

A Robust Adaptive Estimation Algorithm for Hamiltonian Multi-agent Networks

Hoorie Giv, Azam Khalili, and Amir Rastegarnia, *Member, IEEE* and Saeid Sanei, *Senior Member, IEEE*

Abstract—In this paper a robust incremental adaptation algorithm is presented to solve distributed estimation for a Hamiltonian network, where the measurements at each node may be corrupted by heavy-tailed impulsive noise. In the proposed algorithm, each node employs an error-nonlinearity into the update equation to mitigate the detrimental effects of impulsive noise. Moreover, the algorithm estimates both the optimal error non-linearity and the unknown parameter together, which in turn, obviates the requirement of prior knowledge about the statistical characteristics of measurement noise. In addition to algorithm development, its steady-state performance as well as convergence analysis have been provided. Simulation results validate the correctness of the analysis and reveal the superiority of the proposed algorithm over some existing algorithms.

Index terms—Adaptive network, Hamilton, incremental, robust estimation.

I. INTRODUCTION

The problem of estimating an unknown deterministic parameter by nodes of a connected network is considered here. In particular, we are interested in the case where the measurements at each node are corrupted by impulsive noise. This problem (robust distributed estimation) appears in many practical applications. For example, in the tracking of a ground target using an acoustic wireless sensor network, the measurements of RF/acoustic sensors are prone to exhibit false measurements due to multi-path reflections [1]. Typical existing solutions include adaptive networks and consensus schemes. Here, we focus on adaptive networks, as they exhibit superior performances compared to consensus-based solutions [2]. Two strategies have gained the most popularity in adaptive networks, including incremental strategy [3], [4] and diffusion strategy. In the incremental strategy, a cyclic path (Hamilton cycle) is established among the

nodes, in which every agent is visited exactly once and each agent passes the information to its adjacent node. In the diffusion strategy, the nodes communicate with their immediate neighbours [5]–[8].

The most traditional cost function in adaptive networks is the minimum mean square error (MSE) criterion. Under the Gaussian assumption, the algorithms that employ MSE can achieve optimal performance. However, their performance severely degrades in the presence of non-Gaussian noise, specially heavy-tailed impulsive noise cases [9]. To overcome this issue, robust incremental adaptive networks with alternative cost functions such as Wilcoxon norm, Huber norm, and maximum correntropy criterion have been reported in [10]–[13]. Similar robust diffusion adaptive networks are available in the literature [14]–[18].

Although these algorithms exhibit superior performance compared to MSE-based algorithms, they need some parameter tuning, which in turn, requires some prior knowledge about the measurement noise distribution. So, developing a new robust algorithm that obviates the requirement of such prior information is crucial. In this paper, we extend the framework in [9] and [19] and propose a robust adaptive Hamiltonian network. Unlike the given algorithm in [19] which requires a connected network with diffusion topology, the proposed algorithm relies on the incremental mode of cooperation. A Hamilton cycle may not be available in large networks. Nevertheless, in comparison with diffusion adaptive networks, incremental-based algorithms have their own advantages (low power consumption and low communication load), especially in networks with a small number of nodes or networks where the network topology can be controlled [20]. Moreover, neither the algorithm derivation procedure nor the obtained results in [19] can be extended to the incremental networks.

In the proposed algorithm, each node incorporates an error non-linearity into its update equation to handle the heavy-tailed impulsive noise. On the other hand, it tunes the required parameters adaptively. A detailed steady-state performance and convergence analysis is carried out and verified by simulation. Compared to the existing algorithms [10]–[13], the proposed algorithm

Manuscript received December 2020. H. Giv, A. Khalili and A. Rastegarnia are with the Department of Electrical Engineering, Malayer University, Malayer 65719-95863, Iran (email: hgiv92@gmail.com, khalili@malayeru.ac.ir, rastegar@malayeru.ac.ir). S. Sanei is with the School of Science and Technology, Nottingham Trent University, Clifton Lane, Nottingham, U.K. (e-mail: saeid.sanei@ntu.ac.uk)

Digital Object Identifier 2020/XX

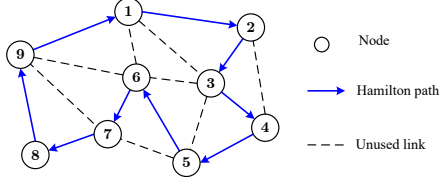


Fig. 1. A Hamiltonian network with $K = 9$ nodes.

has a higher computational complexity. However, as an advantage, the parameters are adjusted adaptively during the learning process, which obviates any prior information for precise parameter setting.

We adopt normal font letters, small boldface letters and bold capital letters for scalars, vectors and matrices, respectively. The notations $\text{tr}(\cdot)$, $(\cdot)^\top$ and $\|\cdot\|_\infty$ denote the trace of a matrix, matrix transposition and maximum absolute entry of its vector argument. The weighted Euclidean norm of \mathbf{x} with a weighting matrix Ξ is $\|\mathbf{x}\|_\Xi^2 = \mathbf{x}^\top \Xi \mathbf{x}$. $\text{col}\{x_1, x_2\}$ is a column vector whose entries are x_1 and x_2 .

II. ROBUST INCREMENTAL ALGORITHM

Consider a Hamiltonian network with K nodes as shown in Fig. 1. The scalar noisy measurement by node k at discrete time instant i is described by a linear regression model as

$$d_{k,i} = \mathbf{u}_{k,i} \mathbf{w}^o + v_{k,i} \quad (1)$$

where $v_{k,i}$ denotes the zero-mean measurement noise term. The $1 \times M$ regression vectors $\{\mathbf{u}_{k,i}\}$ are realizations of wide-sense stationary zero-mean random process with positive definite covariance matrix $\mathbf{R}_{\mathbf{u}_k}$.

