Transient Analysis of Diffusion Recursive Least Squares via Signed Error Algorithm for Cyclostationary Colored Inputs

Wei Gao, Yi Xu, Jie Chen, Cédric Richard

TRANSIENT ANALYSIS OF DIFFUSION RECURSIVE LEAST SQUARES VIA SIGNED ERROR ALGORITHM FOR CYCLOSTATIONARY COLORED INPUTS

Wei Gao † *Yi Xu* † *Jie Chen* ⋆ *Cédric Richard* ‡

†School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang 212013, China

⋆School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China

‡Université Côte d'Azur, OCA, CNRS, 06108 Nice, France

Email: wei_gao@ujs.edu.cn 2211808020@stmail.ujs.edu.cn dr.jie.chen@ieee.org cedric.richard@unice.fr

ABSTRACT

In this paper, we perform the transient theoretical analysis of diffusion recursive least squares via signed error (DRLS-SE) algorithm over networks in the presence of impulsive noise. The obtained analytical models allow us to investigate the impacts of nonstationary system and cyclostationary colored inputs on the network transient convergence behavior. Simulations are provided to highlight the robustness of DRLS-SE algorithm against impulsive noise, and corroborate the correctness and accuracy of obtained theoretical findings.

***Index Terms*—** Diffusion RLS, signed error, transient theoretical analysis, cyclostationary colored inputs, impulsive noise

1. INTRODUCTION

Adaptive networks have been introduced in recent years and intensively studied within the contexts of estimation [1,2], inference [3,4], tracking [5,6], and active noise control application [7], to cite a few. It is well known that the recursive least squares (RLS) algorithm possesses inherent merits of much faster convergence rate and smaller steady-state misadjustment error even for colored (correlated) input signals compared with that of the least-mean-square (LMS) algorithm [8,9]. Consequently, there have been considerable research efforts on developing distributed RLS-type algorithms over networks using the diffusion adaptation.

The diffusion RLS (DRLS) algorithm via incremental update was primitively proposed to address the problem of distributed estimation in [10]. The steady-state mean-squared convergence performance of the adaptive network was also analyzed in this paper. Diffusion adaptation was applied with the bias-compensated RLS algorithm in [11] in order to reduce the residual bias. The authors also derived closed-form expressions describing the steady-state mean and mean-square performances. Several distributed sparse RLS algorithms were presented in [12] with their analyses at steady-state in the mean and mean-square sense. Specifically, the distributed sparse multitask RLS problem over networks was recently studied in [13]. Additionally, the partial diffusion recursive least squares (PDRLS) algorithm was developed in [14] to reach a trade-off between estimation accuracy and communication burden. Furthermore, the convergence performance of the PDRLS algorithm was analyzed in the mean and mean-square sense by using the energy conservation principle. More recently, the reduced-communication DRLS algorithm and its steady-state analytical models were provided in [15].

This work was supported in part by the National Natural Science Foundation of China Grants (62171205, 62171380).

Cyclostationary signals with periodical variation widely exist in real-world applications [16, 17]. As a consequence, the theoretical performance of many classical adaptive filtering algorithms have been extensively studied within this context [18–20]. In particular, the convergence behavior of the diffusion LMS (DLMS) was analyzed for cyclostationary inputs in [21–23]. On the other hand, it was shown that DRLS-type algorithms are readily interfered by impulsive noises which can lead to severe performance degradation. To circumvent this problem, the authors in [24] devised robust DRLS algorithms with side information to operate in environments subject to impulsive disturbances. They also presented mean-square analyses at steady-state, as well as transient semi-analytic models involving terms that were not explicitly evaluated. In this paper, we present a simple and robust diffusion recursive least squares algorithm based on signed error (DRLS-SE) that can cope with impulsive noise. Inspired by recent works on RLS algorithm [25, 26], we analyze the transient behavior of the DRLS-SE for nonstationary systems and cyclostationary colored inputs corrupted by impulsive noise. Actually, those scenarios can be used to describe a large range of real-world systems. Simulations support the robust performance of DRLS-SE algorithm, as well as the accuracy of the analytical models derived for characterizing the network transient behavior.

Notation: Matrices $\boldsymbol{I}_N$ and $\mathbf{1}_N$ denote the $N \times N$ identity matrix and the $N \times N$ matrix with all its entries equal to 1, respectively. Notation $\otimes$ denotes Kronecker product. The operator bdiag$\{\cdot\}$ formulates its arguments as a (block) diagonal matrix, and col$\{\cdot\}$ stacks its vector arguments on top of each other to generate a column vector. The operator tr$\{\cdot\}$ represents the matrix trace, and sgn$\{\cdot\}$ represents the signum function.

2. PRELIMINARIES AND DRLS-SE ALGORITHM

We consider an adaptive network consisting of K nodes and some communication links between these nodes. We assume that every node k has access to the input-output data pairs $\{\boldsymbol{x}_k(n), d_k(n)\}_{n=1}^N$, where $\boldsymbol{x}_k(n) = [x_k(n), x_k(n-1), \ldots, x_k(n-L+1)]^\top$ is the input regression vector and $d_k(n)$ is the desired scalar output assumed to be zero-mean. At each time instant $n \geq 0$ and for each node k, $d_k(n)$ is assumed to be generated from the input vector $\boldsymbol{x}_k(n)$ passing through a linear regression model that varies over time, corrupted by an impulsive noise, that is,

$$d_k(n) = \boldsymbol{x}_k^\top(n)\boldsymbol{w}^\star(n) + z_k(n) \tag{1}$$

where $\boldsymbol{w}^\star(n) \in \mathbb{R}^L$ is an unknown optimal time-varying weight vector to be estimated. Here, $z_k(n)$ represents the additive observa-

978-1-6654-8547-0/22//$31.00 ©2022 IEEE

tion noise with impulsive interference that is often modeled as the contaminated-Gaussian (CG) noise [22]:

$$z_k(n) = v_k(n) + b_k(n)\epsilon_k(n) \tag{2}$$

where $v_k(n)$ and $\epsilon_k(n)$ are temporally and spatially independent zero-mean Gaussian noises with variances $\sigma^2_{v,k}$ and $\sigma^2_{\epsilon,k} = \zeta\sigma^2_{v,k}$, with $\zeta \gg 1$. Moreover, $b_k(n)$ is a Bernoulli random variable with probabilities $\Pr(b_k = 1) = p_r$ and $\Pr(b_k = 0) = 1 - p_r$. We assume further that $z_k(n)$ is temporally and spatially independent of any other signals.

In order to examine the tracking ability of DRLS-SE algorithm for linear nonstationary systems, the random walk process is usually used to model the slowly time-varying optimal weight vector $\boldsymbol{w}^\star(n)$ given by (1) for all nodes in networks [8, 9]:

$$\boldsymbol{w}^\star(n+1) = \boldsymbol{w}^\star(n) + \boldsymbol{q}(n) \tag{3}$$

with $\boldsymbol{q}(n) \in \mathbb{R}^L$, which is a zero-mean white Gaussian random perturbation vector with covariance matrix $\mathbb{E}\{\boldsymbol{q}(n)\boldsymbol{q}^\top(n)\} = \sigma_q^2\boldsymbol{I}_L$. Moreover, $\boldsymbol{q}(n)$ is assumed to be temporally independent of $\boldsymbol{x}_k(n)$ and $z_k(n)$. Each node k of the network seeks to estimate an identical unknown parameter vector $\boldsymbol{w}^\star(n)$ in a collaborative manners.

