ROBUST DISTRIBUTED DETECTION OVER ADAPTIVE DIFFUSION NETWORKS

Sara Al-Sayed⋆*, Abdelhak M. Zoubir*⋆*, and Ali H. Sayed*†

⋆Signal Processing Group, Technische Universität Darmstadt, 64283 Darmstadt, Germany
Email: {salsayed, zoubir}@spg.tu-darmstadt.de
†Department of Electrical Engineering, University of California, Los Angeles, CA 90095
Email: sayed@ee.ucla.edu

ABSTRACT

Diffusion adaptation techniques based on the least-mean-squares criterion have been proposed for distributed detection of a signal in Gaussian-distributed noise, forgoing the need for a fusion center. However, least-mean-squares solutions are generally non-robust against impulsive noise. In this work, we combine non-linear filtering with diffusion adaptation and propose a strategy for distributed detection in the presence of impulsive noise. The superiority of the algorithm is validated experimentally.

***Index Terms*—** Adaptive networks, diffusion LMS, robust distributed detection, hypothesis testing, error nonlinearity.

1. INTRODUCTION

We consider the problem of distributed detection over adaptive networks in the presence of impulsive noise. In the absence of a fusion center, each node cooperates with its neighbors, diffusing information through the network, in order to establish the presence or absence of a known signal using measurements that are corrupted by impulsive noise. By relying solely on local interactions and in-network processing, distributed detection techniques render networks reliable, resilient to node and link failure, scalable and resource efficient. Furthermore, by deploying adaptive techniques for distributed detection, networks are endowed with online learning and tracking abilities in non-stationary environments [1,2]. In [2], a distributed detection technique was proposed under a Gaussian noise assumption and based on the diffusion least-mean-squares (LMS) algorithm, previously developed in [3]. The presence of impulsive noise, however, degrades the performance of the solution [4,5]. An impulsive noise process can be described as one whose realizations contain sparse, random samples of amplitude much higher than nominally accounted for, and hence best modeled by heavy-tailed distributions [5–7]. The incorporation of an error nonlinearity into adaptive filter updates is useful in mitigating the adverse effects of impulsive noise [8–14]. Motivated by these observations, in this work, we develop a robust diffusion strategy for distributed detection that is able to deliver enhanced performance under impulsive noise conditions.

This work was performed within the LOEWE Priority Program Cocoon (http://www.cocoon.tu-darmstadt.de). The work of Ali H. Sayed was supported in part by NSF grant CCF-1011918.

2. DETECTION PROBLEM FORMULATION

2.1. Data model

The network under consideration is composed of N nodes distributed over some region in space. Two nodes that can exchange data are said to be connected. The set of nodes connected to Node k, including itself, is referred to as its neighborhood, and is denoted by $\mathcal{N}_k$. The degree of Node k, denoted by n_k, is the number of its neighbors. At each time instant $i \geq 0$, each node k has access to a real-valued scalar measurement $d_k(i)$ arising from realizations of a random process $\boldsymbol{d}_k(i)$, where the boldface notation is used to denote random variables. These measurements relate to an unknown real-valued vector parameter w^o of size M according to

$$\boldsymbol{d}_k(i) = u_{k,i} w^o + \boldsymbol{v}_k(i) \tag{1}$$

where $u_{k,i}$ is a known deterministic real-valued row regression vector of size M; and $\boldsymbol{v}_k(i)$ is a real-valued scalar wide-sense stationary zero-mean impulsive noise process with variance $\sigma^2_{v,k}$. The random variables $\boldsymbol{v}_k(i)$ and $\boldsymbol{v}_l(j)$ are spatially and temporally independent, for $k \neq l$ or $i \neq j$. It is assumed that the noise probability density functions, $f_{\boldsymbol{v}_k}(v_k)$, are symmetric, for all k, i.e., $\mathbb{E}\,\boldsymbol{v}_k^{2p-1}(i) = 0$, $p = 1, 2, \ldots$, where $\mathbb{E}$ denotes expectation. Model (1) was used in [2]; however, in [2], the noise was assumed to be Gaussian distributed.

The objective is for every node in the network to establish the presence or absence of a known signal given noisy observations, which relates to the simple hypothesis testing problem: $\mathcal{H}_0$: $w^o = 0$; $\mathcal{H}_1$: $w^o = w_s$, where w_s is known.