Assumption 1: The measurement noise $v_{k,i}$ is a white process with variance $\sigma_{v,k}^2$. Moreover, $\{\mathbf{u}_{k,i}\}$ and $\{v_{\ell,j}\}$ are independent for all $k \neq \ell$ and $i \neq j$.

The objective is to develop an algorithm to adaptively estimate the unknown $M \times 1$ parameter vector \mathbf{w}^o in a fully distributed manner. To this end, the incremental least mean-squares (LMS) algorithm was developed [3]. The update equation for the incremental LMS algorithm at node k and iteration i is

$$\boldsymbol{\theta}_{k,i} = \boldsymbol{\theta}_{k-1,i} + \mu_k \mathbf{u}_{k,i}^* (d_{k,i} - \mathbf{u}_{k,i} \boldsymbol{\theta}_{k-1,i}) \quad (2)$$

with $\boldsymbol{\theta}_{1,i} = \boldsymbol{\theta}_{K,i-1}$. The algorithm exhibits good performance in the presence of Gaussian noise, but is not robust against heavy-tailed impulsive noise [9], [12].

One useful way to mitigate the effect of heavy-tailed impulsive noise is to use an error non-linearity function in the update equation. Thus, inspired by the discussion in [9], the robust incremental algorithm with error non-linearity is given by

$$\mathbf{w}_{k,i} = \mathbf{w}_{k-1,i} + \mu_k \mathbf{u}_{k,i}^\top h_{k,i}(e_{k,i}) \quad (3)$$

In [9] it is proven that the optimum function which minimizes the MSE at every node k is given by

$$h_{k,i}^o = -\frac{p'_{e_{k,i}}}{p_{e_{k,i}}} \quad (4)$$

where $p_{e_{k,i}}$ and $p'_{e_{k,i}}$ denote the PDF of error signal $e_{k,i}$ and its first derivative, respectively. Clearly, in most practical applications this PDF is not available. A useful alternative for (4) is to define $h_{k,i}(e_{k,i})$ in terms of a set of preselected sign-preserving basis functions $\{f_{k,b}(\cdot)\}$, $b = 1, \dots, B$ as

$$h_{k,i}(e_{k,i}) = \mathbf{a}_{k,i}^\top \mathbf{b}_{k,i} \quad (5)$$

Here, the $B \times 1$ vector $\mathbf{b}_{k,i}$ is defined as

$$\mathbf{b}_{k,i} = [f_{k,1}(e_{k,i}), f_{k,2}(e_{k,i}), \dots, f_{k,B}(e_{k,i})]^\top \quad (6)$$

A reasonable choice for $f_{k,b}$ to handle impulsive noise is [9]

$$f_{k,b}(x) = \begin{cases} x & b = 1 \\ \tanh((b-1)x) & b = 2, \dots, B \end{cases} \quad (7)$$

The $B \times 1$ vector $\mathbf{a}_{k,i}$ in (5) consists of the non-negative combination weights as

$$\mathbf{a}_{k,i} = [\mathbf{a}_{k,i}(1), \mathbf{a}_{k,i}(2), \dots, \mathbf{a}_{k,i}(B)]^\top \quad (8)$$

The optimum value of $\mathbf{a}_{k,i}$ (denoted by $\mathbf{a}_{k,i}^o$) is obtained by solving the following minimization problem

$$\begin{aligned} & \arg \min_{\mathbf{a}_{k,i}} \mathbb{E}[h_{k,i}^o(e_{k,i}) - h_{k,i}(e_{k,i})]^2 \\ & \text{subject to } \mathbf{a}_{k,i}(b) \geq 0, \sum_{b=1}^B \mathbf{a}_{k,i}(b) = 1 \end{aligned} \quad (9)$$

Note that, the convexity constraint in (9) is required to ensure the boundedness of $\mathbf{a}_{k,i}$. To solve the convex optimization problem (9) we apply the technique introduced in [9]. The idea is to eliminate the constraint $\mathbf{a}_{k,i}(b) \geq 0$ and then transform the solution appropriately to accommodate this constraint. Doing so, a fully distributed robust incremental algorithm can be obtained as presented in Algorithm 1.

Note that f' denotes the first derivative and $\text{sgm}(x) \triangleq \frac{1}{1+e^{-x}}$. Moreover, $R_{\mathbf{b}_{k,i}} = \mathbb{E}[\mathbf{b}_{k,i} \mathbf{b}_{k,i}^\top]$ and this matrix is recursively estimated by (10d). The vector $\mathbf{b}'_{k,i}$ denotes entry-wise differentiation as defined in (10f) and is estimated by (10g). Moreover, $0 < \nu_k < 1$ and $\beta > 0$ are constant parameters. We further define $\mathbf{P}_k = \mathbf{I} - \frac{\mathbf{1}\mathbf{1}^\top}{B}$ of size B where

$$\mathcal{S}_{+,k} = \{\mathbf{x} \in \mathbb{R}_+^B | \mathbf{x}^\top \mathbf{1} = 1\} \quad (11)$$

where \mathbb{R}_+^B is the set of $B_k \times 1$ vectors on the set of positive real numbers \mathbb{R}_+ .