In addition, we also consider the nonstationary input regression vector $\boldsymbol{x}_k(n)$ in (1), which is modeled as a colored random process with the periodically time-varying variance defined as [25]:

$$x_k(n) = \sigma_{x,k}(n)\, u_k(n) \tag{4}$$

where $\sigma_{x,k}(n)$ is defined as a deterministic sequence with repetition period T, and $u_k(n)$ is a colored random sequence generated from a Gaussian distribution. Correspondingly, the time-varying autocorrelation matrix of the cyclostationary colored input vector $\boldsymbol{x}_k(n)$ can be constructed as [25]:

$$\boldsymbol{R}_{x,k}(n) = \mathbb{E}\{\boldsymbol{x}_k(n)\boldsymbol{x}_k^\top(n)\} = \boldsymbol{\Sigma}_{x,k}\boldsymbol{R}_{u,k}\boldsymbol{\Sigma}_{x,k} \tag{5}$$

with the diagonal matrix $\boldsymbol{\Sigma}_{x,k} = \text{bdiag}\big\{\sigma_{x,k}(n), \ldots, \sigma_{x,k}(n-L+1)\big\}$, and the autocorrelation matrix $\boldsymbol{R}_{u,k} = \mathbb{E}\big\{\boldsymbol{u}_k(n)\boldsymbol{u}_k^\top(n)\big\}$. The sinusoidal variation model [18–20, 25], is adopted for periodic sequences $\sigma^2_{x,k}(n)$ to survey the impact of cyclostationary colored inputs on the transient convergence behavior of DRLS-SE algorithm.

Without loss of generality, we only focus on the DRLS-SE algorithm with adapt-then-combine (ATC) diffusion strategy. Define the local estimate $\boldsymbol{w}_k(n)$ and the intermediate estimate $\boldsymbol{\psi}_k(n)$ of $\boldsymbol{w}^\star(n)$, respectively. In the first adaptation step, we present the recursive update relation of $\boldsymbol{w}_k(n)$ and $\boldsymbol{\psi}_k(n+1)$ for each node k:

$$\boldsymbol{\psi}_k(n+1) = \boldsymbol{w}_k(n) + \boldsymbol{P}_k(n+1)\boldsymbol{x}_k(n)\text{sgn}\big\{e_k(n)\big\} \tag{6}$$

with the inverse autocorrelation matrix of input data

$$\boldsymbol{P}_k(n+1) = \boldsymbol{\Phi}_k^{-1}(n+1) \tag{7}$$

and the instantaneous estimation error of node k at time instant n

$$e_k(n) = d_k(n) - \boldsymbol{x}_k^\top(n)\boldsymbol{w}_k(n). \tag{8}$$

Here, $\boldsymbol{\Phi}_k(n+1)$ is the time-averaged correlation matrix of input data for node k defined by [8, 9]

$$\boldsymbol{\Phi}_k(n+1) = \sum_{i=0}^{n+1}\lambda^{n+1-i}\boldsymbol{x}_k(i)\boldsymbol{x}_k^\top(i) = \lambda\boldsymbol{\Phi}_k(n) + \boldsymbol{x}_k(n+1)\boldsymbol{x}_k^\top(n+1) \tag{9}$$

with $0 \ll \lambda < 1$ denoting the forgetting factor. The initial condition of $\boldsymbol{\Phi}_k(n)$ is given by $\boldsymbol{\Phi}_k(0) = \delta\boldsymbol{I}_L$ with a small positive value δ. As given in (6), the signed error leads to the robustness of suppressing the impulsive noise. Applying the matrix inversion lemma to the right hand side (r.h.s.) of (9), the update equation of the inverse of the autocorrelation matrix at node k is given by [8, 9]:

$$\boldsymbol{P}_k(n+1) = \lambda^{-1}\left[\boldsymbol{P}_k(n) - \frac{\lambda^{-1}\boldsymbol{P}_k(n)\boldsymbol{x}_k(n)\boldsymbol{x}_k^\top(n)\boldsymbol{P}_k(n)}{1+\lambda^{-1}\boldsymbol{x}_k^\top(n)\boldsymbol{P}_k(n)\boldsymbol{x}_k(n)}\right] \tag{10}$$

where matrix $\boldsymbol{P}_k(n)$ is initialized by $\boldsymbol{P}_k(0) = \delta^{-1}\boldsymbol{I}_L$.

In the second combination step, we attempt to further improve the estimation precision of intermediate estimate $\boldsymbol{\psi}_k(n)$ for each node k by sharing local data within its neighborhood. Therefore, the weight vector combination is given by [26]:

$$\boldsymbol{w}_k(n+1) = \sum_{\ell\in\mathcal{N}_k} a_{\ell k}\,\boldsymbol{\psi}_\ell(n+1). \tag{11}$$

The set $\{a_{\ell k}\}$ for $k = 1, 2, \ldots, K$ are nonnegative combination coefficients, which satisfy $a_{\ell k} \geq 0$, $\sum_{\ell=1}^K a_{\ell k} = 1$, and $a_{\ell k} = 0$, if $\ell \notin \mathcal{N}_k$, where neighborhood $\mathcal{N}_k$ denotes the set of nodes connected to node k, including itself. Moreover, the coefficient $a_{\ell k}$ is the (ℓ, k)-th entry of a left-stochastic matrix $\boldsymbol{A}$, i.e., $\boldsymbol{A}^\top\boldsymbol{1}_K = \boldsymbol{1}_K$. For simplicity, we refer to the DRLS-SE algorithm with ATC diffusion strategy as DRLS-SE algorithm hereafter.

3. TRANSIENT ANALYSIS OF DRLS-SE ALGORITHM WITH ATC DIFFUSION STRATEGY

In this section, we study the stochastic behavior of DRLS-SE algorithm in the mean and mean-square error senses. We need to introduce the intermediate weight error vector and the weight error vector for node k at time instant n, namely

$$\widetilde{\boldsymbol{\psi}}_k(n) = \boldsymbol{\psi}_k(n) - \boldsymbol{w}^\star(n), \quad \widetilde{\boldsymbol{w}}_k(n) = \boldsymbol{w}_k(n) - \boldsymbol{w}^\star(n). \tag{12}$$

Furthermore, let $\widetilde{\boldsymbol{\psi}}(n)$ and $\widetilde{\boldsymbol{w}}(n)$ denote the block weight error vectors by collecting (12) for all nodes as follows:

$$\widetilde{\boldsymbol{\psi}}(n) = \text{col}\big\{\widetilde{\boldsymbol{\psi}}_1(n), \ldots, \widetilde{\boldsymbol{\psi}}_K(n)\big\} \in \mathbb{R}^{KL\times 1}, \tag{13}$$

$$\widetilde{\boldsymbol{w}}(n) = \text{col}\big\{\widetilde{\boldsymbol{w}}_1(n), \ldots, \widetilde{\boldsymbol{w}}_K(n)\big\} \in \mathbb{R}^{KL\times 1}. \tag{14}$$

