

ROBUST SET-THEORETIC DISTRIBUTED DETECTION IN DIFFUSION NETWORKS

R. L. G. Cavalcante and S. Stanczak

Fraunhofer Heinrich Hertz Institute
Einsteinufer 37, 10587 Berlin, Germany
email: {renato.cavalcante, slawomir.stanczak}@hhi.fraunhofer.de

ABSTRACT

We propose novel set-theoretic distributed adaptive filters for cooperative signal detection in diffusion networks, a problem that has been gaining attention owing to its application to cooperative cognitive radio networks. In the proposed method, nodes in a network detect the presence of a signal of interest by means of an inner product between the current term of a series and a known reference vector. Each term of the series is computed from information fusion among neighboring nodes and projections onto closed convex sets, which are constructed with *a priori* knowledge of the signal of interest and measurements obtained by nodes. In particular, we show that sets based on *a priori* knowledge are useful to decrease the communication overhead and to provide good detection performance. Our results are rigorous in the sense that no approximations are used to prove convergence properties. In particular, we show conditions to guarantee that the series converge to a point that can reliably identify the signal of interest. Furthermore, we also show that recent results in distributed optimization for dynamic systems can be used to derive algorithms where nodes exchange not only the current vectors of their sequences (as in previous distributed set-theoretic filters), but also side information that influences the above-mentioned sets.

Index Terms— Distributed set-theoretic adaptive filtering, cognitive radios, hypothesis testing

1. INTRODUCTION

Algorithms for robust signal detection have been increasingly gaining importance owing to recent developments in wireless systems. For example, in many proposed cognitive radio systems, secondary users (unlicensed users) have to detect from local measurements the presence of primary users (licensed users), with the intent to increase data rate by using licensed frequency bands in periods of inactivity of licensees. Algorithms for such a purpose are commonly known as spectrum sensing algorithms, which have been typically based on standard signal detection schemes (e.g., correlators, energy detectors, etc.). Unfortunately, in wireless communication systems, conventional signal detection algorithms may have poor performance owing to local attenuation of signals caused by deep fading.

To improve the robustness of detection algorithms, researchers have been studying cooperative detection mechanisms [1,2]. In these schemes, nodes in a network take local measurements of the environment and exchange information with local neighbors to reach a reliable decision of whether a signal is present. In doing so, the resulting algorithms provide robustness by spatial diversity. To date, cooperative signal detection algorithms are typically based on consensus algorithms or on algorithms that solve (non-adaptive) distributed optimization problems [1]. However, more recently the study in [2] has shown that adaptive methods may be preferable because, among other benefits, they can process measurements in real time. These algorithms are based on distributed adaptive filters (e.g., distributed

versions of the LMS or RLS algorithm), and they use standard assumptions (e.g., Gaussian noise) to make the analysis tractable. Unfortunately, if different assumptions are used or the update rule is modified to use *a priori* knowledge of the signal of interest, convergence analyses using ideas similar to those in [2] can become intractable.

We address the above shortcomings by using the results in [3,4] to derive novel distributed solutions. The objective of the proposed algorithm is to produce, in every node, a sequence of vectors (filters) that converge to a point in the intersection of a possibly infinite number of closed convex sets that are dispersed in both time and space in the network. As in [2], the presence of the signal of interest can be verified in every node by means of an inner product between the current filter and a reference vector. As common in set-theoretic adaptive filtering [5], sets are constructed in an online fashion from *a priori* knowledge of the desired signal and local measurements of the nodes. Unlike previous adaptive detection methods, the proposed scheme can easily use *a priori* knowledge to decrease the communication overhead and to provide good detection performance. Furthermore, we make only few assumptions concerning noise samples and regressors to prove convergence, and the analysis does not appeal to any approximations. For example, we only assume that bounds of the noise samples are known, but the samples can be drawn from any distribution satisfying this requirement. We also show by simulations that the algorithm is robust in scenarios where these few assumptions fail. In particular, here we show the probability of misdetection and false alarm in a setting where noise does not follow a Gaussian distribution. An additional contribution of this paper is to show that, in algorithms derived from [3,4], nodes can exchange side-information that influences sets considered by their neighbors.