Remark 1: (Computational complexity) The implementation of Algorithm 1 includes $5(1+B)+4B^2+B^3+$

Algorithm 1 Robust incremental estimation algorithm

- 1: Input parameters: $\mu_k, \beta, B, \{f_{k,b}(x)\}, \mathbf{P}_k, \nu_k$
 - 2: Initialize $\mathbf{w}_{k,0}, \mathbf{a}_{k,0}, \widehat{\mathbf{R}}_{\mathbf{b}_{k,0}}, \widehat{\mathbf{b}}'_{k,0}, \widehat{\lambda}_{k,0}$
 - 3: **for** $i = 1, 2, \dots$ **do**
 - 4: **for** $k = 1 : K$ **do**
 - (10a) $e_{k,i} = d_{k,i} - \mathbf{u}_{k,i} \mathbf{w}_{k-1,i}$
 - (10b) $f_{e|k,i}(b) \triangleq f_{k,b}(e_{k,i}), b = 1, \dots, B$
 - (10c) $\mathbf{b}_{k,i} = \text{col}\{f_{e|k,i}(1), \dots, f_{e|k,i}(B)\}$
 - (10d) $\widehat{\mathbf{R}}_{\mathbf{b}_{k,i}} = \nu_k \widehat{\mathbf{R}}_{\mathbf{b}_{k,i-1}} + (1 - \nu_k) \mathbf{b}_{k,i} \mathbf{b}_{k,i}^\top$
 - (10e) $f'_{e|k,i}(b) \triangleq f'_{k,b}(e_{k,i}), b = 1, \dots, B$
 - (10f) $\mathbf{b}'_{k,i} = \text{col}\{f'_{e|k,i}(1), \dots, f'_{e|k,i}(B)\}$
 - (10g) $\widehat{\mathbf{b}}'_{k,i} = \nu_k \widehat{\mathbf{b}}'_{k,i-1} + (1 - \nu_k) \mathbf{b}'_{k,i}$
 - (10h) $\boldsymbol{\delta}_{k,i} = 2\mathbf{P}_k (\widehat{\mathbf{R}}_{\mathbf{b}_{k,i}} \mathbf{a}_{k,i-1} - \widehat{\mathbf{b}}'_{k,i})$
 - (10i) $\widehat{\lambda}_{k,i} = \nu_k \widehat{\lambda}_{k,i-1} + (1 - \nu_k) \frac{\|\mathbf{u}_{k,i}\|^2}{M}$
 - (10j) $\widehat{\tau}_{k,i} = \text{sgm}((\mathbf{a}_{k,i-1}^\top \widehat{\mathbf{b}}'_{k,i}) \widehat{\lambda}_{k,i})$
 - (10k) $\tau_{k,i} = \widehat{\tau}_{k,i} \frac{\min\{\mathbf{a}_{k,i-1}(b), 1 \leq b \leq B\}}{\|\boldsymbol{\delta}_{k,i}\|_\infty + \beta}$
 - (10l) $\mathbf{a}_{k,i} = \mathbf{a}_{k,i-1} - \tau_{k,i} \boldsymbol{\delta}_{k,i}$
 - (10m) $h_{k,i} = \mathbf{a}_{k,i}^\top \mathbf{b}_{k,i}$
 - 5: update $\mathbf{w}_{k,i}$ as $\mathbf{w}_{k,i} = \mathbf{w}_{k-1,i} + \mu_k \mathbf{u}_{k,i}^\top h_{k,i}$
 - 6: **end for**
 - 7: **end for**
-

$3M$ multiplications, $3B - 1 + 2B^2 + B^3 + M$ additions and one division per iteration at each individual node.

Remark 2: (Parameter tuning) In the algorithm, only ν_k, B and $f_{k,b}(x)$ need to be preselected for each node. Similar to [9] and [19], here, we set $0 \ll \nu_k < 1$, $B = 2$, and $f_{k,b}(x)$ as equation (7). Larger values for B gives better approximates of $h_{k,i}^o$, but increases the computational complexity. Also, optimization of $f_{k,b}(x)$ needs the exact noise distribution. The other variables are tuned adaptively through the learning process. So, unlike other algorithms, the proposed algorithm does not require any precise parameter setting (see Table. 1).

III. PERFORMANCE ANALYSIS

A. Assumptions and Definitions

Due to the nonlinear and stochastic nature of the update relations, we need to introduce the following Assumption to make the analysis more tractable,

Assumption 2:

- (i) The regressors $\{\mathbf{u}_{k,i}\}$ are independently and identically distributed.
- (ii) $\mathbf{a}_{k,i}$ is independent of $u_{\ell,j}$ and $v_{\ell,j}$ for $k \neq \ell$ and $i \neq j$.
- (iii) $\{f_{k,b}(x)\}$ are sign-preserving, odd-symmetric, monotonically increasing, and differentiable.

Remark 3: (Justification of assumptions) Assumption 2(i) is widely used in adaptive filtering approaches. As it is discussed in [21] this assumption not only

simplifies the analysis, but also the results for dependent regressions are close to independent ones for small step-sizes. Assumption 2(ii) is more reasonable under small step-sizes. Finally, Assumption 2(iii) is used to ensure that the successive weight estimates $\mathbf{w}_{k,i}$ decrease the MSE level. Moreover, the odd-symmetric and monotonically increasing properties are used to make the analysis tractable. More details on justifications of these assumptions are available at [9] and [19].

In the analysis, both input signals and noise are considered stationary. We define the following error signals, as they will be used frequently in our analysis:

$$\widetilde{\mathbf{w}}_{k,i} \triangleq \mathbf{w}^o - \mathbf{w}_{k,i} \quad (12)$$

$$\epsilon_{k,i} \triangleq \mathbf{u}_{k,i} \widetilde{\mathbf{w}}_{k-1,i} \text{ (a priori error)} \quad (13)$$

$$e_{k,i} \triangleq d_{k,i} - \mathbf{u}_{k,i} \mathbf{w}_{k-1,i} \text{ (output error)} \quad (14)$$

The mean-square deviation (MSD) and excess mean-square error (EMSE) as the performance metrics are defined as

$$\eta_{k,i} = \mathbb{E}[\|\widetilde{\mathbf{w}}_{k,i}\|_{\mathbf{I}}^2], \text{ (MSD)} \quad (15)$$

$$\zeta_{k,i} = \mathbb{E}[\|\epsilon_{k,i}\|^2] = \mathbb{E}[\|\widetilde{\mathbf{w}}_{k,i}\|_{\mathbf{R}_{\mathbf{u}_k}}^2], \text{ (EMSE)} \quad (16)$$