In order to simplify the notation, let us now introduce the following $K \times K$ block diagonal matrices with individual entries of size $L \times L$:

$$\boldsymbol{R}_x(n) = \text{bdiag}\big\{\boldsymbol{R}_{x,1}(n), \ldots, \boldsymbol{R}_{x,K}(n)\big\} \in \mathbb{R}^{KL\times KL}, \tag{15}$$

$$\boldsymbol{\Phi}(n) = \text{bdiag}\big\{\boldsymbol{\Phi}_1(n), \ldots, \boldsymbol{\Phi}_K(n)\big\} \in \mathbb{R}^{KL\times KL}, \tag{16}$$

$$\boldsymbol{P}(n) = \text{bdiag}\big\{\boldsymbol{P}_1(n), \ldots, \boldsymbol{P}_K(n)\big\} \in \mathbb{R}^{KL\times KL}, \tag{17}$$

and the $K \times 1$ block column vector with vectors of length L:

$$\begin{aligned}\boldsymbol{t}(n) = \text{col}\big\{&\text{sgn}\big\{\boldsymbol{x}_1^\top(n)\widetilde{\boldsymbol{w}}_1(n) - z_1(n)\big\}\boldsymbol{x}_1(n), \ldots, \\ &\text{sgn}\big\{\boldsymbol{x}_K^\top(n)\widetilde{\boldsymbol{w}}_K(n) - z_K(n)\big\}\boldsymbol{x}_K(n)\big\} \in \mathbb{R}^{KL\times 1},\end{aligned} \tag{18}$$

$$\boldsymbol{g}(n) = \text{col}\big\{\boldsymbol{q}(n), \ldots, \boldsymbol{q}(n)\big\} \in \mathbb{R}^{KL\times 1}. \tag{19}$$

Before proceeding, we introduce two important assumptions and lemma as follows.

A1. The input data vectors $\boldsymbol{x}_k(n)$ are zero-mean and spatially independent signals.

A2. The weight error vector $\widetilde{\boldsymbol{w}}_k(n)$ is temporally independent of the input data vectors $\boldsymbol{x}_k(n)$.

These two assumptions are reasonable and classical, moreover, they were successfully used and verified in [21].

Lemma 1 Assume that x_1 and x_2 are zero-mean jointly Gaussian random variables. Given the impulsive CG noise defined by (2), and let $y = x_1 + z_k(n)$, it holds that [27]

$$\mathbb{E}\big\{\mathrm{sgn}\{y\}x_2\big\} = (1-p_r)\,\mathbb{E}\big\{\mathrm{sgn}\{y_1\}x_2\big\} + p_r\,\mathbb{E}\big\{\mathrm{sgn}\{y_2\}x_2\big\}$$

with $y_1 = x_1 + v_k(n)$ and $y_2 = x_1 + v_k(n) + \epsilon_k(n)$. This lemma was also used in the theoretical analysis works of diffusion sign algorithms [22, 24].

3.1. Mean Weight Error Analysis

Based on the definition (7), then (9) becomes

$$\boldsymbol{P}_k^{-1}(n+1) = \lambda \boldsymbol{P}_k^{-1}(n) + \boldsymbol{x}_k(n+1)\boldsymbol{x}_k^\top(n+1). \tag{20}$$

Using (5) and the fact that $0 \ll \lambda < 1$, the steady-state expectation of $\boldsymbol{P}_k^{-1}(n)$ can be approximately computed as follows [11, 14, 15]:

$$\begin{aligned}\lim_{n\to\infty}\mathbb{E}\big\{\boldsymbol{P}_k^{-1}(n)\big\} &\approx \lim_{n\to\infty}\sum_{i=0}^{n+1}\lambda^{n+1-i}\mathbb{E}\big\{\boldsymbol{x}_k(i)\boldsymbol{x}_k^\top(i)\big\}\\ &= (1-\lambda)^{-1}\boldsymbol{R}_{x,k}(n).\end{aligned} \tag{21}$$

Note that the steady-state expected value of $\boldsymbol{P}_k^{-1}(n)$ is deterministic periodic due to the periodic variation of $\boldsymbol{R}_{x,k}(n)$. Considering assumption A1 and using (15) and (21), thus the steady-state expected value of (17) can be approximately calculated as

$$\lim_{n\to\infty}\mathbb{E}\big\{\boldsymbol{P}(n)\big\} \approx \lim_{n\to\infty}\big[\mathbb{E}\big\{\boldsymbol{P}^{-1}(n)\big\}\big]^{-1} = (1-\lambda)\,\boldsymbol{R}_x^{-1}(n). \tag{22}$$

Collecting both sides of (9) for all nodes and using (16) to get the block diagonal matrix of size $KL \times KL$, then taking the expectation of both sides and using (15), it leads to

$$\mathbb{E}\big\{\boldsymbol{\Phi}(n+1)\big\} = \lambda\mathbb{E}\big\{\boldsymbol{\Phi}(n)\big\} + \boldsymbol{R}_x(n) \tag{23}$$

with the initial condition $\mathbb{E}\big\{\boldsymbol{\Phi}(0)\big\} = \delta\boldsymbol{I}_{KL}$. The recursive update relation (23) is very useful and crucial in the following theoretical analysis. Replacing (1) into (8) and using (12), the instantaneous estimation error can be alternatively rewritten as

$$e_k(n) = z_k(n) - \boldsymbol{x}_k^\top(n)\widetilde{\boldsymbol{w}}_k(n). \tag{24}$$

Subtracting (3) from both sides of the recursive update equation (6), and according to Eqs. (12) and (24), leads to the intermediate weight error vector update equation:

$$\begin{aligned}\widetilde{\boldsymbol{\psi}}_k(n+1) = \widetilde{\boldsymbol{w}}_k(n) &- \boldsymbol{P}_k(n+1)\boldsymbol{x}_k(n)\\ &\times \mathrm{sgn}\big\{\boldsymbol{x}_k^\top(n)\widetilde{\boldsymbol{w}}_k(n) - z_k(n)\big\} - \boldsymbol{q}(n).\end{aligned} \tag{25}$$

Subtracting $\boldsymbol{w}^\star(n+1)$ from both sides of combination relation (11), and using (12), we have the weight error vector:

$$\widetilde{\boldsymbol{w}}_k(n+1) = \sum_{\ell\in\mathcal{N}_k} a_{\ell k}\,\widetilde{\boldsymbol{\psi}}_\ell(n+1). \tag{26}$$

Substituting (25) into (26), and using the previously introduced expressions (13), (14) and (17)–(19), then the recursive update equation for the global weight error vector can be formulated as

$$\widetilde{\boldsymbol{w}}(n+1) = \boldsymbol{\mathcal{A}}\big[\widetilde{\boldsymbol{w}}(n) - \boldsymbol{P}(n+1)\boldsymbol{t}(n) - \boldsymbol{g}(n)\big] \tag{27}$$

with the matrix $\boldsymbol{\mathcal{A}} = \boldsymbol{A}^\top \otimes \boldsymbol{I}_L$. Pre-multiplying both sides of (27) by $\boldsymbol{P}(n+1)^{-1}\boldsymbol{\mathcal{A}}^{-1}$, it follows from (7) that

$$\boldsymbol{\Phi}(n+1)\boldsymbol{\mathcal{A}}^{-1}\widetilde{\boldsymbol{w}}(n+1) = \boldsymbol{\Phi}(n+1)\widetilde{\boldsymbol{w}}(n) - \boldsymbol{t}(n) - \boldsymbol{\Phi}(n+1)\boldsymbol{g}(n). \tag{28}$$

Taking the expectation on both sides of (28) and applying the statistical properties of perturbation vector $\boldsymbol{q}(n)$, we have

$$\mathbb{E}\big\{\boldsymbol{\Phi}(n+1)\boldsymbol{\mathcal{A}}^{-1}\widetilde{\boldsymbol{w}}(n+1)\big\} = \mathbb{E}\big\{\boldsymbol{\Phi}(n+1)\widetilde{\boldsymbol{w}}(n)\big\} - \mathbb{E}\big\{\boldsymbol{t}(n)\big\}. \tag{29}$$