We arrange the data from all nodes $1, \ldots, N$ at time instant i into vectors and matrices as follows:

$$\boldsymbol{d}_i = \text{col}\{\boldsymbol{d}_1(i), \ldots, \boldsymbol{d}_N(i)\} \tag{2}$$

$$U_i = \text{col}\{u_{1,i}, \ldots, u_{N,i}\} \tag{3}$$

$$\boldsymbol{v}_i = \text{col}\{\boldsymbol{v}_1(i), \ldots, \boldsymbol{v}_N(i)\} \tag{4}$$

$$R_v = \text{diag}\left\{\sigma^2_{v,1}, \ldots, \sigma^2_{v,N}\right\} \tag{5}$$

where the col and diag operators stack their arguments column-wise and diagonally, respectively. Then, we stack the data $\boldsymbol{d}_i$, U_i and $\boldsymbol{v}_i$ from all time instants $i, i-1 \ldots, 0$ in the same manner to obtain $\boldsymbol{d}_{0:i}$, $U_{0:i}$, $\boldsymbol{v}_{0:i}$ and $R_{v,0:i}$. We may therefore express the data model (1) compactly as

$$\boldsymbol{d}_{0:i} = U_{0:i} w^o + \boldsymbol{v}_{0:i}. \tag{6}$$

2.2. Neyman–Pearson-based detection

Based on the Neyman–Pearson (NP) criterion [15], the detector that maximizes the detection probability $P_{d,i}$ given a target false-alarm probability P_f is given by a comparison test of the form

$$\boldsymbol{T}_i\left(\boldsymbol{d}_{0:i}\right) \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \gamma_i. \tag{7}$$

If the noise random vector $\boldsymbol{v}_{0:i}$ is Gaussian distributed, i.e., $\boldsymbol{v}_{0:i} \sim \mathcal{N}\left(0, R_{v,0:i}\right)$, then the test-statistic $\boldsymbol{T}_i\left(\boldsymbol{d}_{0:i}\right)$ is given by [2]

$$\boldsymbol{T}_i\left(\boldsymbol{d}_{0:i}\right) = w_s^T U_{0:i}^T R_{v,0:i}^{-1} \boldsymbol{d}_{0:i} \tag{8}$$

where $(\cdot)^T$ and $(\cdot)^{-1}$ denote matrix transposition and inversion, respectively. The threshold γ_i is computed from the target false-alarm probability as $\gamma_i = \sigma_i\, \mathrm{Q}^{-1}\left(P_f\right)$, where $\mathrm{Q}\left(\cdot\right)$ is the right-tail Gaussian probability function and $\sigma_i^2 = w_s^T U_{0:i}^T R_{v,0:i}^{-1} U_{0:i} w_s$.

Assuming the matrix $U_{0:i}$ is full-rank with $M \leq N$, the minimum-variance unbiased (MVU) estimator of w^o given $\boldsymbol{d}_{0:i}$ in (6) is given by the Gauss–Markov theorem [1]:

$$\boldsymbol{w}_i^{\text{mvu}} = \left(U_{0:i}^T R_{v,0:i}^{-1} U_{0:i}\right)^{-1} U_{0:i}^T R_{v,0:i}^{-1} \boldsymbol{d}_{0:i}. \tag{9}$$

Thus, the NP-optimal test-statistic in (8) can be rewritten in terms of $\boldsymbol{w}_i^{\text{mvu}}$ in (9) as

$$\boldsymbol{T}_i\left(\boldsymbol{w}_i^{\text{mvu}}\right) = w_s^T U_{0:i}^T R_{v,0:i}^{-1} U_{0:i} \boldsymbol{w}_i^{\text{mvu}}. \tag{10}$$

3. DISTRIBUTED DETECTION

The computation, at each node in the network, of the NP-optimal test-statistic $\boldsymbol{T}_i\left(\boldsymbol{d}_{0:i}\right)$ or $\boldsymbol{T}_i\left(\boldsymbol{w}_i^{\text{mvu}}\right)$, using (8) or (10), respectively, and the MVU estimator $\boldsymbol{w}_i^{\text{mvu}}$, using (9), requires that each node have access to the data $\{\boldsymbol{d}_k\left(j\right), u_{k,j}, \sigma_{v,k}^2\}$ from all nodes $1, \ldots, N$ and all time instants $j = 0, \ldots, i$. Since a node can only communicate with its neighbors, adaptive diffusion algorithms present themselves as a viable technique for the approximation of $\boldsymbol{w}_i^{\text{mvu}}$ at each node in the network in a distributed fashion by means of local interactions and in-network processing, as explained in [2,3,16]. Adaptive diffusion algorithms guarantee the dissemination of information from all nodes through the network, so that over time, each node will have incorporated data from beyond its neighborhood's reach. However, the algorithms developed in [2] work well for distributed detection under the Gaussian assumption on the measurement noise. In this work, we consider a more robust adaptive diffusion algorithm, based on the stand-alone counterpart in [5], and show how to extend the distributed formulation of [2] to accommodate impulsive noise scenarios.