B. Steady-State Analysis

To analyze the performance, it firstly suffices to evaluate $\mathbb{E}[\|\widetilde{\mathbf{w}}_{k,i}\|_{\Sigma}^2]$ for positive semi-definite matrix Σ and then particularize the previous performance measure for MSD by setting $\Sigma = \mathbf{I}$ and EMSE by $\Sigma = \mathbf{R}_{\mathbf{u}_k}$. To this end, the update equation (3) is considered. By subtracting \mathbf{w}^o from both sides of (3) we have

$$\widetilde{\mathbf{w}}_{k,i} = \widetilde{\mathbf{w}}_{k-1,i} - \mu_k \mathbf{u}_{k,i}^\top h_{k,i} \quad (17)$$

Applying weighted norm to both sides of (17) and taking the expectations yields

$$\begin{aligned} \mathbb{E}[\|\widetilde{\mathbf{w}}_{k,i}\|_{\Sigma}^2] &= \mathbb{E}[\|\widetilde{\mathbf{w}}_{k-1,i}\|_{\Sigma}^2] - \underbrace{2\mu_k \mathbb{E}[\mathbf{u}_{k,i} \Sigma \widetilde{\mathbf{w}}_{k-1,i} h_{k,i}]}_1 \\ &\quad + \underbrace{\mu_k^2 \mathbb{E}[\|\mathbf{u}_{k,i}\|_{\Sigma}^2 h_{k,i}^2]}_2 \end{aligned} \quad (18)$$

In order to evaluate the moments 1 and 2, firstly $h_{k,i}$ is approximated by using a Taylor series expansion for $\{f_{k,b}(x)\}$ around $\epsilon_{k,i} = 0$ for all $i \geq 0$ as

$$\begin{aligned} h_{k,i} &= \sum_{b=1}^B \mathbf{a}_{k,i}(b) f_{k,b}(e_{k,i}) \\ &\approx \sum_{b=1}^B \mathbf{a}_{k,i}(b) f_{v|k,i}(b) + \epsilon_{k,i} \sum_{b=1}^B \mathbf{a}_{k,i}(b) f'_{v|k,i}(b) \\ &= \mathbf{a}_{k,i}^\top \mathbf{c}_{k,i} + \epsilon_{k,i} \mathbf{a}_{k,i}^\top \mathbf{c}'_{k,i} \end{aligned} \quad (19)$$

where

$$f_{v|k,i}(b) \triangleq f_{k,b}(v_{k,i}), \quad b = 1, \dots, B \quad (20)$$

$$f'_{v|k,i}(b) \triangleq f'_{k,b}(v_{k,i}), \quad b = 1, \dots, B \quad (21)$$

$$\mathbf{c}_{k,i} \triangleq \text{col}\{f_{v|k,i}(1), \dots, f_{v|k,i}(B)\}, \quad (22)$$

$$\mathbf{c}'_{k,i} \triangleq \text{col}\{f'_{v|k,i}(1), \dots, f'_{v|k,i}(B)\}, \quad (23)$$

So, under Assumptions 1 and 2, term 1 in (18) becomes

$$\mathbb{E}[\mathbf{u}_{k,i} \Sigma \tilde{\mathbf{w}}_{k-1,i} h_{k,i}] = p_{k,i} \mathbb{E}[\|\tilde{\mathbf{w}}_{k-1,i}\|_{\mathbf{R}_{\mathbf{u}} \Sigma}^2] \quad (24)$$

In (24) $p_{k,i} \triangleq (\mathbb{E}[\mathbf{a}_{k,i}])^T \mathbf{r}_k$ with

$$\mathbf{r}_k = \mathbb{E}[\mathbf{c}'_{k,i}] = \text{col}\left\{\mathbb{E}\left[f'_{v|k,i}(1)\right], \dots, \mathbb{E}\left[f'_{v|k,i}(B)\right]\right\},$$

By taking the square of both sides of (19) and discarding powers of $\epsilon_{k,i}$ higher than 2, and invoking Assumptions 1 and 2 it follows that

$$\begin{aligned} \mathbb{E}[\|\mathbf{u}_{k,i}\|_{\Sigma}^2 h_{k,i}^2] &= \\ &\mathbb{E}[\|\mathbf{u}_{k,i}\|_{\Sigma}^2] \mathbb{E}\left[\left(\sum_{b=1}^B \mathbf{a}_{k,i}(b) f_{v|k,i}(b)\right)^2\right] \\ &+ \mathbb{E}[\|\mathbf{u}_{k,i}\|_{\Sigma}^2 \epsilon_{k,i}^2] \left\{\mathbb{E}\left[\left(\sum_{b=1}^B \mathbf{a}_{k,i}(b) f'_{v|k,i}(b)\right)^2\right]\right. \\ &\left. + \mathbb{E}\left[\sum_{b=1}^B \mathbf{a}_{k,i}(b) f_{v|k,i}(b) + \mathbf{a}_{k,i}(b) f''_{v|k,i}(b)\right]\right\} \quad (25) \end{aligned}$$

where $f''_{v|k,i}(b) \triangleq f''_{k,b}(v_{k,i})$, $b = 1, \dots, B$. Note that (25) can be simplified as

$$\begin{aligned} \mathbb{E}[\|\mathbf{u}_{k,i}\|_{\Sigma}^2 h_{k,i}^2] &= s_{k,i} \text{tr}(\mathbf{R}_{\mathbf{u},k} \Sigma) \\ &+ t_{k,i} \mathbb{E}\left[\|\tilde{\mathbf{w}}_{k-1,i}\|_{\mathbb{E}[\|\mathbf{u}_{k,i}\|_{\Sigma}^2 \mathbf{u}_{k,i}^T \mathbf{u}_{k,i}]}^2\right] \quad (26) \end{aligned}$$