As given in (18), the k-th subvector of $\mathbb{E}\big\{\boldsymbol{t}(n)\big\}$ is the $L\times 1$ column vector given by

$$\mathbb{E}\big\{\boldsymbol{t}_k(n)\big\} = \mathbb{E}\big\{\mathrm{sgn}\big\{\boldsymbol{x}_k^\top(n)\widetilde{\boldsymbol{w}}_k(n) - z_k(n)\big\}\boldsymbol{x}_k(n)\big\}. \tag{30}$$

More specifically, the i-th element of $\mathbb{E}\big\{\boldsymbol{t}_k(n)\big\}$ is given by

$$\big[\mathbb{E}\big\{\boldsymbol{t}_k(n)\big\}\big]_i = \mathbb{E}\big\{\mathrm{sgn}\big\{\boldsymbol{x}_k^\top(n)\widetilde{\boldsymbol{w}}_k(n) - z_k(n)\big\}x_k(n-i+1)\big\}. \tag{31}$$

We now evaluate (31) under assumptions A1 and A2. Taking the conditional expectation of (31) on $\widetilde{\boldsymbol{w}}_k(n)$, applying lemma 1 and Price's theorem [28], then taking the expectation again, we obtain

$$\begin{aligned}\big[\mathbb{E}\big\{\boldsymbol{t}_k(n)\big\}\big]_i &= \mathbb{E}\Big\{\mathrm{sgn}\big\{\boldsymbol{x}_k^\top(n)\widetilde{\boldsymbol{w}}_k(n) - z_k(n)\big\}x_k(n-i+1)\big|\widetilde{\boldsymbol{w}}_k(n)\Big\}\\ &= \sqrt{\frac{2}{\pi}}\big[\boldsymbol{R}_{x,k}(n)\big]_i^\top\mathbb{E}\big\{\widetilde{\boldsymbol{w}}_k(n)\big\}\Bigg[\frac{1-p_r}{\sqrt{\sigma_{v,k}^2 + \widetilde{\boldsymbol{w}}_k^\top(n)\boldsymbol{R}_{x,k}(n)\widetilde{\boldsymbol{w}}_k(n)}}\\ &\quad + \frac{p_r}{\sqrt{(\zeta+1)\sigma_{v,k}^2 + \widetilde{\boldsymbol{w}}_k^\top(n)\boldsymbol{R}_{x,k}(n)\widetilde{\boldsymbol{w}}_k(n)}}\Bigg].\end{aligned} \tag{32}$$

The random variable in the denominators of two terms on the r.h.s. of (32) can be reasonably approximated as [19]

$$\begin{aligned}\widetilde{\boldsymbol{w}}_k^\top(n)\boldsymbol{R}_{x,k}(n)\widetilde{\boldsymbol{w}}_k(n) &\approx \mathbb{E}\big\{\widetilde{\boldsymbol{w}}_k^\top(n)\boldsymbol{R}_{x,k}(n)\widetilde{\boldsymbol{w}}_k(n)\big\}\\ &= \mathrm{tr}\big\{\boldsymbol{K}_k(n)\boldsymbol{R}_{x,k}(n)\big\}\end{aligned} \tag{33}$$

with the autocorrelation matrix of weight error vector at node k, i.e., $\boldsymbol{K}_k(n) = \mathbb{E}\big\{\widetilde{\boldsymbol{w}}_k(n)\widetilde{\boldsymbol{w}}_k^\top(n)\big\}$. Substituting (33) into (32), and collecting all the elements into a column vector, (30) can be evaluated as

$$\mathbb{E}\big\{\boldsymbol{t}_k(n)\big\} = \sqrt{2/\pi}\,\theta_k\boldsymbol{R}_{x,k}(n)\mathbb{E}\big\{\widetilde{\boldsymbol{w}}_k(n)\big\} \tag{34}$$

with

$$\theta_k = \frac{1-p_r}{\sqrt{\sigma_{v,k}^2 + \mathrm{tr}\big\{\boldsymbol{K}_k(n)\boldsymbol{R}_{x,k}(n)\big\}}} + \frac{p_r}{\sqrt{(\zeta+1)\sigma_{v,k}^2 + \mathrm{tr}\big\{\boldsymbol{K}_k(n)\boldsymbol{R}_{x,k}(n)\big\}}}. \tag{35}$$

In view of (34), the expectation of (18) can be evaluated as

$$\mathbb{E}\big\{\boldsymbol{t}(n)\big\} = \boldsymbol{S}(n)\mathbb{E}\big\{\widetilde{\boldsymbol{w}}(n)\big\} \tag{36}$$

with

$$\boldsymbol{S}(n) = \sqrt{2/\pi}\,\mathrm{bdiag}\big\{\theta_1\boldsymbol{I}_L,\ldots,\theta_K\boldsymbol{I}_L\big\}\boldsymbol{R}_x(n). \tag{37}$$

For the sake of mathematical tractability, we assume that the following two approximations hold, namely

$$\mathbb{E}\big\{\boldsymbol{\Phi}(n)\boldsymbol{\mathcal{A}}^{-1}\widetilde{\boldsymbol{w}}(n)\big\} \approx \mathbb{E}\big\{\boldsymbol{\Phi}(n)\big\}\boldsymbol{\mathcal{A}}^{-1}\mathbb{E}\big\{\widetilde{\boldsymbol{w}}(n)\big\} \tag{38}$$

$$\mathbb{E}\big\{\boldsymbol{\Phi}(n+1)\widetilde{\boldsymbol{w}}(n)\big\} \approx \mathbb{E}\big\{\boldsymbol{\Phi}(n+1)\big\}\mathbb{E}\big\{\widetilde{\boldsymbol{w}}(n)\big\}. \tag{39}$$

The proofs of two approximations (38) and (39) are not presented explicitly due to the constrained space, whereas their rationality and effectiveness will be validated by simulation results later. Substituting (36) into (29), and using the above presented approximations (38) and (39), yields

$$\mathbb{E}\{\boldsymbol{\Phi}(n+1)\}\boldsymbol{\mathcal{A}}^{-1}\mathbb{E}\{\widetilde{\boldsymbol{w}}(n+1)\}=\left[\mathbb{E}\{\boldsymbol{\Phi}(n+1)\}-\boldsymbol{S}(n)\right]\mathbb{E}\{\widetilde{\boldsymbol{w}}(n)\}. \tag{40}$$

Pre-multiplying both sides of (40) by $\boldsymbol{\mathcal{A}}\left[\mathbb{E}\{\boldsymbol{\Phi}(n+1)\}\right]^{-1}$, the mean weight error behavior of DRLS-SE algorithm is given by

$$\mathbb{E}\{\widetilde{\boldsymbol{w}}(n+1)\}=\boldsymbol{\mathcal{A}}\left[\boldsymbol{I}_{KL}-\left[\mathbb{E}\{\boldsymbol{\Phi}(n+1)\}\right]^{-1}\boldsymbol{S}(n)\right]\mathbb{E}\{\widetilde{\boldsymbol{w}}(n)\} \tag{41}$$

which utilizes the recursive relation (23) for final evaluation.