2. PROBLEM FORMULATION

We start by reviewing basic concepts in convex analysis. In more detail, a set C is said to be *convex* if $\mathbf{v} = \nu \mathbf{v}_1 + (1 - \nu) \mathbf{v}_2 \in C$ for every $\mathbf{v}_1, \mathbf{v}_2 \in C$ and $0 \leq \nu \leq 1$. If, in addition to being convex, C contains all its boundary points, then C is a *closed convex set*. The *metric projection* $P_C : \mathbb{R}^N \rightarrow C$ of a closed convex set C maps $\mathbf{v} \in \mathbb{R}^N$ to the uniquely existing vector $P_C(\mathbf{v}) \in C$ satisfying $\|\mathbf{v} - P_C(\mathbf{v})\| = \min_{\mathbf{y} \in C} \|\mathbf{v} - \mathbf{y}\| =: d(\mathbf{v}, C)$, where $\|\cdot\|$ is the norm induced by the inner product $\langle \mathbf{v}, \mathbf{y} \rangle := \mathbf{v}^T \mathbf{y}$ for every $\mathbf{v}, \mathbf{y} \in \mathbb{R}^N$.

We now turn to the problem formulation. We consider a network with N nodes (cognitive radio networks as a particular case), which are represented by a graph $\mathcal{G} := (\mathcal{N}, \mathcal{E})$, where $\mathcal{N} := \{1, \dots, N\}$ is the node set, and $\mathcal{E} \subset \mathcal{N} \times \mathcal{N}$ is the edge set. An edge $(k, l) \in \mathcal{E}$ indicates that, if required by the communication protocol, node k is able to send information to node l directly (we assume (k, k) belongs to \mathcal{E} for every $k \in \mathcal{N}$). To simplify the exposition, we also assume that $(k, l) \in \mathcal{E}$ if and only if $(l, k) \in \mathcal{E}$. Inward neighbors of node k are denoted by $\mathcal{N}_k := \{l \in \mathcal{N} \mid (l, k) \in \mathcal{E}\}$.

In the problem addressed here, as in [2], node k obtains measurements of the form $\mathbb{R} \ni d_k[i] = \mathbf{u}_k[i]^T \mathbf{w}^\circ + v_k[i]$, where i is the time index, $\mathbf{u}_k[i] \in \mathbb{R}^M$ is a known regression vector, $\mathbf{w}^\circ \in \mathbb{R}^M$ is an unknown vector, and $v_k[i] \in \mathbb{R}$ is the noise measurement. We assume that there are two hypotheses for the vector $\mathbf{w}^\circ \in \mathbb{R}^M$:

$$\mathbf{w}^\circ = \begin{cases} \mathbf{0}, & \text{under } \mathcal{H}_0 \\ \mathbf{w}_s, & \text{under } \mathcal{H}_1 \end{cases}. \quad (1)$$

The objective of the proposed algorithm is to distinguish, in every node $k \in \mathcal{N}$, whether hypothesis \mathcal{H}_0 or \mathcal{H}_1 is true from the measurements $d_k[i]$. Note that spectrum sensing in cognitive radio networks is a particular case of this problem [2, Sect. V-B]. (Note: extending the proposed algorithm to complex baseband signals is straightforward.)

With the settings above, the active hypothesis can be identified with distributed adaptive filters [2]. In this paradigm, each node produces at every time instant i an estimate $\mathbf{w}_k[i]$ of \mathbf{w}° , and nodes evaluate the active hypothesis with the following test:

$$\mathbf{w}_s^T \mathbf{w}_k[i] \stackrel{\mathcal{H}_0}{\underset{\mathcal{H}_1}{\gtrless}} \gamma_k[i], \quad (2)$$

where $\gamma_k[i] \in \mathbb{R}$ is the decision threshold for node k at time i .

3. PROPOSED ALGORITHM

This study is based on the premise that, to have good detection performance, we need a good estimate $\mathbf{w}_k[i]$ of \mathbf{w}° . To this end, we use the results in [3,4] to develop novel distributed set-theoretic adaptive filters. To simplify the exposition, we first present the basic idea of the algorithm, and then we proceed with the mathematical details.