with

$$s_{k,i} = \mathbb{E}[(\mathbf{a}_{k,i}^T \mathbf{c}_{k,i})^2] = \text{tr}(\mathbf{R}_{\mathbf{a}_{k,i}} \mathbf{R}_{\mathbf{c}_k}) \quad (27)$$

$$\begin{aligned} t_{k,i} &= \mathbb{E}[(\mathbf{a}_{k,i}^T \mathbf{c}'_{k,i})^2] + \mathbb{E}[(\mathbf{a}_{k,i}^T \mathbf{c}_{k,i})(\mathbf{a}_{k,i}^T \mathbf{c}''_{k,i})] \\ &= \text{tr}(\mathbf{R}_{\mathbf{a}_{k,i}} \mathbf{R}_{\mathbf{c}'_k} + \mathbf{R}_{\mathbf{a}_{k,i}} \mathbf{R}_{\mathbf{c}''_k}) \quad (28) \end{aligned}$$

where¹

$$\mathbf{c}''_{k,i} \triangleq \text{col}\{f''_{v|k,i}(1), \dots, f''_{v|k,i}(B)\}$$

$$\begin{aligned} \mathbf{R}_{\mathbf{a}_{k,i}} &\triangleq \mathbb{E}[\mathbf{a}_{k,i} \mathbf{a}_{k,i}^T] & \mathbf{R}_{\mathbf{c}_k} &\triangleq \mathbb{E}[\mathbf{c}_{k,i} \mathbf{c}_{k,i}^T] \\ \mathbf{R}_{\mathbf{c}'_k} &\triangleq \mathbb{E}[\mathbf{c}'_{k,i} \mathbf{c}'_{k,i}^T] & \mathbf{R}_{\mathbf{c}''_k} &\triangleq \mathbb{E}[\mathbf{c}''_{k,i} \mathbf{c}''_{k,i}^T] \end{aligned}$$

Using the required moments, (18) changes to

$$\mathbb{E}[\|\tilde{\mathbf{w}}_{k,i}\|_{\Sigma}^2] = \mathbb{E}[\|\tilde{\mathbf{w}}_{k-1,i}\|_{\mathbf{Q}}^2] + \mu_k^2 s_{k,i} \text{tr}(\mathbf{R}_{\mathbf{u},k} \Sigma) \quad (29)$$

¹Note that the subscript i has been dropped in $\mathbf{R}_{\mathbf{c}_k}$, $\mathbf{R}_{\mathbf{c}'_k}$ and $\mathbf{R}_{\mathbf{c}''_k}$ since noise is a wide-sense stationary process.

with the new weighing matrix \mathbf{Q} is given by

$$\mathbf{Q} = \Sigma - 2\mu_k p_{k,i} \mathbf{R}_{\mathbf{u}_k} \Sigma + \mu_k^2 t_{k,i} \mathbb{E}[\|\mathbf{u}_{k,i}\|_{\Sigma}^2 \mathbf{u}_{k,i}^T \mathbf{u}_{k,i}] \quad (30)$$

To further simplify the analysis, the following assumption is considered.

Assumption 3: The regressor vectors $\{\mathbf{u}_{k,i}\}$ are sampled from Gaussian distribution.

This assumption enables us to derive a closed-form for the fourth-order moment $\mathbb{E}\{\|\mathbf{u}_{k,i}\|_{\Sigma}^2 \mathbf{u}_{k,i}^T \mathbf{u}_{k,i}\}$ that appears in (30).

To proceed, let $\mathbf{R}_{\mathbf{u}_k} = \mathbf{U}_k \mathbf{\Gamma}_k \mathbf{U}_k^T$ be the eigendecomposition of $\mathbf{R}_{\mathbf{u}_k}$, where $\mathbf{\Gamma}_k$ is a diagonal matrix with the eigenvalues of $\mathbf{R}_{\mathbf{u}_k}$ and \mathbf{U}_k is a unitary matrix. We further define the following transformed quantities

$$\bar{\mathbf{w}}_{k,i} \triangleq \mathbf{U}_k^T \tilde{\mathbf{w}}_{k,i}, \quad \bar{\mathbf{u}}_{k,i} \triangleq \mathbf{u}_{k,i} \mathbf{U}_k, \quad \bar{\Sigma} \triangleq \mathbf{U}^T \Sigma \mathbf{U}_k \quad (31)$$

It should be noted that for the unitary matrix \mathbf{U}_k we have

$$\|\tilde{\mathbf{w}}_{k,i}\|_{\Sigma}^2 = \|\bar{\mathbf{w}}_{k,i}\|_{\bar{\Sigma}}^2, \quad \|\mathbf{u}_{k,i}\|_{\Sigma}^2 = \|\bar{\mathbf{u}}_{k,i}\|_{\bar{\Sigma}}^2 \quad (32)$$

Using the transformed quantities in (31), expressions in (29) and (30) change to

$$\mathbb{E}[\|\bar{\mathbf{w}}_{k,i}\|_{\bar{\Sigma}}^2] = \mathbb{E}[\|\bar{\mathbf{w}}_{k-1,i}\|_{\bar{\mathbf{Q}}}^2] + \mu_k^2 s_{k,i} \text{tr}(\mathbf{\Gamma}_k \bar{\Sigma}) \quad (33)$$

$$\begin{aligned} \bar{\mathbf{Q}} &= \bar{\Sigma} - 2\mu_k p_{k,i} \mathbf{\Gamma}_k \bar{\Sigma} \\ &+ \mu_k^2 t_{k,i} (\mathbf{\Gamma}_k \text{tr}(\mathbf{\Gamma}_k \bar{\Sigma}) + 2\mathbf{\Gamma}_k \bar{\Sigma} \mathbf{\Gamma}_k) \quad (34) \end{aligned}$$