Theorem 1 (Convergence in the mean) Given a left-stochastic matrix $\boldsymbol{A}$, the weight error vector of DRLS-SE algorithm converges to a zero vector as $n \to \infty$, that is

$$\lim_{n\to\infty}\mathbb{E}\{\widetilde{\boldsymbol{w}}(n)\}=\boldsymbol{0}_{KL} \tag{42}$$

which ensures that the estimate of DRLS-SE is asymptotically unbiased and convergent in the mean sense, i.e., $\lim_{n\to\infty}\mathbb{E}\{\boldsymbol{w}_k(n)\}=\boldsymbol{w}^\star(n)$ for all nodes k.

Proof: When $n \to \infty$, based on (41) the steady-state mean weight error behavior of DRLS-SE algorithm can be expressed as

$$\lim_{n\to\infty}\mathbb{E}\{\widetilde{\boldsymbol{w}}(n+1)\}=\boldsymbol{\mathcal{A}}\left[\boldsymbol{I}_{KL}-\lim_{n\to\infty}\left[\mathbb{E}\{\boldsymbol{\Phi}(n+1)\}\right]^{-1}\boldsymbol{S}(n)\right] \times \lim_{n\to\infty}\mathbb{E}\{\widetilde{\boldsymbol{w}}(n)\}. \tag{43}$$

Inserting (22) in (43) and using (37), we can immediately obtain

$$\lim_{n\to\infty}\mathbb{E}\{\widetilde{\boldsymbol{w}}(n+1)\}=\boldsymbol{\mathcal{A}}\big(\boldsymbol{I}_{KL}-\boldsymbol{H}\big)\lim_{n\to\infty}\mathbb{E}\{\widetilde{\boldsymbol{w}}(n)\} \tag{44}$$

with

$$\boldsymbol{H}=(1-\lambda)\sqrt{2/\pi}\,\mathrm{bdiag}\{\theta_1\boldsymbol{I}_L,\ldots,\theta_K\boldsymbol{I}_L\}\in\mathbb{R}^{KL\times KL}. \tag{45}$$

Note that the steady-state mean stability of DRLS-SE algorithm does not depend on the variation of $\boldsymbol{R}_x(n)$. Consequently, the k-th entry of (44) can be given as follow:

$$\lim_{n\to\infty}\mathbb{E}\{\widetilde{\boldsymbol{w}}_k(n+1)\}=\sum_{\ell\in\mathcal{N}_k}a_{\ell k}\Big[1-(1-\lambda)\sqrt{2/\pi}\theta_k\Big]\boldsymbol{I}_L \times\lim_{n\to\infty}\mathbb{E}\{\widetilde{\boldsymbol{w}}_k(n)\}. \tag{46}$$

Taking $0 \ll \lambda < 1$ into account, it is easy to get

$$0<(1-\lambda)\sqrt{2/\pi}\ll 1. \tag{47}$$

In view of (47), (46) implies that the steady-state mean stability (44) is determined by all the values of θ_k for $k=1,2,\ldots,K$. Thus, we shall discuss three possible values of θ_k in (35) as below:

Case 1: When $\sigma_{v,k}^2 \gg \max\{\mathrm{tr}\{\boldsymbol{K}_k(n)\boldsymbol{R}_{x,k}(n)\}\} > 1$, we may approximate θ_k only with the noise variance $\sigma_{v,k}$ as

$$\theta_k \approx (1-p_r)/\sigma_{v,k}+p_r/(\sqrt{\zeta+1}\sigma_{v,k})\approx\sigma_{v,k}^{-1} \tag{48}$$

where the probability p_r is small, and ζ is very large value. Accordingly, we have $0<\theta_k\ll 1$ due to $\sigma_{v,k}\gg 1$. It then follows from (47) and (48) that

$$0 \ll 1-\sqrt{2/\pi}(1-\lambda)\theta_k < 1 \tag{49}$$

for all nodes k, which implies that (42) holds true.

Case 2: When $\sigma_{v,k}^2 \approx \min\{\mathrm{tr}\{\boldsymbol{K}_k(n)\boldsymbol{R}_{x,k}(n)\}\}$, because of the small probability p_r and very large ζ, we also have the following approximation

$$\theta_k \approx \frac{1-p_r}{\sqrt{2\min\{\mathrm{tr}\{\boldsymbol{K}_k(n)\boldsymbol{R}_{x,k}(n)\}\}}}+\frac{p_r}{\sqrt{(\zeta+2)\min\{\mathrm{tr}\{\boldsymbol{K}_k(n)\boldsymbol{R}_{x,k}(n)\}\}}} \approx 1/\sqrt{2\min\{\mathrm{tr}\{\boldsymbol{K}_k(n)\boldsymbol{R}_{x,k}(n)\}\}}. \tag{50}$$

Case 3: When $\min\{\mathrm{tr}\{\boldsymbol{K}_k(n)\boldsymbol{R}_{x,k}(n)\}\} > \sigma_{v,k}^2=0$, i.e., the ideal noise free case, it follows that

$$\theta_k = 1/\sqrt{\min\{\mathrm{tr}\{\boldsymbol{K}_k(n)\boldsymbol{R}_{x,k}(n)\}\}}. \tag{51}$$

The realizations of random variable given by (50) and (51) are much greater than 1, whereas, they are dramatically attenuated below 1 by (47) as the numerical simulations validated in Section 4. Finally, we can still arrive at (49). Therefore, we conclude that the solution of DRLS-SE algorithm is asymptotically unbiased in the mean. ■

3.2. Mean-Square Weight Error Analysis

We shall perform the mean-square transient analysis of DRLS-SE algorithm in terms of the network mean-square-deviation (MSD) [1], i.e., $\mathrm{MSD}(n)=\mathrm{tr}\{\boldsymbol{K}(n)\}/K=\sum_{k=1}^{K}\mathrm{MSD}_k(n)/K$, where $\boldsymbol{K}(n)=\mathbb{E}\{\widetilde{\boldsymbol{w}}(n)\widetilde{\boldsymbol{w}}^\top(n)\}$ is the second-order moment matrix of global weight error vector across all the nodes.

Post-multiplying (28) by its transpose, and taking the expectation of its both sides, then applying the statistical properties of perturbation vector $\boldsymbol{q}(n)$, we obtain

$$\begin{aligned}&\mathbb{E}\{\boldsymbol{\Phi}(n+1)\boldsymbol{\mathcal{A}}^{-1}\widetilde{\boldsymbol{w}}(n+1)\widetilde{\boldsymbol{w}}^\top(n+1)(\boldsymbol{\mathcal{A}}^{-1})^\top\boldsymbol{\Phi}(n+1)\}\\&=\mathbb{E}\{\boldsymbol{\Phi}(n+1)\widetilde{\boldsymbol{w}}(n)\widetilde{\boldsymbol{w}}^\top(n)\boldsymbol{\Phi}(n+1)\}+\mathbb{E}\{\boldsymbol{t}(n)\boldsymbol{t}^\top(n)\}\\&\quad-\mathbb{E}\{\boldsymbol{t}(n)\widetilde{\boldsymbol{w}}^\top(n)\boldsymbol{\Phi}(n+1)\}-\mathbb{E}\{\boldsymbol{\Phi}(n+1)\widetilde{\boldsymbol{w}}(n)\boldsymbol{t}^\top(n)\}\\&\quad+\mathbb{E}\{\boldsymbol{\Phi}(n+1)\}\mathbb{E}\{\boldsymbol{g}(n)\boldsymbol{g}^\top(n)\}\mathbb{E}\{\boldsymbol{\Phi}(n+1)\}.\end{aligned} \tag{52}$$

