcal{M}\mathcal{H}_i]$, the network mean weight-error recursion is given by $\mathbb{E}\,\widetilde{\boldsymbol{w}}_i = B_i\,\mathbb{E}\,\widetilde{\boldsymbol{w}}_{i-1}$. Transforming the recursion by pre-multiplying it with $\mathcal{V}_\epsilon$ in (89), and using (110) and (108) yields

$$\underbrace{\begin{bmatrix}\mathbb{E}\,\bar{\boldsymbol{w}}_i\\ \mathbb{E}\,\check{\boldsymbol{w}}_i\end{bmatrix}}_{\triangleq z_i} = \underbrace{\begin{bmatrix}I_M - D_{i,11}^\top & -D_{i,21}^\top\\ -D_{i,12}^\top & \mathcal{J}_\epsilon^\top - D_{i,22}^\top\end{bmatrix}}_{\triangleq \bar{B}_i}\underbrace{\begin{bmatrix}\mathbb{E}\,\bar{\boldsymbol{w}}_{i-1}\\ \mathbb{E}\,\check{\boldsymbol{w}}_{i-1}\end{bmatrix}}_{\triangleq z_{i-1}}. \quad (140)$$

Let $\bar{z}_i \triangleq \mathbb{E}\,\bar{\boldsymbol{w}}_i$ and $\check{z}_i \triangleq \mathbb{E}\,\check{\boldsymbol{w}}_i$ such that $z_i = \begin{bmatrix}\bar{z}_i^\top & \check{z}_i^\top\end{bmatrix}^\top$ in (140). Following a similar procedure as in Appendix A, $\|\bar{z}_i\|^2$ and $\|\check{z}_i\|^2$ can be bounded as

$$\|\bar{z}_i\|^2 \le (1-\sigma_{11}\mu_{\max})\mathbb{E}\|\bar{z}_{i-1}\|^2 + \tfrac{\sigma_{21}^2\mu_{\max}}{\sigma_{11}}\|\check{z}_{i-1}\|^2 \quad (141a)$$

$$\|\check{z}_i\|^2 \le \Big(\rho(J_{\epsilon,c_1}) + \epsilon + \tfrac{2\sigma_{22}^2\mu_{\max}^2}{1-\rho\left(J_{\epsilon,c_1}\right)-\epsilon}\Big)\|\check{z}_{i-1}\|^2 + \tfrac{2\sigma_{12}^2\mu_{\max}^2}{1-\rho\left(J_{\epsilon,c_1}\right)-\epsilon}\|\bar{z}_{i-1}\|^2 \quad (141b)$$

1053-587X (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2017.2771731, IEEE Transactions on Signal Processing

The inequalities (141a)–(141b) can be combined into a system of inequalities of the form

$$\begin{bmatrix} \|\bar{z}_i\|^2 \\ \|\check{z}_i\|^2 \end{bmatrix} \preceq \underbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}_{\Gamma} \begin{bmatrix} \|\bar{z}_{i-1}\|^2 \\ \|\check{z}_{i-1}\|^2 \end{bmatrix} \tag{142}$$

where the scalar coefficients a, b, c, and d, and the coefficient matrix Γ are the same as in Appendix A—see (130) and the preceding discussion. If Γ is stable—see the discussion following (130), then, if (142) is iterated, it can be concluded that $\limsup_{i\to\infty}\|\bar{z}_i\|^2 = 0$ and $\limsup_{i\to\infty}\|\check{z}_i\|^2 = 0$, implying that

$$\limsup_{i\to\infty}\left\|\begin{bmatrix}\mathbb{E}\bar{\boldsymbol{w}}_i \\ \mathbb{E}\check{\boldsymbol{w}}_i\end{bmatrix}\right\| = 0. \tag{143}$$

This leads to

$$\begin{aligned}\limsup_{i\to\infty}\mathbb{E}\|\widetilde{\boldsymbol{w}}_{k,i}\| &\le \limsup_{i\to\infty}\mathbb{E}\|\widetilde{\boldsymbol{w}}_i\| \\ &= \limsup_{i\to\infty}\left\|\mathcal{V}_\epsilon^{-1}\begin{bmatrix}\mathbb{E}\bar{\boldsymbol{w}}_i \\ \mathbb{E}\check{\boldsymbol{w}}_i\end{bmatrix}\right\| \\ &\le \limsup_{i\to\infty}\|\mathcal{V}_\epsilon^{-1}\|\left\|\begin{bmatrix}\mathbb{E}\bar{\boldsymbol{w}}_i \\ \mathbb{E}\check{\boldsymbol{w}}_i\end{bmatrix}\right\| = 0,\end{aligned} \tag{144}$$

implying that, for each k, $\limsup_{i\to\infty}\|\mathbb{E}\widetilde{\boldsymbol{w}}_{k,i}\| = 0$, thus concluding the proof. ■

REFERENCES

[1] S. Chouvardas, M. Muma, K. Hamaidi, S. Theodoridis, and A. M. Zoubir, "Distributed robust labeling of audio sources in heterogeneous wireless sensor networks," in *Proc. IEEE ICASSP*, Brisbane, Australia, Apr. 2015, pp. 5783–5787.