3.1. Robust diffusion adaptation

Consider an $N \times N$ matrix A with non-negative real entries $a_{l,k}$ satisfying

$$a_{l,k} = 0 \text{ if } l \notin \mathcal{N}_k, \quad \mathbb{1}^T A = \mathbb{1}^T \tag{11}$$

where $\mathbb{1}$ is the all-one column vector of appropriate size. Let $e_k\left(i\right) \triangleq d_k\left(i\right) - u_{k,i} w_{k,i-1}$ denote the output error of the k^{th} node at time instant i. The update equations for each node k of the adapt-then-combine (ATC) version of the diffusion LMS algorithm are given by [3]

$$\begin{aligned} \psi_{k,i} &= w_{k,i-1} + \mu_k u_{k,i}^T e_k\left(i\right) \\ w_{k,i} &= \sum_{l \in \mathcal{N}_k} a_{l,k} \psi_{l,i} \end{aligned} \tag{12}$$

where μ_k is a positive step-size parameter. Motivated by the discussion in [5], we introduce an error nonlinearity, $h_{k,i}\left(e_k\left(i\right)\right)$, into the adaptation step:

$$\psi_{k,i} = w_{k,i-1} + \mu_k u_{k,i}^T h_{k,i}\left(e_k\left(i\right)\right). \tag{13}$$

The error nonlinearity $h_{k,i}\left(\cdot\right)$ is chosen to be a linear combination of preselected nonlinear basis functions:

$$h_{k,i}\left(e_k\left(i\right)\right) = \alpha_{k,i}^T \varphi_{k,i} \tag{14}$$

where $\alpha_{k,i}$ and $\varphi_{k,i}$, both vectors of length B_k, are Node k's vector of non-negative combination weights at time instant i and vector of nonlinear basis functions evaluated at its output error at time instant i, respectively:

$$\begin{aligned} \alpha_{k,i} &\triangleq \left[\alpha_{k,i}\left(1\right), \ldots, \alpha_{k,i}\left(B_k\right)\right]^T \\ \varphi_{k,i} &\triangleq \left[\phi_{k,1}\left(e_k\left(i\right)\right), \ldots, \phi_{k,B_k}\left(e_k\left(i\right)\right)\right]^T \end{aligned} \tag{15}$$

If Node k were to run the stand-alone counterpart of the adaptive filter in (13), by setting $w_{k,i}$ to $\psi_{k,i}$, then the optimal nonlinearity that minimizes Node k's mean-square error (MSE) was given in [14] as

$$h_{k,i}^{\text{opt}}\left(x\right) = -\frac{f'_{\boldsymbol{e}_k(i)}\left(x\right)}{f_{\boldsymbol{e}_k(i)}\left(x\right)} \tag{16}$$

in terms of the probability density function (pdf) of the error signal, where $f'\left(x\right) \triangleq \frac{df(x)}{dx}$. Here, instead, the nonlinearity is chosen according to (14), and the vector $\alpha_{k,i}$ is found by minimizing the MSE between $h_{k,i}\left(e_k\left(i\right)\right)$ and the optimal nonlinearity:

$$\alpha_{k,i}^{\text{opt}} = \arg\min_{\alpha_{k,i}} \mathbb{E}\left(h_{k,i}^{\text{opt}}\left(\boldsymbol{e}_k\left(i\right)\right) - h_{k,i}\left(\boldsymbol{e}_k\left(i\right)\right)\right)^2. \tag{17}$$

For online adaptation purposes, each node k estimates $\alpha_{k,i}^{\text{opt}}$ adaptively and jointly with w^o, by recourse to a stochastic-gradient recursion and subject to a non-negativity constraint [5,17]. The ensuing moments, $R_{\varphi_{k,i}} \triangleq \mathbb{E}\, \boldsymbol{\varphi}_{k,i} \boldsymbol{\varphi}_{k,i}^T$ and $\mathbb{E}\, \boldsymbol{\varphi}'_{k,i}$, are estimated by means of smoothing recursions.