Note that in (34) we used the following property for Gaussian regressors [3]

$$\mathbb{E}[\|\bar{\mathbf{u}}_{k,i}\|_{\bar{\Sigma}}^2 \bar{\mathbf{u}}_{k,i}^T \bar{\mathbf{u}}_{k,i}] = \mathbf{\Gamma}_k \text{tr}(\mathbf{\Gamma}_k \bar{\Sigma}) + 2\mathbf{\Gamma}_k \bar{\Sigma} \mathbf{\Gamma}_k \quad (35)$$

According to (15) and (16), to derive a general expression for MSD and EMSE, we choose Σ to be a diagonal matrix. In this case, $\bar{\mathbf{Q}}$ will be a diagonal matrix as well. Thus, by applying $\text{diag}\{\cdot\}$ to (33) and (34) we have

$$\mathbb{E}[\|\bar{\mathbf{w}}_{k,i}\|_{\bar{\Sigma}}^2] = \mathbb{E}[\|\bar{\mathbf{w}}_{k-1,i}\|_{\bar{\mathbf{F}}_{k,i}}^2] + \mathbf{g}_k \bar{\sigma}_k \quad (36)$$

$$\bar{\mathbf{F}}_{k,i} = \mathbf{I} - 2\mu_k p_{k,i} \mathbf{\Gamma}_k + 2\mu_k^2 t_{k,i} \mathbf{\Gamma}_k^2 + \mu_k^2 t_{k,i} \mathbf{\Gamma}_k \mathbf{\Gamma}_k^T \quad (37)$$

where $\bar{\sigma} = \text{diag}\{\bar{\Sigma}\}$ and $\mathbf{g}_k = \mu_k^2 s_{k,i} \mathbf{\Gamma}_k^T$.

Expression (37) is a coupled equation that involves both $\bar{\mathbf{w}}_{k,i}$ and $\bar{\mathbf{w}}_{k-1,i}$. By iterating (37) a set of K coupled equalities arise which can be solved to obtain $\{\eta_{k,\infty}, \zeta_{k,\infty}\}$ by choosing $\{\bar{\sigma}_k\}$ and proper manipulation of the equations (see [3] for details). Following the same argument and derivation, the closed-form expressions for steady-state values of MSD and EMSE are obtained. In this way, the following proposition holds.

Proposition 1: (Stead-state performance) Under Assumptions 1-3, the closed-form expressions for steady-state values of MSD and EMSE are

$$\eta_{k,\infty} = \mathbf{z}_k (\mathbf{I} - \mathbf{\Omega}_{k,1})^{-1} (\text{diag}\{\mathbf{I}\}) \quad (38)$$

$$\zeta_{k,\infty} = \mathbf{z}_k (\mathbf{I} - \mathbf{\Omega}_{k,1})^{-1} (\text{diag}\{\mathbf{\Gamma}_k\}) \quad (39)$$

with

$$\begin{aligned}\Omega_{k,\ell} &\triangleq \bar{\mathbf{F}}_{k+l-1,\infty} \bar{\mathbf{F}}_{k+l,\infty} \dots \bar{\mathbf{F}}_{N,\infty} \bar{\mathbf{F}}_{1,\infty} \dots \bar{\mathbf{F}}_{k-1,\infty}, \\ \mathbf{z}_k &\triangleq \mathbf{g}_k \Omega_{k,2} + \mathbf{g}_{k+1} \Omega_{k,3} + \dots + \mathbf{g}_{k-2} \Omega_{k,N} + \mathbf{g}_{k-1}\end{aligned}$$

where the subscripts are all mod K and

$$\bar{\mathbf{F}}_{k,\infty} = \mathbf{I} - 2\mu_k p_{k,\infty} \mathbf{\Gamma}_k + 2\mu_k^2 t_{k,\infty} \mathbf{\Gamma}_k^2 + \mu_k^2 t_{k,\infty} \mathbf{\Gamma}_k \mathbf{\Gamma}_k^\top$$

where

$$\begin{aligned}p_{k,\infty} &= \lim_{i \rightarrow \infty} p_{k,i} = (\mathbb{E}[\mathbf{a}_{k,i}])^\top \mathbf{r}_k \\ s_{k,\infty} &= \lim_{i \rightarrow \infty} s_{k,i} = \text{tr}(\mathbf{R}_{\mathbf{a}_{k,\infty}} \mathbf{R}_{\mathbf{c}_k}) \\ t_{k,\infty} &\triangleq \lim_{i \rightarrow \infty} t_{k,i} = \text{tr}(\mathbf{R}_{\mathbf{a}_{k,\infty}} \mathbf{R}_{\mathbf{c}'_k} + \mathbf{R}_{\mathbf{a}_{k,\infty}} \mathbf{R}_{\mathbf{c}''_k})\end{aligned}$$

In the next section, the accuracy of (38) and (39) are verified through simulation.

C. Stability Analysis

In the sequel, we use recursion (17) to obtain the conditions for mean stability condition.

Proposition 2: (Mean Stability) Let the linear model (1) and Assumption 1-3 hold. Then, the proposed algorithm is asymptotically unbiased if, and only if, the following condition holds

$$|1 - \mu_k p_{k,i} \text{tr}(\mathbf{R}_{\mathbf{u}_k})| < 1 \quad (40)$$

Proof: Taking the expectation of (17) gives

$$\mathbb{E}[\tilde{\mathbf{w}}_{k,i}] = \mathbb{E}[\tilde{\mathbf{w}}_{k-1,i}] - \mu_k \mathbb{E}[\mathbf{u}_{k,i}^\top h_{k,i}] \quad (41)$$

Replacing $\epsilon_{k,i}$ and $h_{k,i}$ from (13) and (19), and applying Assumptions 1 and 2 we have

$$\mathbb{E}[\mathbf{u}_{k,i}^\top h_{k,i}] = p_{k,i} \mathbf{R}_{\mathbf{u}_k} \mathbb{E}[\tilde{\mathbf{w}}_{k-1,i}] \quad (42)$$