3.1. Overview of the proposed set-theoretic adaptive filter

In the set-theoretic paradigm, at time i every node k builds closed convex sets that are likely to contain the *estimandum* \mathbf{w}° . Nodes also build additional sets to incorporate any *a priori* knowledge. In the discussion that follows, sets constructed from measurements are denoted by $C_k[i] \ni \mathbf{w}^\circ$, and the set corresponding to available *a priori* knowledge is denoted by $C(\epsilon) \ni \mathbf{w}^\circ$, where ϵ is a design parameter that is described later.¹ If all sets are appropriately chosen, a reasonable estimate of \mathbf{w}° is any vector that lies in the intersection of as many sets as possible. Ideally, the estimate should belong to all sets built by different nodes; i.e., belong to $\Omega^* := \bigcap_{i=0}^{\infty} \kappa[i] \ni \mathbf{w}^\circ$, where $\kappa[i] := (\bigcap_{k \in \mathcal{N}_k} C_k[i]) \cap C(\epsilon)$. Note that the set Ω^* corresponds to all information we could possibly have about \mathbf{w}° from the constructed sets, and in the set-theoretic paradigm we assume that any two points $\mathbf{p}_1, \mathbf{p}_2 \in \Omega^*$ are equally good estimates of \mathbf{w}° ; i.e., we do not have enough information to prefer \mathbf{p}_1 over \mathbf{p}_2 . Unfortunately, obtaining a point in Ω^* is often impractical because, among other problems, i) sets $C_k[i]$ are constructed as information arrives, ii) data is dispersed in a network with sparse communication, and iii) nodes have limited memory. However, by imposing reasonable assumptions, we can derive a low-complexity algorithm that, loosely speaking, produces sequences of estimates $\mathbf{w}_k[i]$ that converge asymptotically to a point that belongs to all but finitely many membership sets of \mathbf{w}° . More precisely, the algorithm can produce sequences of estimates $\mathbf{w}_k[i]$ that converge to a point in

$$\Omega := \overline{\liminf_{i \rightarrow \infty} \kappa[i]} = \overline{\bigcup_{i=0}^{\infty} \bigcap_{n \geq i} \kappa[n]} \ni \mathbf{w}^\circ, \quad (3)$$

¹For simplicity, we assume that all nodes have the same *a priori* information; the techniques presented here extend to more general cases where nodes have different *a priori* information.

where the overline operator denotes the closure of a set. To derive such an algorithm, we use the results in [3,4] by proceeding as follows. We first define membership sets $C_k[i]$ and $C(\epsilon)$ that are expected to produce a set Ω where all points are good estimates of \mathbf{w}° (Sect. 3.2). Then we recast the feasibility problem as a time-varying convex optimization problem (Sect. 3.3). In doing so, by also assigning appropriate weights to the edges of the graph (Sect. 3.4), we can use the results in [3,4] to produce sequences $\mathbf{w}_k[i]$ with the required properties (Sect. 3.5).

3.2. Membership sets of \mathbf{w}°

To build membership sets $C_k[i] \ni \mathbf{w}^\circ$ based on measurements $d_k[i]$, we use the following assumption, which is discarded later in Sect. 4.

Assumption 1. *The sequences of noise samples $v_k[i]$ are bounded above and below, and the bounds are known; i.e., there exist known bounds $l, u \in \mathbb{R}$ such that $l \leq v_k[i] \leq u$ for every $i \in \mathbb{N}$ and $k \in \mathcal{N}$.²*

Therefore, given measurements $d_k[0], \dots, d_k[i]$ at node k , we know that \mathbf{w}° belongs to the following local sets

$$L_k[i] := \left\{ \mathbf{w} \in \mathbb{R}^M \mid l \leq \bar{d}_k[i] - \mathbf{w}^T \bar{\mathbf{u}}_k[i] \leq u \right\} \ni \mathbf{w}^\circ, \quad (4)$$