The choice of basis functions should conform to prior knowledge about the nature of the noise in the data model (1) [18–21]. For example, if we know the noise to be of an impulsive nature, a sensible choice would be

$$\phi_{k,b}\left(x\right) = \tanh\left(bx\right), \; b = 1, \ldots, B_k \tag{18}$$

for every node k, where $\tanh\left(\cdot\right)$ above and $\operatorname{sech}\left(\cdot\right)$ in the table further ahead denote the hyperbolic tangent and secant functions, respectively.

The resulting algorithm is listed in Table 1, where ν_k, $\lambda_k \in (0,1)$ and $\epsilon > 0$ are constants, with ν_k usually close to 1 for the smoothing recursions, and ϵ very small to prevent division by zero; and $\|.\|_\infty$ denotes the maximum absolute entry of its vector argument.

3.2. Robust diffusion detection algorithm

We focus our attention on the incremental update for the k^{th} node in the ATC robust diffusion algorithm:

$$\psi_{k,i} = w_{k,i-1} + \mu_k u_{k,i}^T \sum_{b=1}^{B_k} \alpha_{k,i}(b) \tanh\left(b e_k\left(i\right)\right). \tag{19}$$

Linearizing the error nonlinearities $\phi_{k,b}\left(i\right) = \tanh\left(b e_k\left(i\right)\right)$, $b = 1, \ldots, B_k$, by a Taylor series around $e_k\left(i\right) = 0$ gives $\phi_{k,b}\left(i\right) \approx$

Table 1. Robust Diffusion Detection Algorithm

Initializations: w_s, B_k, $\alpha_{k,-1} \in \mathbb{R}_{++}^{B_k}$, $\hat{R}_{\varphi_{k,-1}}$, $\hat{\varphi}'_{k,-1}$, ν_k, λ_k, ϵ, μ_k.
Start with $w_{k,-1} = 0$ for every node k. For every time instant $i \geq 0$, repeat

Error nonlinearity update: for every node k, repeat

$$\begin{aligned}
e_k(i) &= d_k(i) - u_{k,i} w_{k,i-1} \\
\phi_{k,b}(i) &= \tanh(b e_k(i)),\ b = 1, \ldots, B_k \\
\varphi_{k,i} &= \text{col}\{\phi_{k,1}(i), \ldots, \phi_{k,B_k}(i)\} \\
\hat{R}_{\varphi_{k,i}} &= \nu_k \hat{R}_{\varphi_{k,i-1}} + (1-\nu_k)\varphi_{k,i}\varphi_{k,i}^T \\
\phi'_{k,b}(i) &= b\,\text{sech}^2(b e_k(i)),\ b = 1, \ldots, B_k \\
\varphi'_{k,i} &= \text{col}\left\{\phi'_{k,1}(i), \ldots, \phi'_{k,B_k}(i)\right\} \\
\hat{\varphi}'_{k,i} &= \nu_k \hat{\varphi}'_{k,i-1} + (1-\nu_k)\varphi'_{k,i} \\
\delta_{k,i} &= 2(\hat{R}_{\varphi_{k,i}} \alpha_{k,i-1} - \hat{\varphi}'_{k,i}) \\
\lambda_k(i) &= \lambda_k \frac{\min\{\alpha_{k,i-1}(b), 1 \leq b \leq B_k\}}{\|\delta_{k,i}\|_\infty + \epsilon} \\
\alpha_{k,i} &= \alpha_{k,i-1} - \lambda_k(i)\delta_{k,i} \\
h_k(i) &= \alpha_{k,i}^T \varphi_{k,i}
\end{aligned}$$

Incremental update: for every node k, repeat
$\psi_{k,i} = w_{k,i-1} + \mu_k u_{k,i}^T h_k(i)$

Diffusion update: for every node k, repeat
$w_{k,i} = \sum_{l \in \mathcal{N}_k} a_{l,k} \psi_{l,i}$

Decision: for every node k, repeat
$T_{k,i} = w_s^T Q_{k,i} w_{k,i}$ (see (25) and the note thereafter)
$T_{k,i} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \gamma_{k,i}$ (see (30)–(32))