So, (41) becomes

$$\mathbb{E}[\tilde{\mathbf{w}}_{k,i}] = (\mathbf{I} - \mu_k p_{k,i} \mathbf{R}_{\mathbf{u}_k}) \mathbb{E}[\tilde{\mathbf{w}}_{k-1,i}] \quad (43)$$

Iterating (43) gives the following equation that explains how $\mathbb{E}[\tilde{\mathbf{w}}_{k,i}]$ evolves over time:

$$\begin{aligned}\mathbb{E}[\tilde{\mathbf{w}}_{k,i}] &= \left(\prod_{\ell=1}^K (\mathbf{I} - \mu_\ell p_{\ell,i} \mathbf{R}_{\mathbf{u}_\ell}) \right) \mathbb{E}[\tilde{\mathbf{w}}_{k,i-1}] \\ &= \left(\prod_{\ell=1}^K \mathbf{G}_\ell \right) \mathbb{E}[\tilde{\mathbf{w}}_{k,i-1}] = \mathcal{G} \mathbb{E}[\tilde{\mathbf{w}}_{k,i-1}]\end{aligned} \quad (44)$$

Thus, a necessary and sufficient condition for robust incremental algorithm to converge in the mean is to have matrix \mathcal{G} stable, or equivalently, all its eigenvalues inside the unit circle, i.e. $\rho(\mathcal{G}) < 1$. As the spectral radius of a matrix is bounded by its norms (for any induced matrix norm), i.e. $\rho(\mathcal{G}) < \|\mathcal{G}\|$, we have

$$\begin{aligned}\rho(\mathcal{G}) &\leq \|\mathcal{G}\| \leq \|\mathbf{G}_1\| \|\mathbf{G}_2\| \dots \|\mathbf{G}_K\| \\ &= \rho(\mathbf{G}_1) \rho(\mathbf{G}_2) \dots \rho(\mathbf{G}_K)\end{aligned}$$

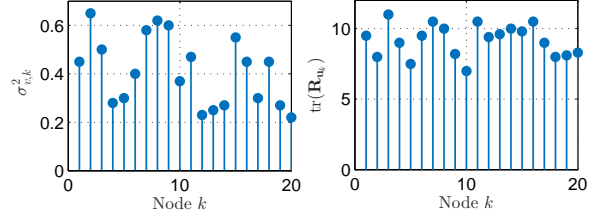


Fig. 2. $\sigma_{v,k}^2$ and $\text{tr}(\mathbf{R}_{\mathbf{u}_k})$ for every node.

Note that, the last equality stems from the fact that every \mathbf{G}_k is a Hermitian matrix (due to Assumption 2 and $p_{k,k} > 0$), so its 2-induced norm agrees with its spectral radius. So if $\rho(\mathbf{G}_k) \leq 1$ is established, the constraint $\rho(\mathcal{G}) < 1$ will also be satisfied. On the other hand, $\rho(\mathbf{G}_k) \leq 1$ holds when $|1 - \mu_k p_{k,i} \text{tr}(\mathbf{R}_{\mathbf{u}_k})| < 1$. ■

IV. SIMULATION RESULTS

The simulation results are presented to evaluate the performance of proposed algorithm and the accuracy of theoretical expressions (38) and (39). We consider a network with $K = 20$ nodes, seeking to estimate $\mathbf{w}^o = [11111]^\top$ ($M = 5$). The regressors $\{\mathbf{u}_{k,i}\}$ are independently generated from a multivariate zero-mean Gaussian distribution with diagonal covariances $\mathbf{R}_{\mathbf{u}_k}$. The noise samples are independently generated according to an γ -contaminated Gaussian mixture model with the following PDF

$$p_{v_k}(v) = (1 - \gamma) \mathcal{N}(0, \sigma_{v,k}^2) + \gamma \mathcal{N}(0, \kappa \sigma_{v,k}^2) \quad (45)$$

where $\{\sigma_{v,k}^2\}$ are the nominal noise variances where $\kappa = 100$. The regressor covariance traces and nominal noise variances are shown in Fig. 2. In the simulations we set $\mu_k = 0.01$, $\gamma = 0.1$, $\nu_k = 0.9$ and $\beta = 10^{-6}$. For the proposed algorithm, we set $B = 2$ with the following basis functions: $f_{k,1}(x) = x$ and $f_{k,2}(x) = \tanh(x)$. Each element of the initial estimates of basis weights, $\mathbf{a}_{k,0}$ is set to 0.5. For the smoothing recursions, zero initial conditions are assumed, and ν_k is set to 0.9 for every node K . The moments $\mathbb{E}[\mathbf{a}_{k,\infty}]$ and $\mathbf{R}_{\mathbf{a}_{k,\infty}}$ are approximated with Mont Carlo (MC) simulation (averaging over the last 100 samples and across trials). The steady-state values of MSD and EMSE are obtained by running the algorithms for 2500 iterations and then averaging over 100 independent trials. Fig. 3 shows the steady-state values of MSD and EMSE for the incremental LMS algorithm and the proposed algorithm. Both theoretical values (using expressions (38) and (39)) and simulated values are compared. It is clear that the proposed algorithm outperforms the incremental LMS algorithm. Moreover, there is a good match between simulations and theory. Table I shows the average steady-state MSD values over network for different algorithms

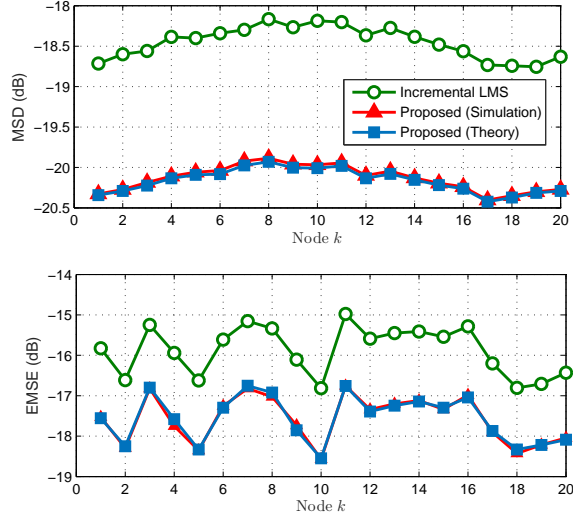


Fig. 3. Simulated and theoretical steady-state values for the incremental LMS and the proposed algorithm.