where $\bar{d}_k[i] := \eta[i]^{-1} \sum_{j=i-\eta[i]+1}^i d_k[j]$, $\bar{\mathbf{u}}_k[i] := \eta[i]^{-1} \sum_{j=i-\eta[i]+1}^i \mathbf{u}_k[j]$, $\eta[i] := \min\{m, i+1\}$, and the parameter $m \geq 1$ defines the memory of the algorithm (the maximum number of samples $d_k[i]$ considered at every time i). This last parameter trades the detection performance at steady state against the tracking speed in nonstationary environments. More precisely, if \mathbf{w}° changes, the algorithm requires m samples of $d_k[i]$ to “forget” the previous value of \mathbf{w}° . Note that, to construct $L_k[i]$, we are computing averages of possibly noisy measurements $d_k[i]$, which can increase the reliability of $L_k[i]$ when Assumption 1 is not valid. For example, if noise samples $v_k[i]$ are i.i.d., with mild additional assumptions, the noise variance decreases by increasing m , which increases the probability that the relation $\mathbf{w}^\circ \in L_k[i]$ is valid. As a result, the performance of the algorithm improves.

The set $L_k[i]$ is local information obtained by samples $d_k[i]$, but node k can also obtain information about $L_j[i]$ of the local neighbors $j \in \mathcal{N}_k$ by simply obtaining the averages $\bar{d}_j[i]$ and $\bar{\mathbf{u}}_j[i]$. We can incorporate this readily available information into the algorithm by using $C_k[i] := \bigcap_{k \in \mathcal{N}_k} L_k[i] \ni \mathbf{w}^\circ$ as the set described in Sect. 3.1.

We now define $C(\epsilon) \ni \mathbf{w}^\circ$, which is the set based on *a priori* knowledge. In particular, here we use

$$C(\epsilon) := \left\{ \mathbf{w} \in \mathbb{R}^M \mid \|\mathbf{w} - \alpha \mathbf{w}_s\| \leq \epsilon \text{ for some } \alpha \in [0, 1] \right\}, \quad (5)$$

where $\epsilon \geq 0$ is a design parameter used to simplify the convergence analysis. Note that the relation $\mathbf{w}^\circ \in C(\epsilon)$ is valid even with $\epsilon = 0$. This property is useful to decrease the communication overhead among nodes because all vectors in $C(0)$ are of the form $\alpha \mathbf{w}_s$, where \mathbf{w}_s is known. As a result, if we guarantee that the estimates $\mathbf{w}_k[i]$ of \mathbf{w}° belong to $C(0)$, then nodes can exchange estimates (which, as seen later, is a requirement of the proposed algorithm) by simply sending scalars instead of vectors of dimension M . Unfortunately, by using $\epsilon = 0$, we cannot use directly the results in [4] to prove that the sequences of estimates $\mathbf{w}_k[i]$ produced by proposed algorithm

²Contrasting with existing work, we do not assume that $v_k[i]$ is a sample of a random variable with a particular distribution (e.g., Gaussian). For the proposed algorithm to work well in practice, l and u should be sufficiently small.

converge to a point in the set Ω defined in (3). However, we can still prove that the estimates of nodes, when considered together, improves at every iteration (c.f. Proposition 1).

3.3. The estimation task as a time-varying optimization problem

To derive the proposed algorithm with the scheme in [3,4], we now recast the feasibility problem described in Sect. 3.1 with the sets defined in Sect. 3.2 as a time-varying optimization problem having the sought estimates as time-invariant solutions. In more detail, we define a sequence of optimization problems indexed by i having the following form:

$$\begin{aligned} \min. \quad & \sum_{k \in \mathcal{N}} \Theta_k[i](\mathbf{w}_k), \\ \text{s.t.} \quad & \mathbf{w}_1 = \dots = \mathbf{w}_N \end{aligned} \quad (6)$$

where $\Theta_k[i] : \mathbb{R}^M \rightarrow [0, \infty)$ is a local convex function of node k attaining the value 0 only at points in $C_k[n_i]$ for i even or at points in $C(\epsilon)$ for i odd, and $n_i := \lfloor i/2 \rfloor$ ($\lfloor \cdot \rfloor$ denotes the floor function). By constructing the cost functions in this way, the set of time-invariant solutions of (6) (solutions that do not depend on the index i) can be given in terms of the ideal set of estimates Ω^* ; the time-invariant solutions are $\mathbf{w}_1, \dots, \mathbf{w}_N \in \Omega^*$ with $\mathbf{w}_1 = \dots = \mathbf{w}_N$. In particular, here we use the following functions, which have the desired characteristics and often give rise to fast and low-complexity set-theoretic adaptive filters [5]:

