Study of Robust Distributed Diffusion RLS Algorithms with Side Information for Adaptive Networks

Yi Yu, Haiquan Zhao, Senior Member, IEEE, Rodrigo C. de Lamare, Senior Member, IEEE, Yury Zakharov, Senior Member, IEEE, and Lu Lu

Abstract—This work develops robust diffusion recursive least squares algorithms to mitigate the performance degradation often experienced in networks of agents in the presence of impulsive noise. The first algorithm minimizes an exponentially weighted least-squares cost function subject to a time-dependent constraint on the squared norm of the intermediate update at each node. A recursive strategy for computing the constraint is proposed using side information from the neighboring nodes to further improve the robustness. We also analyze the mean-square convergence behavior of the proposed algorithm. The second proposed algorithm is a modification of the first one based on the dichotomous coordinate descent iterations. It has a performance similar to that of the former, however its complexity is significantly lower especially when input regressors of agents have a shift structure and it is well suited to practical implementation. Simulations show the superiority of the proposed algorithms over previously reported techniques in various impulsive noise scenarios.

Index Terms—Distributed algorithms, diffusion cooperation, dichotomous coordinate-descent, impulsive noises, recursive least squares algorithms.

I. INTRODUCTION

OVER the past decade, distributed parameter estimation over wireless sensor networks with multiple nodes (agents) has attracted much attention. It only relies on the local data exchange between interconnected nodes, and therefore removes the requirement of a powerful central processor and, as such, reduces communications bandwidth of the traditional centralized estimation whilst retaining similar estimation performance [2], [3]. Distributed estimation has been applied to target localization [4], clustering [5], frequency estimation [6] and spectrum estimation in Cognitive radio (CR) [7], [8].

This work is partially supported by National Nature Science Foundation of P.R. China (Nos: 61871461, 61571374, 61433011, 11472297). The work of Y. Zakharov is partly supported by the UK Engineering and Physical Sciences Research Council (EPSRC) through Grant EP/R003297/1. The work of L. Lu is supported by China Postdoctoral Science Foundation Funded Project under Grant 2018M640916. An earlier version of this work was reported in the conference presentation IEEE International Conf. Acoustics, Speech and Signal Processing (ICASSP), Alberta, Canada, April 2018 [1].

Y. Yu is with the School of Information Engineering, Southwest University of Science and Technology, Mianyang, 621010, China (e-mail: yuyi_yxuan@163.com)
H. Zhao is with the School of Electrical Engineering, Southwest Jiaotong University, Chengdu, 610031, China. (e-mail: hqzhao_xjtu@126.com)
R. C. de Lamare is with the CETUC, PUC-Rio, Rio de Janeiro 22451-900, Brazil, and also with the Department of Electronic Engineering, University of York, York YO10 5DD, U.K. (e-mail: rcld500@ohm.york.ac.uk)
Y. Zakharov is with the Department of Electronic Engineering, University of York, York YO10 5DD, U.K. (e-mail: yury.zakharov@york.ac.uk)
L. Lu is with the School of Electronics and Information Engineering, Sichuan University, Chengdu, China. (lulu19900303@126.com).

A. Prior and Related Work

According to the cooperation strategies between interconnected nodes, existing algorithms can be categorized as incremental [9], consensus [10], and diffusion [11], [12], [13], [14] types. Among these, the diffusion strategy is popular, because it does not require a Hamiltonian cycle path as in the incremental type, thereby it is more robust to nodes/links failures; it is stable and shows a faster convergence rate and a lower mean-square error than that of the consensus approach. Several diffusion algorithms were proposed, e.g., diffusion least mean square (dLMS) algorithm [11] and its variable step size variants [15], [16].

In practice, the measurements can be corrupted by non-Gaussian noise with impulsiveness. Impulsive noise has small occurrence probability but much higher amplitude than the nominal measurements. It may occur due to atmospheric phenomena, or man-made due to either electric machinery in the operation environment [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43]. Other examples are keyboard clicking or pen dropping in teleconference [44], double-talk in echo cancellation [45], biological noise [46] or ice cracking [47] in various underwater signals, out-of-band spectral leakage in CR [48], etc. In such scenarios, the conventional algorithms like the dLMS designed for Gaussian noise would undergo a significant performance deterioration. To this end, many robust distributed algorithms have been proposed. Some algorithms are based on the instantaneous gradient-descent method to minimize different robust criteria, for instance, the diffusion error nonlinearity (dEN) [49], diffusion least mean p-th power (dLMP) [50], diffusion sign error LMS (dSE-LMS) [51], and diffusion maximum correntropy [52] algorithms. Moreover, because of the insensitivity of correntropy to impulsive noise, the maximum total correntropy diffusion algorithm was proposed in [53] for the case of large outliers in communication links. Nevertheless, their main limitation is slow convergence especially when the nodes’ input signals are colored (highly correlated). As shown in [49], the dEN algorithm converges slower than the dSE-LMS algorithm. In [54], by resorting to the adaptive projected subgradient method, a robust diffusion algorithm was developed which projects the output errors onto halfspaces defined by Huber’s error function at each node, thereby speeding up the convergence. However, the
setting of the parameters controlling the algorithm’s robustness requires prior knowledge of the noise distribution which is often unavailable.