$b e_k(i)$, $b = 1, \ldots, B_k$. The incremental and diffusion updates in the algorithm can be combined as

$$\begin{aligned} w_{k,i} \approx &\sum_{l \in \mathcal{N}_k} a_{l,k} \left[I_M - \mu_l \left(\alpha_{l,i}^T \beta_l\right) u_{l,i}^T u_{l,i}\right] w_{k,i-1} \\ &+ \sum_{l \in \mathcal{N}_k} a_{l,k} \mu_l \left(\alpha_{l,i}^T \beta_l\right) u_{l,i}^T d_l(i) \end{aligned} \tag{20}$$

where $\beta_k = [1, \ldots, B_k]^T$ and I_M is the identity matrix of size M. Let $\mu_k(i) = \mu_k\left(\alpha_{k,i}^T \beta_k\right)$, $C_{k,i} = I_M - \mu_k(i) u_{k,i}^T u_{k,i}$ and $E_k = \text{diag}\{e_k\}$, where e_k is the all-zero vector of length N and k^{th} entry equal to 1. By induction, it can be verified that $w_{k,i} \approx K_{k,i} d_{0:i}$, where

$$K_{k,i} = \left[\sum_{l \in \mathcal{N}_k} a_{l,k}\mu_l(i) U_i^T E_l \quad \sum_{l \in \mathcal{N}_k} a_{l,k} C_{l,i} K_{l,i-1}\right]. \tag{21}$$

By the central limit theorem, the estimates $w_{k,i}$ are therefore asymptotically, as $i \to \infty$, approximately Gaussian distributed. In this case, if $K_{k,i}$ is full-rank with $M \leq N$, and motivated by (6) and (8), a near-optimal NP detector at the k^{th} node is given by

$$\boldsymbol{T}_{k,i}(\boldsymbol{w}_{k,i}) \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \gamma_{k,i} \tag{22}$$

with the local test-statistic given by

$$T_{k,i}(w_{k,i}) = w_s^T Q_{k,i}^{\text{opt}} w_{k,i} \tag{23}$$

where

$$Q_{k,i}^{\text{opt}} = (K_{k,i} U_{0:i})^T \left(K_{k,i} R_{v,0:i} K_{k,i}^T\right)^{-1}. \tag{24}$$

The threshold at the k^{th} node, $\gamma_{k,i}$, is to be computed in a distributed manner as well in terms of the target false-alarm probability. This is addressed in Sec. 4.

In order to reduce the communication and computational burden at each node, we may approximate $Q_{k,i}^{\text{opt}}$ in (24). If we overlook the diffusion operation by setting A to the identity matrix in (21), a reasonable substitute for $Q_{k,i}^{\text{opt}}$ under small step-sizes μ_k is

$$Q_{k,i} = \left(\sum_{j=0}^{i} \mu_k(j) u_{k,j}^T u_{k,j}\right)\left(\sum_{j=0}^{i} \mu_k^2(j) u_{k,j}^T u_{k,j}\right)^{-1} \tag{25}$$

for $i \geq M-1$, assuming invertibility. For $i < M-1$, $Q_{k,i}$ is set to I_M. Asymptotically, as $i \to \infty$, the random variable $\boldsymbol{Q}_{k,i}$ can be shown to tend, in expectation, to a scaled identity matrix:

$$\lim_{i \to \infty} \mathbb{E}\, \boldsymbol{Q}_{k,i} = \eta_k I_M. \tag{26}$$

where η_k is the limiting value of $\mathbb{E}\, \boldsymbol{\mu}_k^{-1}(i)$, subject to algorithm stability. Since the inverted expression in (25) constitutes a running sum over time of unit-rank matrices, we may appeal to the Sherman–Morrison formula for matrix inversion to simplify the computation [22].

4. PERFORMANCE ANALYSIS

In this section, we analyze briefly the detection performance of the robust diffusion detection algorithm subject to the data model described in Sec. 2.1. Let $\tilde{\boldsymbol{w}}_{k,i} \triangleq w^o - \boldsymbol{w}_{k,i}$ denote Node k's weight-error vector at time instant i. We make the following additional assumptions:

(*A1*) The step-size μ_k is sufficiently small, for all k.
(*A2*) $\boldsymbol{\alpha}_{k,i}$ is independent of $\boldsymbol{v}_l(i)$ and $\tilde{\boldsymbol{w}}_{l,i}$, for all k, l, and i.

The second assumption is reasonable under small step-size μ_k, more so when ν_k is close to 1, and asymptotically, as $i \to \infty$ [1, 23].