[2] F. K. Teklehaymanot, M. Muma, B. Béjar, P. Binder, A. Zoubir, and M. Vetterli, "Robust diffusion-based unsupervised object labelling in distributed camera networks," in *Proc. AFRICON*, Addis Ababa, Ethiopia, Sep. 2015, pp. 1–6.

[3] A. H. Sayed, *Adaptive Filters*. NJ: Wiley, 2008.

[4] ——, "Adaptation, learning, and optimization over networks," in *Foundations and Trends in Machine Learning*, vol. 7, no. 4-5, 2014, pp. 311–801.

[5] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Autom. Control*, vol. 31, no. 9, pp. 803–812, Sep. 1986.

[6] P. Braca, S. Marano, and V. Matta, "Enforcing consensus while monitoring the environment in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3375–3380, Jul. 2008.

[7] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Trans. Autom. Control*, vol. 54, no. 1, pp. 48–61, Jan. 2009.

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

[9] F. S. Cattivelli and A. H. Sayed, "Diffusion LMS strategies for distributed estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1035–1048, Mar. 2010.

[10] X. Zhao and A. H. Sayed, "Performance limits for distributed estimation over LMS adaptive networks," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5107–5124, Oct. 2012.

[11] 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.

[12] A. H. Sayed, "Diffusion adaptation over networks," in *Academic Press Library in Signal Processing*, R. Chellapa and S. Theodoridis, Eds., vol. 3. Academic Press, Elsevier, 2014, pp. 323–454.

[13] S.-Y. Tu and A. H. Sayed, "Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6217–6234, Dec. 2012.

[14] J. Chen and A. H. Sayed, "On the learning behavior of adaptive networks—Part I: Transient analysis," *IEEE Trans. Inform. Theory*, vol. 61, no. 6, pp. 3487–3517, Jun. 2015.

[15] ——, "On the learning behavior of adaptive networks—Part II: Performance analysis," *IEEE Trans. Inform. Theory*, vol. 61, no. 6, pp. 3518–3548, Jun. 2015.

[16] P. Braca, S. Marano, V. Matta, and P. Willett, "Asymptotic optimality of running consensus in testing binary hypotheses," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 814–825, Feb. 2010.

[17] D. Bajović, D. Jakovetić, J. Xavier, B. Sinopoli, and J. M. F. Moura, "Distributed detection via Gaussian running consensus: Large deviations asymptotic analysis," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4381–4396, Sep. 2011.

[18] D. Jakovetić, J. M. F. Moura, and J. Xavier, "Distributed detection over noisy networks: Large deviations analysis," *IEEE Trans. on Signal Process.*, vol. 60, no. 8, pp. 4306–4320, Aug. 2012.

[19] D. Bajović, D. Jakovetić, J. M. F. Moura, J. Xavier, and B. Sinopoli, "Large deviations performance of consensus+innovations distributed detection with non-Gaussian observations," *IEEE Trans. Signal Process.*, vol. 60, no. 11, Nov. 2012.

[20] A. K. Sahu and S. Kar, "Recursive distributed detection for composite hypothesis testing: Nonlinear observation models in additive Gaussian noise," *IEEE Trans. Inform. Theory*, vol. 63, no. 8, pp. 4797–4828, Aug. 2017.

[21] F. S. Cattivelli and A. H. Sayed, "Distributed detection over adaptive networks using diffusion adaptation," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 1917–1932, May 2011.

[22] S. Al-Sayed, A. M. Zoubir, and A. H. Sayed, "Robust distributed detection over adaptive diffusion networks," in *Proc. IEEE ICASSP*, Florence, Italy, May 2014, pp. 7233–7237.

[23] S. Al-Sayed, "Robust adaptation and learning over networks," Ph.D. dissertation, Technische Universität Darmstadt, Darmstadt, Germany, 2016.

[24] V. Matta, P. Braca, S. Marano, and A. H. Sayed, "Diffusion-based adaptive distributed detection: Steady-state performance in the slow adaptation regime," *IEEE Trans. Inform. Theory*, vol. 62, no. 8, pp. 4710–4732, Aug. 2016.