In order to make the mean-square analysis tractable, we introduce the following necessary approximations:

$$\begin{aligned}&\mathbb{E}\{\boldsymbol{\Phi}(n+1)\boldsymbol{\mathcal{A}}^{-1}\widetilde{\boldsymbol{w}}(n+1)\widetilde{\boldsymbol{w}}^\top(n+1)(\boldsymbol{\mathcal{A}}^{-1})^\top\boldsymbol{\Phi}(n+1)\}\\&\approx\mathbb{E}\{\boldsymbol{\Phi}(n+1)\}\boldsymbol{\mathcal{A}}^{-1}\boldsymbol{K}(n+1)(\boldsymbol{\mathcal{A}}^{-1})^\top\mathbb{E}\{\boldsymbol{\Phi}(n+1)\},\end{aligned} \tag{53}$$

$$\begin{aligned}&\mathbb{E}\{\boldsymbol{\Phi}(n+1)\widetilde{\boldsymbol{w}}(n)\widetilde{\boldsymbol{w}}^\top(n)\boldsymbol{\Phi}(n+1)\}\\&\approx\mathbb{E}\{\boldsymbol{\Phi}(n+1)\}\boldsymbol{K}(n)\mathbb{E}\{\boldsymbol{\Phi}(n+1)\},\end{aligned} \tag{54}$$

$$\mathbb{E}\{\boldsymbol{t}(n)\widetilde{\boldsymbol{w}}^\top(n)\boldsymbol{\Phi}(n+1)\}\approx\mathbb{E}\{\boldsymbol{t}(n)\widetilde{\boldsymbol{w}}^\top(n)\}\mathbb{E}\{\boldsymbol{\Phi}(n+1)\}. \tag{55}$$

Likewise, the corresponding proofs of the above presented approximations (53)–(55) are omitted due to the limited space. Substituting (53)–(55) into (52), yields

$$\begin{aligned}&\mathbb{E}\{\boldsymbol{\Phi}(n+1)\}\boldsymbol{\mathcal{A}}^{-1}\boldsymbol{K}(n+1)(\boldsymbol{\mathcal{A}}^{-1})^\top\mathbb{E}\{\boldsymbol{\Phi}(n+1)\}\\&=\mathbb{E}\{\boldsymbol{\Phi}(n+1)\}\boldsymbol{K}(n)\mathbb{E}\{\boldsymbol{\Phi}(n+1)\}-\boldsymbol{U}(n)\mathbb{E}\{\boldsymbol{\Phi}(n+1)\}+\boldsymbol{T}(n)\\&\quad-\mathbb{E}\{\boldsymbol{\Phi}(n+1)\}\boldsymbol{U}^\top(n)+\mathbb{E}\{\boldsymbol{\Phi}(n+1)\}\,\boldsymbol{Q}\,\mathbb{E}\{\boldsymbol{\Phi}(n+1)\}.\end{aligned} \tag{56}$$

with

$$\boldsymbol{T}(n)=\mathbb{E}\{\boldsymbol{t}(n)\boldsymbol{t}^\top(n)\}, \tag{57}$$

$$\boldsymbol{U}(n)=\mathbb{E}\{\boldsymbol{t}(n)\widetilde{\boldsymbol{w}}^\top(n)\}, \tag{58}$$

$$\boldsymbol{Q}=\mathbb{E}\{\boldsymbol{g}(n)\boldsymbol{g}^\top(n)\}. \tag{59}$$

By assumption A1, it is easy to obtain

$$\boldsymbol{T}(n)=\mathbb{E}\{\boldsymbol{t}(n)\boldsymbol{t}^\top(n)\}=\boldsymbol{R}_x(n). \tag{60}$$

Then, the (i,j)-th block of matrix $\boldsymbol{U}(n)$ is given by

$$\begin{aligned}\boldsymbol{U}_{ij}(n) &= \mathbb{E}\{\boldsymbol{t}_i(n)\widetilde{\boldsymbol{w}}_j^\top(n)\} \\ &= \mathbb{E}\{\mathrm{sgn}\{\boldsymbol{x}_i^\top(n)\widetilde{\boldsymbol{w}}_i(n) - z_i(n)\}\boldsymbol{x}_i(n)\widetilde{\boldsymbol{w}}_j^\top(n)\}.\end{aligned} \tag{61}$$

Furthermore, the submatrix $\boldsymbol{U}_{ij}(n)$ can be written as

$$\big[\boldsymbol{U}_{ij}(n)\big]_{\ell p} = \mathbb{E}\{\mathrm{sgn}\{\boldsymbol{x}_i^\top(n)\widetilde{\boldsymbol{w}}_i(n) - z_i(n)\}x_i(n-\ell+1)\big[\widetilde{\boldsymbol{w}}_j(n)\big]_p\}. \tag{62}$$

Subsequently, we consider assumptions A1 and A2, and Price's theorem [28] for the following derivation. Taking the conditional expectation of (62) on $\widetilde{\boldsymbol{w}}_i(n)$, applying the lemma 1, then using the approximation (33), we obtain

$$\begin{aligned}&\big[\boldsymbol{U}_{ij}(n)\big]_{\ell p} \\ &= \mathbb{E}\{\mathrm{sgn}\{\boldsymbol{x}_i^\top(n)\widetilde{\boldsymbol{w}}_i(n) - z_i(n)\}x_i(n-\ell+1)\big[\widetilde{\boldsymbol{w}}_j(n)\big]_p|\widetilde{\boldsymbol{w}}_i(n)\} \\ &= \sqrt{2/\pi}\theta_i\big[\boldsymbol{R}_{x,i}(n)\big]_\ell^\top\mathbb{E}\{\widetilde{\boldsymbol{w}}_i(n)\big[\widetilde{\boldsymbol{w}}_j(n)\big]_p\}\end{aligned} \tag{63}$$

From (63), thus $\boldsymbol{U}_{ij}(n)$ defined in (61) can be calculated as

$$\boldsymbol{U}_{ij}(n) = \sqrt{2/\pi}\theta_i\boldsymbol{R}_{x,i}(n)\mathbb{E}\{\widetilde{\boldsymbol{w}}_i(n)\widetilde{\boldsymbol{w}}_j(n)\}. \tag{64}$$

In view of (64), (58) can be expressed in compact matrix form as

$$\boldsymbol{U}(n) = \boldsymbol{S}(n)\boldsymbol{K}(n). \tag{65}$$

According to the temporal independence of perturbation vector $\boldsymbol{q}(n)$, the matrix $\boldsymbol{Q}$ is computed as

$$\boldsymbol{Q} = \mathbb{E}\{\boldsymbol{g}(n)\boldsymbol{g}^\top(n)\} = \sigma_q^2\mathbf{1}_K \otimes \boldsymbol{I}_L. \tag{66}$$

Substituting Eqs. (60), (65), and (66) into (56), we then arrive at

$$\begin{aligned}&\mathbb{E}\{\boldsymbol{\Phi}(n+1)\}\boldsymbol{\mathcal{A}}^{-1}\boldsymbol{K}(n+1)(\boldsymbol{\mathcal{A}}^{-1})^\top\mathbb{E}\{\boldsymbol{\Phi}(n+1)\} \\ &= \mathbb{E}\{\boldsymbol{\Phi}(n+1)\}\boldsymbol{K}(n)\mathbb{E}\{\boldsymbol{\Phi}(n+1)\} + \boldsymbol{R}_x(n) - \boldsymbol{S}(n)\boldsymbol{K}(n)\mathbb{E}\{\boldsymbol{\Phi}(n+1)\} \\ &- \mathbb{E}\{\boldsymbol{\Phi}(n+1)\}\boldsymbol{K}^\top(n)\boldsymbol{S}^\top(n) + \mathbb{E}\{\boldsymbol{\Phi}(n+1)\}\,\boldsymbol{Q}\,\mathbb{E}\{\boldsymbol{\Phi}(n+1)\}.\end{aligned} \tag{67}$$