$$\Theta_k[i](\mathbf{w}) = \begin{cases} \sum_{j \in \mathcal{N}_k} c_{k,j}[i] \|\mathbf{w} - P_{L_j[n_i]}(\mathbf{w})\|, & i \text{ even,} \\ \|\mathbf{w} - P_{C(\epsilon)}(\mathbf{w})\|, & i \text{ odd,} \end{cases} \quad (7)$$

where $c_{k,j}[i]$ is the weight given by

$$c_{k,j}[i] = \begin{cases} \frac{\omega_{k,j}[i]}{M_k[i]} \|\mathbf{w}_k[i] - P_{L_j[n_i]}(\mathbf{w}_k[i])\| & \text{if } M_k[i] \neq 0, \\ 0 & \text{otherwise,} \end{cases}$$

$M_k[i]$ is given by $M_k[i] := \sum_{j \in \mathcal{N}_k} \omega_{k,j}[i] \|\mathbf{w}_k[i] - P_{L_j[n_i]}(\mathbf{w}_k[i])\|$, $\mathbf{w}_k[i]$ is the current estimate of \mathbf{w}^o at iteration i and at node k , and $\omega_{k,j}[i] > 0$ is the weight that node k gives to the set $L_j[n_i]$. The weights $\omega_{k,j}[i]$ should also satisfy $\sum_{j \in \mathcal{N}_k} \omega_{k,j}[i] = 1$. The projections onto the sets $L_k[i]$ and $C(\epsilon)$ are easy to compute [6], but we omit the details owing to the space limitation.

3.4. Weights assigned to the edges of the graph

To find a solution to all but finitely many optimization problems in (6), the iterative approach in [3,4] also requires possibly random matrices $\mathbf{A}_{k,j}[i]$ ($k, j \in \mathcal{N}$) taking values on $\mathbb{R}^{M \times M}$, each of which represents the weight that node k assigns to the edge (j, k) at iteration i ($\mathbf{A}_{k,j}[i] = \mathbf{0}$ if $(j, k) \notin \mathcal{E}$). Furthermore, the block matrix

$$\mathbf{P}[i] = \begin{bmatrix} \mathbf{A}_{1,1}[i] & \cdots & \mathbf{A}_{1,N}[i] \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{N,1}[i] & \cdots & \mathbf{A}_{N,N}[i] \end{bmatrix}, \quad i = 0, 1, \dots,$$

which takes values on $\mathbb{R}^{MN \times MN}$, should satisfy the following properties for every i [4, Theorem 2]:

1. $\|E[\mathbf{P}[i]^T \mathbf{P}[i]]\|_2 = 1$;
2. $\mathbf{P}[i]\mathbf{v} = \mathbf{v}$ for every $\mathbf{v} \in \mathcal{C} \subset \mathbb{R}^{MN}$;

where \mathcal{C} is the consensus subspace $\mathcal{C} := \text{span}\{\mathbf{b}_1, \dots, \mathbf{b}_M\}$, $\mathbf{b}_k = (\mathbf{1}_N \otimes \mathbf{e}_k)/\sqrt{N} \in \mathbb{R}^{MN}$, $\mathbf{1}_N \in \mathbb{R}^N$ is the vector of ones, $\mathbf{e}_k \in$

\mathbb{R}^M is the k th standard basis vector, and \otimes denotes the Kronecker product. Moreover, there should exist $I \in \mathbb{N}$ and $\epsilon_1 > 0$, such that, for every $p \in \mathbb{N}$, we can always find a matrix $\mathbf{P}[i]$ satisfying $\|E[\mathbf{P}[i]^T(\mathbf{I} - \mathbf{J})\mathbf{P}[i]]\|_2 \leq (1 - \epsilon_1)$ for some $i \in [p, p+I]$, where $\mathbf{J} \in \mathbb{R}^{MN \times MN}$ is the orthogonal projection matrix onto \mathcal{C} .