[25] ——, "Distributed detection over adaptive networks: Refined asymptotics and the role of connectivity," *Trans. Signal Inf. Process. Netw.*, vol. 2, no. 4, pp. 442–460, Dec. 2016.

[26] A. Bertrand and M. Moonen, "Distributed adaptive node-specific signal estimation in fully connected sensor networks—Part I: Sequential node updating," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5277–5291, Oct. 2010.

[27] ——, "Distributed adaptive estimation of node-specific signals in wireless sensor networks with a tree topology," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2196–2210, May 2011.

[28] J. Szurley, A. Bertrand, and M. Moonen, "Distributed adaptive node-specific signal estimation in heterogeneous and mixed-topology wireless sensor networks," *Signal Processing*, vol. 117, pp. 44–60, Dec. 2015.

[29] A. Hassani, A. Bertrand, and M. Moonen, "GEVD-based low-rank approximation for distributed adaptive node-specific signal estimation in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 64, no. 10, pp. 2557–2572, May 2016.

[30] N. Bogdanović, J. Plata-Chaves, and K. Berberidis, "Distributed incremental-based LMS for node-specific adaptive parameter estimation," *IEEE Trans. Signal Process.*, vol. 62, no. 20, pp. 5382–5397, Oct. 2014.

[31] J. Plata-Chaves, N. Bogdanović, 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.

[32] J. Plata-Chaves, M. H. Bahari, M. Moonen, and A. Bertrand, "Unsupervised diffusion-based LMS for node-specific parameter estimation over wireless sensor networks," in *Proc. IEEE ICASSP*, Shanghai, China, Mar. 2016, pp. 4159–4163.

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

[34] ——, "Diffusion LMS over multitask networks," *IEEE Trans. Signal Process.*, vol. 63, no. 11, pp. 2733–2748, Jun. 2015.

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

[36] ——, "Proximal multitask learning over networks with sparsity-inducing coregularization," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6329–6344, Dec. 2016.

[37] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. NJ: Prentice Hall, 1998.

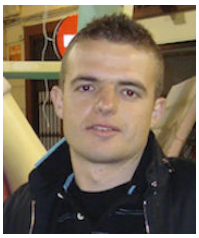
1053-587X (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2017.2771731, IEEE Transactions on Signal Processing

[38] S. Al-Sayed, A. M. Zoubir, and A. H. Sayed, "Robust distributed estimation by networked agents," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 3909–3921, Aug. 2017.
[39] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge: Cambridge University Press, 2003.
[40] S. Boyd and L. Vanderberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.



Sara Al-Sayed (S'06-M'17) received the B.Sc. degree from the German University in Cairo, New Cairo, Egypt, in 2009, the M.Sc. degree from Universität Ulm, Ulm, Germany, in 2010, and the Ph.D. degree from Technische Universität Darmstadt, Darmstadt, Germany, in 2016, all in electrical engineering. From 2011 to 2015, she was a Research Associate at the Signal Processing Group, Technische Universität Darmstadt. She is currently a Postdoctoral Researcher at the Bioinspired Communication Systems Group, Technische Universität Darmstadt. Her research interests include adaptive filtering, adaptation and learning over networks, and distributed statistical signal processing.



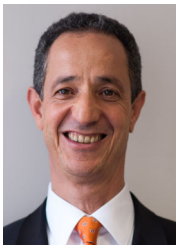
Jorge Plata-Chaves (S'09-M'13) was born in Madrid, Spain, in 1984. He received the B.Sc., M.Sc., and Ph.D. degrees from the Universidad Carlos III de Madrid, Spain, in 2007, 2009, and 2012, respectively, all in electrical engineering, with a specialization in Multimedia and Communications. From 2007 to 2012, he was with the Department of Signal Theory and Communications, Universidad de Carlos III de Madrid, Spain. From 2012 to 2013, he was an ER Marie Curie fellow at the Research Academic Computer Technology Institute (RACTI) in Patras, Greece. He is currently working as a postdoctoral fellow with the Electrical Engineering Department (ESAT), KU Leuven. His research interests are in information theory and statistical signal processing with application to interference management in cellular networks as well as detection, estimation, and localization in wireless mobile sensor networks and smart grids.