Pre-multiplying (67) by $\boldsymbol{\mathcal{A}}[\mathbb{E}\{\boldsymbol{\Phi}(n+1)\}]^{-1}$ and post-multiplying (67) by $[\mathbb{E}\{\boldsymbol{\Phi}(n+1)\}]^{-1}\boldsymbol{\mathcal{A}}^\top$ simultaneously, we finally obtain the recursive update equation of $\boldsymbol{K}(n)$ as

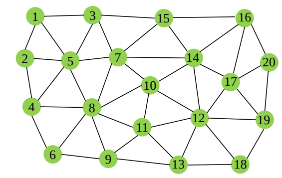
$$\begin{aligned}&\boldsymbol{K}(n+1) \\ &= \boldsymbol{\mathcal{A}}\Big[\boldsymbol{K}(n) + \boldsymbol{Q} + [\mathbb{E}\{\boldsymbol{\Phi}(n+1)\}]^{-1}\boldsymbol{R}_x(n)[\mathbb{E}\{\boldsymbol{\Phi}(n+1)\}]^{-1} \\ &- [\mathbb{E}\{\boldsymbol{\Phi}(n+1)\}]^{-1}\boldsymbol{S}(n)\boldsymbol{K}(n) - \boldsymbol{K}^\top(n)\boldsymbol{S}^\top(n)[\mathbb{E}\{\boldsymbol{\Phi}(n+1)\}]^{-1}\Big]\boldsymbol{\mathcal{A}}^\top\end{aligned} \tag{68}$$

with the initialization $\boldsymbol{K}(0) = \big[\mathbf{1}_K \otimes \boldsymbol{w}^\star(0)\big]\big[\mathbf{1}_K \otimes \boldsymbol{w}^\star(0)\big]^\top$. Note that the recursive relation (23) is also used for the evaluation of $\boldsymbol{K}(n)$. Eq. (68) allows us to investigate the network transient convergence behavior of DRLS-SE algorithm in the mean-square sense.

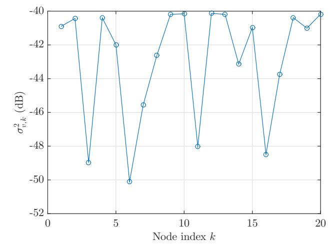
4. SIMULATION RESULTS

In this section, we present numerical simulations to testify the robustness of DRLS-SE algorithm against impulsive noise and the accuracy of derived theoretical analysis results. Each empirical curve was averaged over 200 independent trails to obtain the smoothness.

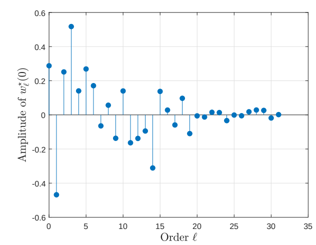
We considered a network consisting of 20 nodes with the topology illustrated in Fig. 1(a). Fig. 1(b) depicts the noise variances $\sigma_{v,k}^2$ at each node. Fig. 1(c) depicts the initialized system impulse response$\boldsymbol{w}^\star(0)$ generated from a standard normal distribution scaling by a exponential decaying factor 0.5 for all nodes.



(a) The network topology.

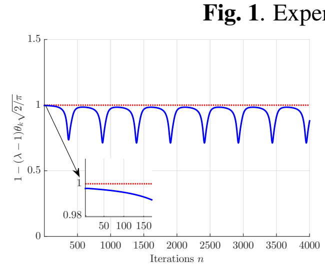


(b) $\sigma_{v,k}^2$ for all node k.

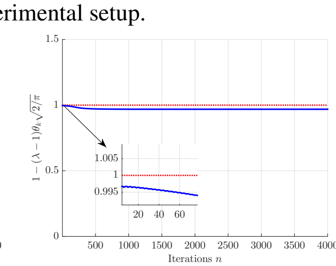


(c) $\boldsymbol{w}^\star(0)$ for all node k.

Fig. 1. Experimental setup.



(a) Slow variation ($T = 512$).



(b) Fast variation ($T = 4$).

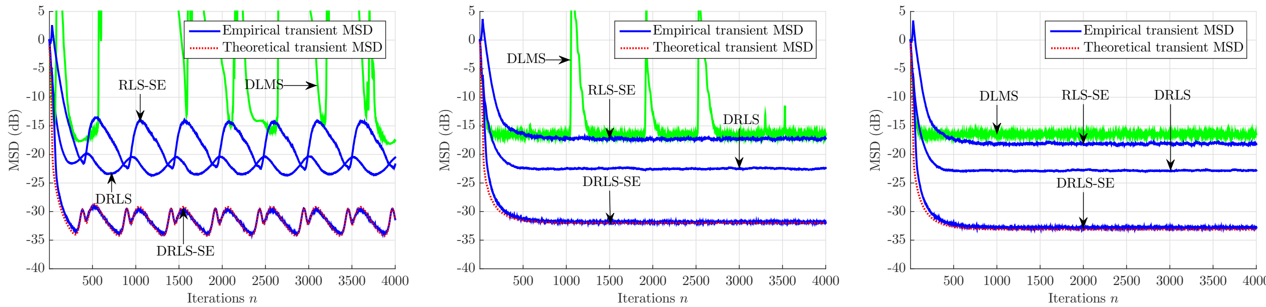
Fig. 2. The realizations of random variable for $\rho = 0.5$.

The parameters of sinusoidal variation model were set to be same as those in references [20, 25]. Three different periods of sinusoidal cyclostationary inputs variances were set to $T = 512$, $T = 32$, and $T = 4$, respectively. The variance σ_q^2 of perturbation vector was set to 1×10^{-6}. We set the step-size η of DLMS algorithm to be 0.05 for all nodes. The parameter ζ and probability p_r of impulsive CG noise were selected as 5×10^4 and 0.1, respectively. We chose the forgetting factor λ and parameter δ as 0.995 and 0.1 for RLS-SE, DRLS, and DRLS-SE algorithms. The stationary colored sequences $u_k(n)$ are generated independently across all nodes k according to the first-order autoregression, i.e., $u_k(n) = \rho u_k(n-1) + \sigma_{u,k}\sqrt{1-\rho^2}w_k(n)$, where ρ is the normalized correlation factor, and $w_k(n)$ is generated from the standard normal distribution. Hence, the (i,j)-th entry of autocorrelation matrix of vector $\boldsymbol{u}_k(n)$ can be determined as

$$[\boldsymbol{R}_{u,k}]_{ij} = \mathbb{E}\{u_k(n-i+1)u_k(n-j+1)\} = \sigma_{u,k}^2\rho^{|i-j|} \tag{69}$$

with $1 \leq i, j \leq L$. The realizations of random variable $[1 - (1-\lambda)\sqrt{2/\pi}\theta_k]$ given by (46) are depicted in Fig. 2 for slow and fast sinusoidal variation of cyclostationary input variance. It can be clearly seen that the realized values are very close to 1 during the initial stage, however, the entire dynamic realizations versus time instant are bounded and always less than 1. Therefore, the simulation results actually indicate that the weight error vector of DRLS-SE algorithm asymptotically approaches to zero in the mean sense.