Fortunately, constructing matrices with the above properties is easy. Indeed, as shown in [4], we can use most existing consensus algorithms for this task. In particular, to simplify the exposition, we use here the following deterministic matrices:

$$\mathbf{A}_{k,j}[i] = \begin{cases} \mathbf{0} & \text{if } (j, k) \notin \mathcal{E} \text{ and } i \text{ even,} \\ \mathbf{I} & \text{if } j = k \text{ and } i \text{ even,} \\ a_{k,j}\mathbf{I} & \text{otherwise,} \end{cases} \quad (8)$$

where $a_{k,j}$ are the Metropolis-Hastings weights

$$a_{k,j} = \begin{cases} 1/\max\{g_k, g_j\}, & \text{if } k \neq j \text{ and } (j, k) \in \mathcal{E} \\ 1 - \sum_{j \in \mathcal{N}_k \setminus \{k\}} 1/\max\{g_k, g_j\}, & \text{if } k = j, \\ 0, & \text{otherwise,} \end{cases}$$

and $g_k = |\mathcal{N}_k|$ denotes the degree of node k in the graph \mathcal{G} .

3.5. The set-theoretic adaptive filter

Substituting the functions in (7) and $\mathbf{P}[i]$ (composed of the submatrices $\mathbf{A}_{k,j}[i]$ in (8)) into [4, Eq. (11)], we obtain the following algorithm:³

Algorithm 1. 1. *Initialization:* Set i to 0 and estimates $\mathbf{w}_k[i]$ to arbitrary values in \mathbb{R}^M . Select m , the memory of the algorithms, and the parameter ϵ for the set $C(\epsilon)$. In addition, select two arbitrarily small constants $\epsilon_2, \epsilon_3 \in \mathbb{R}$.

2. *Node k chooses* $\omega_{k,j}[i] > 0$ satisfying $\sum_{j \in \mathcal{N}_k} \omega_{k,j}[i] = 1$.

3. *For all* $k \in \mathcal{N}$, *update the estimates according to* $\mathbf{w}_k[i+1] = \sum_{j \in \mathcal{N}_k} a_{k,j} P_{C(\epsilon)}(\mathbf{w}'_j[i])$, *where*

$$\mathbf{w}'_k[i] = \mathbf{w}_k[i] + \mu_k[i] \left(\sum_{j \in \mathcal{N}_k} \omega_{k,j}[i] P_{L_j[n_i]}(\mathbf{w}_k[i]) - \mathbf{w}_k[i] \right),$$

$\mu_k[i] \in [\epsilon_2, (2 - \epsilon_3)\lambda_k[i]]$ *is an arbitrarily chosen step size, and*

$$1 \leq \lambda_k[i] := \begin{cases} \frac{\sum_{j \in \mathcal{N}_k} \omega_{k,j}[i] \|P_{L_j[n_i]}(\mathbf{w}_k[i]) - \mathbf{w}_k[i]\|^2}{\left\| \sum_{j \in \mathcal{N}_k} \omega_{k,j}[i] P_{L_j[n_i]}(\mathbf{w}_k[i]) - \mathbf{w}_k[i] \right\|^2} \\ 1, & \text{if } \mathbf{w}_k[i] \notin \bigcap_{j \in \mathcal{N}_k} L_j[n_i] \\ & \text{otherwise,} \end{cases}$$

4. *In every node k , decide whether the signal is present by using the scheme in (2) (with the new estimate* $\mathbf{w}_k[i+1]$ *).*

5. *Increment i and go to step 2.*

The theorem below shows sufficient conditions to guarantee that Algorithm 1 has the desired properties described in Sect. 3.1.

Proposition 1. *Let Assumption 1 be valid. By Defining* $\psi[i] := [\mathbf{w}_1^T[i] \cdots \mathbf{w}_N^T[i]]^T$ *and* $\psi^* := [(\mathbf{w}^o)^T \cdots (\mathbf{w}^o)^T]^T$, *the following holds:*

1. *At every iteration the network performance improves in the sense that* $\|\psi[i+1] - \psi^*\| \leq \|\psi[i] - \psi^*\|$.

³Here we use a unity step size when using projections onto $C(\epsilon)$, and we omit the derivation details owing to the space limitation.