For sufficiently large i, the test-statistics are distributed as

$$\boldsymbol{T}_{k,i}(\boldsymbol{w}_{k,i}) \sim \mathcal{N}\left(\eta_k w_s^T \mathbb{E}\, \boldsymbol{w}_{k,i}, \sigma_{k,i}^2\right) \tag{27}$$

where $\sigma_{k,i}^2 = \eta_k^2 w_s^T R_{\tilde{w}_{k,i}} w_s$, with $R_{\tilde{w}_{k,i}}$ denoting Node k's weight-error covariance matrix at time instant i:

$$R_{\tilde{w}_{k,i}} \triangleq \mathbb{E}\left(\tilde{\boldsymbol{w}}_{k,i} - \mathbb{E}\,\tilde{\boldsymbol{w}}_{k,i}\right)\left(\tilde{\boldsymbol{w}}_{k,i} - \mathbb{E}\,\tilde{\boldsymbol{w}}_{k,i}\right)^T. \tag{28}$$

Hence, the detection, false-alarm and miss probabilities at each node k and time instant i are asymptotically given by

$$\begin{aligned} P_{d,k,i} &= \mathrm{Q}\left(\frac{\gamma_{k,i} - \eta_k w_s^T w_s + \eta_k w_s^T \mathbb{E}\,\tilde{\boldsymbol{w}}_{k,i}}{\sigma_{k,i}}\right) \\ P_{f,k,i} &= \mathrm{Q}\left(\frac{\gamma_{k,i} + \eta_k w_s^T \mathbb{E}\,\tilde{\boldsymbol{w}}_{k,i}}{\sigma_{k,i}}\right) \end{aligned} \tag{29}$$

and $P_{m,k,i} = 1 - P_{d,k,i}$. Given target false-alarm probabilities at each node k and time instant i, under the assumption of asymptotic unbiasedness of the weight estimates $\boldsymbol{w}_{k,i}$, the corresponding detection thresholds may subsequently be approximated, in a distributed manner, as

$$\gamma_{k,i} = \frac{1}{\sqrt{g}} \hat{\sigma}_{k,i}^{A=I}\, \mathrm{Q}^{-1}(P_{f,k,i}) \tag{30}$$

where $\left(\hat{\sigma}_{k,i}^{A=I}\right)^2 = w_s^T Q_{k,i} \hat{R}_{\tilde{w}_{k,i}}^{A=I} Q_{k,i} w_s$, with $\hat{R}_{\tilde{w}_{k,i}}^{A=I}$ given by the following recursion:

$$\begin{aligned} \hat{R}_{\tilde{w}_{k,i}}^{A=I} = &\left[I_M - \mu_k \hat{p}_k(i) u_{k,i}^T u_{k,i}\right] \hat{R}_{\tilde{w}_{k,i-1}}^{A=I} \left[I_M - \mu_k \hat{p}_k(i) u_{k,i}^T u_{k,i}\right] \\ &+ \mu_k^2 \hat{s}_k(i) u_{k,i}^T u_{k,i}, \qquad \hat{R}_{\tilde{w}_{k,-1}}^{A=I} = 0. \end{aligned} \tag{31}$$

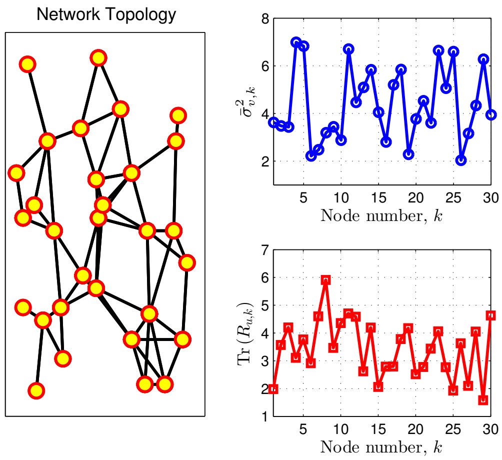


Fig. 1. Network topology, node nominal noise variances $\bar{\sigma}^2_{v,k}$ and regressor covariance traces $\mathrm{Tr}\,(R_{u,k})$, for $N = 30$ nodes.

The estimated moments $\hat{p}_k\,(i)$ and $\hat{s}_k\,(i)$ are stochastic approximations of their true counterparts, reusing smoothed estimates from the algorithm:

$$\hat{p}_k\,(i) = \alpha^T_{k,i}\hat{\varphi}'_{k,i} \qquad \hat{s}_k\,(i) = \alpha^T_{k,i}\hat{R}_{\varphi_{k,i}}\alpha_{k,i} \tag{32}$$

The approximate recursion in (31) follows from straightforward analysis subject to the aforementioned assumptions. A sketch of the derivation follows. From model (1), it holds that $\boldsymbol{e}_k\,(i) = u_{k,i}\tilde{\boldsymbol{w}}_{k,i-1} + \boldsymbol{v}_k\,(i)$. The nonlinearities $\tanh\left(b\boldsymbol{e}_k\,(i)\right)$, $b = 1,\ldots,B_k$, are expanded around 0 using the corresponding Taylor series. The binomial theorem is then used to expand the resulting powers of $\boldsymbol{e}_k\,(i)$, discarding powers of $u_{k,i}\tilde{\boldsymbol{w}}_{k,i-1}$ higher than 2, which are negligible under (*A1*) towards steady-state (cf. [9, 10]).