Michael Muma (S'10-M'14) received the Dipl.-Ing. (2009) and the Dr.-Ing. degree with summa cum laude in Electrical Engineering and Information Technology (2014), both from Technische Universität Darmstadt, Darmstadt, Germany. He completed his diploma thesis with the Contact Lens and Visual Optics Laboratory, School of Optometry, Brisbane, Australia, on the role of cardiopulmonary signals in the dynamics of the eye's wavefront aberrations. His PhD thesis was on "Robust Estimation and Model Order Selection for Signal Processing". Currently, he is a Postdoctoral Fellow at the Signal Processing Group, Institute of Telecommunications and Athene Young Investigator of Technische Universität Darmstadt. His research is on robust statistics for signal processing, biomedical signal processing, and robust distributed detection, classification, and estimation for wireless sensor networks. Since 2015, he is an Associate Editor for the IEEE Signal Processing eNewsletter representing the Europe region. Dr. Muma was the supervisor of the Technische Universität Darmstadt student team that won the international IEEE Signal Processing Cup 2015, the competition topic was "Heart Rate Monitoring During Physical Exercise Using Wrist-Type Photoplethysmographic (PPG) Signals". Dr. Muma co-organized the 2016 Joint IEEE SPS and EURASIP Summer School on Robust Signal Processing. He organized a Special Session on "Cooperative and Distributed Algorithms for Signal Processing and Self-Organization over Wireless Ad-Hoc/Sensor Networks" at the 2017 European Signal Processing Conference (EUSIPCO).



Marc Moonen (M'94-SM'06-F'07) is a Full Professor at the Electrical Engineering Department of KU Leuven, where he is heading a research team working in the area of numerical algorithms and signal processing for digital communications, wireless communications, DSL, and audio signal processing. He received the 1994 KU Leuven Research Council Award, the 1997 Alcatel Bell (Belgium) Award (with Piet Vandaele), the 2004 Alcatel Bell (Belgium) Award (with Raphael Cendrillon), and was a 1997 Laureate of the Belgium Royal Academy of Science. He received journal best paper awards from the IEEE Transactions on Signal Processing (with Geert Leus and with Daniele Giacobello) and from Elsevier Signal Processing (with Simon Doclo). He was chairman of the IEEE Benelux Signal Processing Chapter (1998–2002), a member of the IEEE Signal Processing Society Technical Committee on Signal Processing for Communications, and President of EURASIP (European Association for Signal Processing, 2007–2008 and 2011–2012). He has served as Editor-in-Chief for the EURASIP Journal on Applied Signal Processing (2003–2005), Area Editor for Feature Articles in IEEE Signal Processing Magazine (2012–2014), and has been a member of the editorial board of IEEE Transactions on Circuits and Systems II, IEEE Signal Processing Magazine, Integration-the VLSI Journal, EURASIP Journal on Wireless Communications and Networking, and Signal Processing. He is currently a member of the editorial board of EURASIP Journal on Advances in Signal Processing.



Abdelhak M. Zoubir (S'87-M'92-SM'97-F'08) is a Fellow of the IEEE and IEEE Distinguished Lecturer (Class 2010–2011). He received his Dr.-Ing. from Ruhr-Universität Bochum, Bochum, Germany, in 1992. He was with Queensland University of Technology, Brisbane, Qld., Australia, from 1992 to 1998 where he was an Associate Professor. In 1999, he joined Curtin University of Technology, Perth, WA, Australia, as a Professor of Telecommunications and was Interim Head of the School of Electrical & Computer Engineering from 2001 until 2003. In 2003, he moved to Technische Universität Darmstadt, Darmstadt, Germany, as Professor of Signal Processing and Head of the Signal Processing Group. His research interest lies in statistical methods for signal processing with emphasis on bootstrap techniques, robust detection and estimation, and array processing applied to telecommunications, radar, sonar, automotive monitoring and safety, and biomedicine. He published over 400 journal and conference papers in these areas. Dr. Zoubir acted as General or Technical Chair of numerous conferences and workshops. Notably, he was the Technical Co-Chair of ICASSP-14 held in May 2014 in Florence, Italy. Dr. Zoubir has also held several positions in editorial boards; most notably, he was the Editor-in-Chief of the IEEE Signal Processing Magazine (2012–2014). He was elected as Chair (2010–2011), Vice-Chair (2008–2009), and Member (2002–2008) of the IEEE SPS Technical Committee Signal Processing Theory and Methods (SPTM), and a Member (2007–2012) of the IEEE SPS Technical Committee Sensor Array and Multi-channel Signal Processing (SAM). He currently serves on the Board of Governors of the IEEE SPS as an elected Member-at-Large (2015–2017), and has been a Member of the Board of Directors of the European Association of Signal Processing (EURASIP) since 2008. He is the current President of EURASIP (2017–2019).

1053-587X (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.