Fig. 3 shows the empirical and theoretical MSD with $\rho = 0.5$ for slow, moderate, and fast sinusoidal variations of the input variance, respectively. First, DRLS-SE algorithm can perform well for the time-variant system and cyclostationary colored inputs in the impulsive noise environment, which results in severe performance degradation and even divergence for DLMS-type algorithms. Second, the DRLS-SE significantly outperforms the noncooperative RLS-SE



(a) Sinusoidal slow variation ($T = 512$). (b) Sinusoidal moderate variation ($T = 32$). (c) Sinusoidal fast variation ($T = 4$).

Fig. 3. The comparisons of empirical and theoretical MSD for $\rho = 0.5$ and sinusoidal variations of input variance.

and the DRLS without robustness in terms of convergence rate and steady-state accuracy of MSD. Third, the empirical and theoretically predicted MSD of DRLS-SE algorithm coincide with each other except of slight mismatch in the initial transient phase, which is caused by the large deviation between the instantaneous $\mathbf{\Phi}(n)$ and expected $\mathbb{E}\{\mathbf{\Phi}(n)\}$ given by (23). Therefore, one can see the good agreement between the empirical and theoretical learning curves, which demonstrates the validity of theoretical findings and supported the rationality of necessary assumptions and used approximations.

5. CONCLUSION

In this paper, we proposed the robust DRLS-SE algorithm against impulsive noise. Its transient theoretical convergence performance was analyzed to examine the change of convergence characteristics under the conditions of time-variant system and cyclostationary colored inputs. Simulation results illustrate the superiority of the proposed algorithm, as well as the sufficient precision of theoretical predictions provided by the obtained analytical models.

6. REFERENCES

[1] A. H. Sayed, “Diffusion adaptation over networks,” in *Academic Press Library in Signal Processing*, vol. 3, pp. 323 – 453. Elsevier, 2014.

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

[3] J. Chen, C. Richard, and A. H. Sayed, “Multitask diffusion adaptation over networks with common latent representations,” *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 3, pp. 563–579, Apr. 2017.

[4] S. Shaghaghian and M. Coates, “Online Bayesian inference of diffusion networks,” *IEEE Trans. Signal Inf. Process. Netw.*, vol. 3, no. 3, pp. 500–512, Sept. 2017.

[5] J. Chen, C. Richard, and A. H. Sayed, “Diffusion LMS over multitask networks,” *IEEE Trans. on Signal Process.*, vol. 63, no. 11, pp. 2733–2748, Jun. 2015.

[6] J. Chen, J. Li, S. Yang, and F. Deng, “Weighted optimization-based distributed Kalman filter for nonlinear target tracking in collaborative sensor networks,” *IEEE Trans. Cybern.*, vol. 47, no. 11, pp. 3892–3905, Nov. 2017.

[7] Y. Dong, J. Chen, and W. Zhang, “Distributed wave-domain active noise control based on the diffusion adaptation,” *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 28, pp. 2374–2385, 2020.

[8] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, New Jersey, 2nd edition, 1991.

[9] A. H. Sayed, *Fundamentals of Adaptive Filtering*, Wiley, New York, 2003.

[10] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed, “Diffusion recursive least-squares for distributed estimation over adaptive networks,” *IEEE Trans. on Signal Process.*, vol. 56, no. 5, pp. 1865–1877, May 2008.

[11] A. Bertrand, M. Moonen, and A. H. Sayed, “Diffusion bias-compensated RLS estimation over adaptive networks,” *IEEE Trans. on Signal Process.*, vol. 59, no. 11, pp. 5212–5224, Nov. 2011.

[12] Z. Liu, Y. Liu, and C. Li, “Distributed sparse recursive least-squares over networks,” *IEEE Trans. on Signal Process.*, vol. 62, no. 6, pp. 1386–1395, Mar. 2014.

[13] X. Cao and K. J. R. Liu, “Decentralized sparse multitask RLS over networks,” *IEEE Trans. on Signal Process.*, vol. 65, no. 23, pp. 6217–6232, Dec. 2017.

[14] R. Arablouei, K. Dogancay, S. Werner, and Y. Huang, “Adaptive distributed estimation based on recursive least-squares and partial diffusion,” *IEEE Trans. on Signal Process.*, vol. 62, no. 14, pp. 3510–3522, Jul. 2014.

[15] A. Rastegarnia, “Reduced-communication diffusion RLS for distributed estimation over multi-agent networks,” *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 67, no. 1, pp. 177–181, Jan. 2020.

[16] W. A. Gardner, “Exploitation of spectral redundancy in cyclostationary signals,” *IEEE Signal Process. Mag.*, vol. 8, no. 2, pp. 14–36, Apr. 1991.

[17] W. A. Gardner, A. Napolitano, and L. Paura, “Cyclostationarity: Half a century of research,” *Signal Processing*, vol. 86, no. 4, pp. 639–697, Apr. 2006.

[18] N. J. Bershad, E. Eweda, and J. C. M. Bermudez, “Stochastic analysis of the LMS and NLMS algorithms for cyclostationary white Gaussian inputs,” *IEEE Trans. on Signal Process.*, vol. 62, no. 9, pp. 2238–2249, May 2014.

[19] E. Eweda and N. J. Bershad, “Stochastic analysis of the signed LMS algorithms for cyclostationary white Gaussian inputs,” *IEEE Trans. on Signal Process.*, vol. 65, no. 7, pp. 1673–1684, Apr. 2017.

[20] J. C. M. Bermudez, N. J. Bershad, and E. Eweda, “Stochastic analysis of the LMS algorithm for cyclostationary colored Gaussian inputs,” *Signal Processing*, vol. 160, pp. 127–136, 2019.

[21] W. Wang and H. Zhao, “Performance analysis of diffusion LMS algorithm for cyclostationary inputs,” *Signal Processing*, vol. 150, pp. 33–50, 2018.

[22] W. Gao, Y. Xu, and J. Chen, “Stochastic analysis of diffusion dual sign algorithm for cyclostationary white Gaussian inputs and nonstationary system,” *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 67, no. 7, pp. 1354–1358, Jul. 2020.

[23] Neil J. Bershad, Eweda Eweda, and Jose C.M. Bermudez, “Stochastic analysis of the diffusion LMS algorithm for cyclostationary white Gaussian inputs,” *Signal Processing*, vol. 185, pp. 108081, 2021.

[24] Y. Yu, H. Zhao, R. C. de Lamare, Y. Zakharov, and L. Lu, “Robust distributed diffusion recursive least squares algorithms with side information for adaptive networks,” *IEEE Trans. on Signal Process.*, vol. 67, no. 6, pp. 1566–1581, Mar. 2019.

[25] E. Eweda, N. J. Bershad, and J. C. M. Bermudez, “Stochastic analysis of the recursive least squares algorithm for cyclostationary colored inputs,” *IEEE Trans. on Signal Process.*, vol. 68, pp. 676–686, 2020.

[26] W. Gao, J. Chen, and C. Richard, “Transient theoretical analysis of diffusion RLS algorithm for cyclostationary colored inputs,” *IEEE Signal Process. Lett.*, vol. 28, pp. 1160–1164, 2021.

[27] S. C. Bang, S. Ann, and I. Song, “Performance analysis of the dual sign algorithm for additive contaminated-Gaussian noise,” *IEEE Signal Process. Lett.*, vol. 1, no. 12, pp. 196–198, Dec. 1994.

[28] R. Price, “A useful theorem for nonlinear devices having Gaussian inputs,” *IEEE Trans. Inf. Theory*, vol. 4, no. 2, pp. 69–72, Jun. 1958.