2. The (time-varying) functions $\Theta_k[i]$ are asymptotically minimized (see also the definition in [5]); i.e., $\lim_{i \rightarrow \infty} \Theta_k[i](\mathbf{w}_k[i]) = 0$ for every $k \in \mathcal{N}$.

3. Nodes reach consensus asymptotically; i.e., $\lim_{i \rightarrow \infty} \|\mathbf{w}_k[i] - \mathbf{w}_j[i]\| = 0$ for every $k, j \in \mathcal{N}$.⁴

4. If the parameters ϵ , l , and u in the sets $C(\epsilon)$ and $L_k[i]$ are strictly positive, then every sequence $\mathbf{w}_k[i]$ ($k \in \mathcal{N}$) converges to a point $\hat{\mathbf{w}} \in \mathbb{R}^M$.

5. In addition to the above, we also have $\hat{\mathbf{w}} \in \Omega$ if

$$(\forall \epsilon_4 > 0, \forall r > 0, \exists \xi > 0)$$

$$\sum_{k \in \mathcal{N}} \inf_{\substack{d(\mathbf{w}_k[i], C_k[i]) \geq \epsilon_4 \\ \sum_{k \in \mathcal{N}} \|\mathbf{w}^o - \mathbf{w}_k[i]\| \leq r}} \Theta_k[i](\mathbf{w}_k[i]) \geq \xi.$$

Proof (Sketch) Properties 1-3 are a direct application of [4, Theorem 2]. Properties 4-5 can also be proved with [4, Theorem 2] by observing that \mathbf{w}^o is an interior point of the ideal set Ω^* .

4. EMPIRICAL EVALUATION

We simulate a system similar to that in [2]. The network has $N = 20$ nodes distributed uniformly at random in a unit grid. Two nodes $k, j \in \mathcal{N}$ are neighbors if their Euclidean distance is less than $\sqrt{(\log N)/N}$ (we discard networks not strongly connected). The vector \mathbf{w}_s has dimension 10; it is first drawn from a Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{I})$, and then it is normalized to $\|\mathbf{w}_s\| = 1$. As common in cooperative cognitive radios [2], the regressors $\mathbf{u}_k[i]$ are time invariant, and they are drawn from a Gaussian distribution with p.d.f. $\mathcal{N}(\mathbf{0}, \mathbf{I})$. Noise $v_k[i]$ is i.i.d. in both space and time, and its p.d.f. is $f = (1 - \beta)\mathcal{N}(0, \nu^2) + \beta\mathcal{N}(0, \kappa\nu^2)$, where $\nu > 0$, $0 \leq \beta < 1$, and $\kappa > 1$. This distribution is often used to reproduce impulsive noise in radio channels [7, Ch. 4]. The term $\mathcal{N}(0, \nu^2)$ represents the Gaussian background noise, and the term $\mathcal{N}(0, \kappa\nu^2)$ models the impulsive component, which has probability β of occurring. Here, $\beta = 0.1$ and $\kappa = 1,000$. The parameter ν is adjusted so that the variance of noise at node k , which is given by $\sigma_{v,k}^2 = (1 - \beta)\nu^2 + \beta\kappa\nu^2$, falls within the range [1, 6] in every run. The noise variance for each node is chosen uniformly at random. To show the robustness of the proposed method when many assumptions do not hold, we fix $l = 0$ and $u = 0$ for the sets $L_k[i]$, and we set $\epsilon = 0$ for $C(\epsilon)$. The threshold for detection is $\gamma_k[i] = 0.1$. We mitigate noise by setting the step size of the proposed algorithm to $\mu_k[i] = 0.07 \lambda_k[i]$ and the memory size m to $m = 20$. As in [2], noise is further mitigated by using weights $\omega_{k,j}[i] = \sigma_{v,j}^{-1} / (\sum_{n \in \mathcal{N}_k} \sigma_{v,n}^{-1})$. The proposed method is compared with the LMS approach in [2] with step size 0.06 and detection threshold 0.08. We use as the performance metric the empirical probabilities of misdetection and false alarm in the nodes with the worst performance. They are obtained from 100,000 runs of the simulation. In all runs, $\mathbf{w}_k[0] = \mathbf{0}$ for both algorithms. Fig. 1 shows the performance of the algorithms.