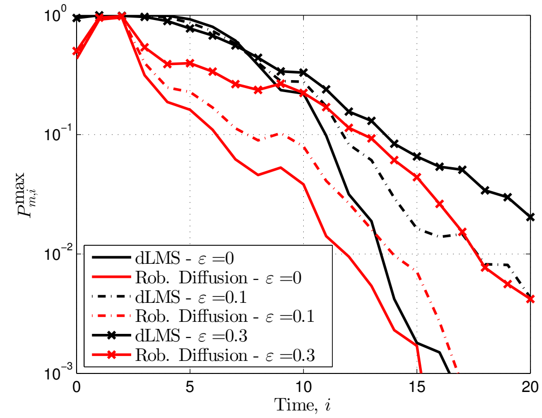
For comparison, the LMS-based algorithm uses the following recursion [2]:

$$\begin{aligned} R^{A=I}_{\tilde{w}_{k,i}} &= \left[I_M - \mu_k u^T_{k,i}u_{k,i}\right] R^{A=I}_{\tilde{w}_{k,i-1}} \left[I_M - \mu_k u^T_{k,i}u_{k,i}\right] \\ &\quad + \mu^2_k\sigma^2_{v,k}u^T_{k,i}u_{k,i}, \qquad R^{A=I}_{\tilde{w}_{k,-1}} = 0. \end{aligned} \tag{33}$$

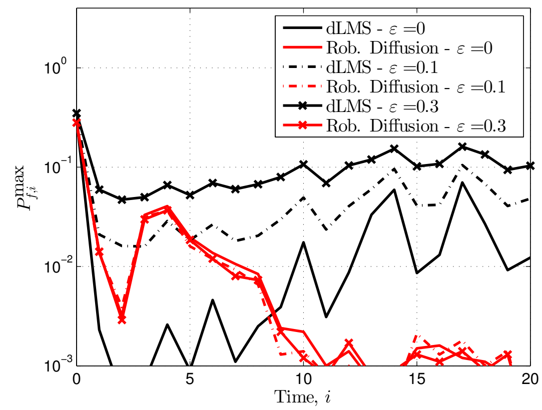
The correction factor $g^{-\frac{1}{2}}$ accounts for the gain incurred by the diffusion process and can be estimated offline (cf. [2]).

5. SIMULATIONS

We consider a network of $N = 30$ nodes, seeking to detect a unit-norm signal vector w_s of size $M = 2$. We compare the worst-case detection and false-alarm performance over time of the ATC diffusion LMS-based detection algorithm of [2] and the robust counterpart developed in this work. The regressors $u_{k,i}$ and noise samples $v_k\,(i)$ are drawn independently across time and space and identically distributed across time: the regressors from a multivariate zero-mean Gaussian distribution with covariances $R_{u,k}$, with the same set maintained throughout the experiments; and the noise samples according to an ε-contaminated Gaussian mixture model with pdf $f_{\boldsymbol{v}_k}\,(v_k) = (1-\varepsilon)\,\mathcal{N}\left(0, \bar{\sigma}^2_{v,k}\right) + \varepsilon\mathcal{N}\left(0, \kappa\bar{\sigma}^2_{v,k}\right)$, where $\bar{\sigma}^2_{v,k}$ are the nominal noise variances, ε is the contamination ratio, and $\kappa \gg 1$. Herein, κ is set to 10. The network topology, regressor covariance traces and nominal noise variances are shown in Fig. 1, depicting a scenario of low signal-to-noise ratio. The weighting coefficients $a_{l,k}$ are chosen according to the relative-degree rule, i.e., $a_{l,k} = n_l/\Sigma_{m\in\mathcal{N}_k} n_m$. The target false-alarm probabilities $P_{f,k,i}$ are set to 10^{-2} for every node k and time instant i. The factor g was estimated offline and set to 4. The step-sizes μ_k are set to be the same across the nodes, but selected uniquely for each algorithm in such a way as to equalize their convergence rates for the case of no contamination ($\varepsilon = 0$) for fair comparison: $\mu^{\text{dLMS}} = 0.025$ and $\mu^{\text{rob}} = 0.09$. For the robust algorithm, we consider two basis functions, i.e., $B_k = B = 2$, for all k. The initial estimates of the basis weights, $\alpha_{k,-1}$, are set to $\frac{1}{B}\mathbb{1}$, and λ_k to 10^{-2} for every node k. For the smoothing recursions, zero initial conditions are assumed, and ν_k is set to 0.9 for every node k. Finally, ϵ is set to 10^{-6}. All simulation results are obtained by averaging over 10,000 experiments.