TABLE I

AVERAGE STEADY-STATE MSD VALUES OVER NETWORK FOR DIFFERENT ALGORITHMS IN TWO DIFFERENT CONDITIONS.

Method	Proper tuning	Improper tuning
[3]	-18.61	32.1
[10]	-20.07	29.7
[12]	-20.23	33.4
[13]	-20.29	25.4
Proposed	-20.31	-

in two different conditions including proper parameter setting and improper parameter setting. As we can see, unlike other algorithms, the performance of the proposed algorithm does not require precise parameter setting.

V. CONCLUSIONS

In this paper we introduced a robust adaptive algorithm that is suitable for parameter estimation in Hamiltonian Networks where the measurement are corrupted by heavy-tailed impulsive noise. The main advantage of the proposed algorithm is that it tunes the parameters of the optimal error-nonlinearity and estimates the unknown parameter at the same time. Therefore, it does not require a priori knowledge about the noise probability density function, which is always required in similar methods. The provided steady-state analysis has been validated through simulations.

REFERENCES

- [1] M. Ekman, K. Davstad, and L. Sjöberg, "Ground target tracking using acoustic sensors," in *2007 Information, Decision and Control*, Feb 2007, pp. 182–187.
- [2] S. Tu and A. H. Sayed, "Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6217–6234, Dec 2012.
- [3] C. G. Lopes and A. H. Sayed, "Incremental adaptive strategies over distributed networks," *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 4064–4077, Aug 2007.
- [4] M. O. B. Saeed and A. Zerguine, "An incremental variable step-size lms algorithm for adaptive networks," *IEEE Transactions on Circuits and Systems II: Express Briefs*, pp. 1–1, 2019.
- [5] F. S. Cattivelli and A. H. Sayed, "Diffusion lms strategies for distributed estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1035–1048, March 2010.
- [6] J. Chen and A. H. Sayed, "Diffusion adaptation strategies for distributed optimization and learning over networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 8, pp. 4289–4305, Aug 2012.
- [7] K. Yuan, B. Ying, X. Zhao, and A. H. Sayed, "Exact diffusion for distributed optimization and learning part i: Algorithm development," *IEEE Transactions on Signal Processing*, vol. 67, no. 3, pp. 708–723, Feb 2019.
- [8] A. Rastegarnia, "Reduced-communication diffusion rls for distributed estimation over multi-agent networks," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 67, no. 1, pp. 177–181, Jan 2020.
- [9] S. Al-Sayed, A. M. Zoubir, and A. H. Sayed, "Robust adaptation in impulsive noise," *IEEE Transactions on Signal Processing*, vol. 64, no. 11, pp. 2851–2865, June 2016.
- [10] T. Panigrahi, G. Panda, B. Mulgrew, and B. Majhi, "Robust incremental lms over wireless sensor network in impulsive noise," in *2010 International Conference on Computational Intelligence and Communication Networks*, Nov 2010, pp. 205–209.
- [11] T. Panigrahi, G. Panda, and B. Mulgrew, "Robust distributed block lms over wsn in impulsive noise," in *Distributed Computing and Internet Technology*, R. Ramanujam and S. Ramaswamy, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 261–262.
- [12] M. K. Islam, A. Rastegarnia, and A. Khalili, "A robust distributed estimation algorithm under alpha-stable noise condition," *Journal of Communication Engineering*, vol. 4, no. 2, pp. 76–85, 2015.
- [13] R. Bao, H. Rong, J. Yang, and B. Chen, "An incremental self-learning algorithm with robustness against impulsive noise," in *2018 IEEE 20th International Conference on High Performance Computing and Communications*, June 2018, pp. 1620–1626.
- [14] W. Ma, B. Chen, J. Duan, and H. Zhao, "Diffusion maximum correntropy criterion algorithms for robust distributed estimation," *Digital Signal Processing*, vol. 58, pp. 10 – 19, 2016.
- [15] W. Huang, L. Li, Q. Li, and X. Yao, "Diffusion robust variable step-size lms algorithm over distributed networks," *IEEE Access*, vol. 6, pp. 47 511–47 520, 2018.
- [16] X. Li, Q. Shi, S. Xiao, S. Duan, and F. Chen, "A robust diffusion minimum kernel risk-sensitive loss algorithm over multitask sensor networks," *Sensors*, vol. 19, no. 10, 2019.
- [17] A. M. Wilson, T. Panigrahi, and A. Dubey, "Robust distributed lorentzian adaptive filter with diffusion strategy in impulsive noise environment," *Digital Signal Processing*, vol. 96, p. 102589, 2020.
- [18] Q. Shi, F. He, J. Wu, and F. Chen, "Distributed adaptive clustering based on maximum correntropy criterion over dynamic multi-task networks," *IEEE Access*, vol. 8, pp. 12 402–12 412, 2020.
- [19] S. Al-Sayed, A. M. Zoubir, and A. H. Sayed, "Robust distributed estimation by networked agents," *IEEE Transactions on Signal Processing*, vol. 65, no. 15, pp. 3909–3921, Aug 2017.
- [20] F. S. Cattivelli and A. H. Sayed, "Analysis of spatial and incremental lms processing for distributed estimation," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1465–1480, 2011.
- [21] A. H. Sayed, *Adaptive Filters*. Wiley-IEEE Press, 2008.