We can see that, at steady state, the proposed algorithm has lower probability of misdetection and false alarm (the probability of false alarm increases initially because we start from the true hypothesis $\mathbf{w}_k[0] = \mathbf{0}$, and the performance cannot be improved by using updates based on noisy measurements). The superior performance of the proposed algorithm is explained by the reduced noise variance in the sets $L_k[i]$ (see the discussion after (4)) and the use *a priori* information. In addition, the proposed method has lower communication overhead because all estimates $\mathbf{w}_k[i]$ belong to $C(0)$ when nodes exchange estimates, so only one scalar per node needs to be reported

⁴Note that items 1, 2, and 3 hold even if we use $C(\epsilon)$ with $\epsilon = 0$.

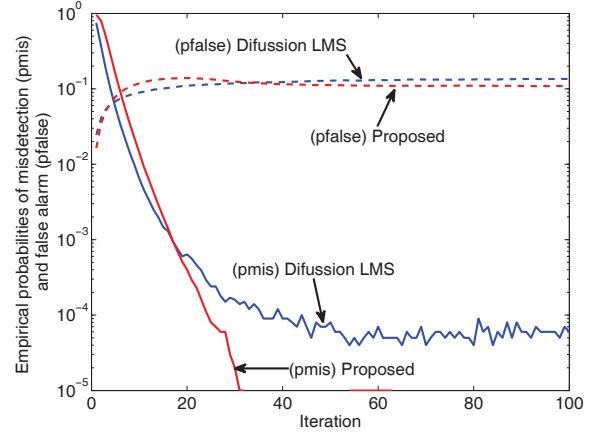


Fig. 1: Empirical probabilities in the nodes with the worst performance as a function of the number of iterations.

in the estimate exchange step (instead of N scalars as in the LMS algorithm).

5. CONCLUSIONS

We have proposed a distributed signal detection algorithm based on distributed set-theoretic adaptive filters. In the proposed method, each node in the network uses *a priori* knowledge of the signal of interest and side-information provided by neighbors. As a result, the algorithm has better detection performance and lower communication overhead than some previous methods in practical scenarios. In addition, by using only few assumptions, we prove rigorously convergence properties of the proposed algorithm; we do not appeal to any approximations as common in the analysis of adaptive filters. We have also showed that, in practice, the algorithm is robust even when the assumptions do not necessarily hold.

Acknowledgement: This research was supported by BMWi - Federal Ministry of Economics and Technology, Germany, under grant 01ME11024

6. REFERENCES

- [1] Z. Tian, E. Blasch, W. Li, G. Chen, and X. Li, "Performance evaluation of distributed compressed wideband sensing for cognitive radio networks," in *ISIF/IEEE Intl. Conf. on Information Fusion (FUSION)*, July 2008, pp. 1–8.
- [2] F. S. Cattivelli and A. H. Sayed, "Distributed detection over adaptive networks using diffusion adaptation," *IEEE Trans. Signal Processing*, vol. 59, no. 5, pp. 1917–1932, May 2011.
- [3] R. L. G. Cavalcante, I. Yamada, and B. Mulgrew, "An adaptive projected subgradient approach to learning in diffusion networks," *IEEE Trans. Signal Processing*, vol. 57, no. 7, pp. 2762–2774, July 2009.
- [4] R. L. G. Cavalcante, A. Rogers, N. R. Jennings, and I. Yamada, "Distributed asymptotic minimization of sequences of convex functions by a broadcast adaptive subgradient method," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 739–753, Aug. 2011.
- [5] I. Yamada and N. Ogura, "Adaptive projected subgradient method for asymptotic minimization of sequence of nonnegative convex functions," *Numerical Functional Analysis and Optimization*, vol. 25, no. 7/8, pp. 593–617, 2004.
- [6] H. Stark and Y. Yang, *Vector Space Projections – A Numerical Approach to Signal and Image Processing, Neural Nets, and Optics*. New York: Wiley, 1998.
- [7] X. Wang and H. V. Poor, *Wireless Communication Systems*. Upper Saddle River, NJ: Prentice Hall, 2004.