(a)



(b)

Fig. 2. Worst-case (a) detection performance (b) false-alarm performance of diffusion LMS-based detection (black) and robust diffusion detection (red) for different contamination ratios, ε.

In Fig. 2, the resulting worst-case performance of both algorithms is displayed for various degrees of contamination. The superiority of the robust algorithm is evident, insensitive as it is to noise impulsiveness. Moreover, the robustness of the algorithm figures prominently with respect to the false-alarm performance, since it is only the robust algorithm that meets the target false-alarm probability over time in the worst case.

6. REFERENCES

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

[2] 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.

[3] 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.

[4] N. J. Bershad, "On error saturation nonlinearities for LMS adaptation in impulsive noise," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4526–4530, Sept. 2008.

[5] S. Al-Sayed, A. M. Zoubir, and A. H. Sayed, "An optimal error nonlinearity for robust adaptation against impulsive noise," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Darmstadt, Germany, June 2013, pp. 415–419.

[6] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, "Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts," *IEEE Signal Process. Mag.*, vol. 29, no. 4, pp. 61–80, July 2012.

[7] A. M. Zoubir and R. F. Brcich, "Multiuser detection in heavy tailed noise," *Digital Signal Processing*, vol. 12, no. 2-3, pp. 262–273, 2002.

[8] J. Chambers and A. Avlonitis, "A robust mixed-norm adaptive filter algorithm," *IEEE Signal Process. Lett.*, vol. 4, no. 2, pp. 46–48, Feb. 1997.

[9] A. K. Barros, J. Principe, Y. Takeuchi, C. H. Sales, and N. Ohnishi, "An algorithm based on the even moments of the error," in *Proc. IEEE Workshop on Neural Networks for Signal Processing (NNSP)*, Toulouse, France, Sept. 2003, pp. 879–885.

[10] C. Boukis, D. P. Mandic, and A. G. Constantinides, "A generalised mixed norm stochastic gradient algorithm," in *Proc. International Conference on Digital Signal Processing*, Cardiff, Wales, UK, July 2007, pp. 27–30.

[11] L. R. Vega, H. Rey, J. Benesty, and S. Tressens, "A new robust variable step-size NLMS algorithm," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1878–1893, May 2008.

[12] T. Gänsler, S. L. Gay, M. M. Sondhi, and J. Benesty, "Double-talk robust fast converging algorithms for network echo cancellation," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 6, pp. 656–663, Nov. 2000.

[13] S. C. Douglas and T. H.-Y. Meng, "Stochastic gradient adaptation under general error criteria," *IEEE Trans. Signal Process.*, vol. 42, no. 6, pp. 1335–1351, June 1994.

[14] T. Y. Al-Naffouri and A. H. Sayed, "Adaptive filters with error nonlinearities: Mean-square analysis and optimum design," *EURASIP Journal on Advances in Signal Processing*, vol. 2001, no. 1, pp. 192–205, Jan. 2001.

[15] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, vol. 1, Prentice Hall, NJ, 1998.

[16] 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.

[17] N. Takahashi, I. Yamada, and A. H. Sayed, "Diffusion least-mean squares with adaptive combiners: Formulation and performance analysis," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4795–4810, Sept. 2010.

[18] A. Taleb, R. Brcich, and M. Green, "Suboptimal robust estimation for signal plus noise models," in *Proc. Asilomar Conference on Signals, Systems and Computers*, Nov. 2000, vol. 2, pp. 837–841.

[19] R. F. Brcich and A. M. Zoubir, "Robust estimation with parametric score function estimation," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Orlando, FL, USA, May 2002, vol. 2, pp. 1149–1152.

[20] P. J. Huber and E. M. Ronchetti, *Robust Statistics*, Wiley, NJ, 2009.

[21] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions*, Wiley, NJ, 1986.

[22] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1990.

[23] R. H. Kwong and E. W. Johnston, "A variable step size LMS algorithm," *IEEE Trans. Signal Process.*, vol. 40, no. 7, pp. 1633–1642, July 1992.