It is well-known that due to the exponentially weighted least squares (EWLS) criterion, the diffusion recursive LS (dRLS) algorithm provides fast convergence even for colored signals [55], [56]. By means of the alternating direction method of multipliers to solve the EWLS problem, Mateos et al. proposed another type of distributed RLS algorithm [57]. Following this algorithm, to reduce computation and communication costs, its variants were presented by censoring observations with small innovations [58]. Likewise, these algorithms might experience convergence issues in impulsive noise environments, because impulsive noise samples are directly involved in the adaptation through output errors of nodes. For the single-agent case, many works have proposed RLS algorithms robust against impulsive noise, e.g., [59], [60]. However, distributed RLS-based techniques that are robust to impulsive noise have not been well investigated. The study in [61] develops the diffusion recursive least p-th power (dRLP) algorithm, while its robustness relies on the value of p as the dLMP does.

Analogous to the RLS, distributed RLS requires high computational complexity. Apart from this, it may also suffer from numerical instability due to accumulation of round-off errors in finite-precision implementations [62]. Aiming to address these problems, an efficient alternative method is the dichotomous coordinate-descent (DCD) that solves a system of normal equations associated with the RLS-type algorithms [63], [64], [65]. In particular, the DCD method only involves shift and addition operations, thus the DCD-based RLS algorithms reduce the computational cost and improve the numerical stability in contrast with the original RLS counterparts, whilst preserving comparable estimation performance. For this reason, reference [66] also explored the use of the DCD in distributed networks, and developed the DCD-dRLS algorithm.

It is worth mentioning that, however, the development of the DCD-based algorithms in impulsive noise environments has not been studied in single nor multi-agent scenarios.

B. Contributions

The focus of this paper is on developing robust distributed RLS algorithms for scenarios with impulsive noise. Specifically, our contributions are listed as follows:

1) A robust dRLS (R-dRLS) algorithm is developed by extending the framework of [59] to multi-agent scenarios with a diffusion distributed strategy. To ensure that the proposed R-dRLS algorithm has good convergence performance after an abrupt change in the set of parameters to be estimated, we also propose a diffusion-based non-stationary control (NC) method.

2) Theoretical insights into the mean square steady-state and evolution behaviors of the R-dRLS algorithm in impulsive noise environment are presented.

3) We employ the DCD method for developing recursions used in the adaptation step of the R-dRLS algorithm, resulting in the DCD-R-dRLS algorithm with similar learning performance. Remarkably, the DCD-R-dRLS algorithm brings a reduction in computational complexity; especially for shift structured input regressors, it reduces the order of complexity from $O(M^2)$ to $O(M)$, where $M$ is the length of the estimated vector.

4) Simulation examples are presented to demonstrate the performance of the proposed algorithms in impulsive noise scenarios described by either Bernoulli-Gaussian (BG) or α-stable processes.

In comparison to the preliminary results [1] related to this work, the current version is further developed due to the main contributions 2) and 3). We slightly improve the NC method by a smoothing operation as shown in [15]. Moreover, the effectiveness of the proposed algorithms are also verified in an application to distributed spectrum estimation.

This paper is organized as follows. In Section II, the problem formulation is described. The R-dRLS algorithm is derived in Section III. Analyses of its mean square behavior are presented in Section IV. In Section V, we review the DCD algorithm and propose the DCD-R-dRLS algorithm. In Section VI, extensive simulations are presented to verify the proposed algorithms. Finally, conclusions are given in Section VII.

Notation: Throughout the paper, all vectors are column vectors. We use the parenthesis on $i$ to denote matrices and vectors, and the subscript on $i$ to denote scalars. The superscript $(\cdot)^T$ denotes the transpose, $\|\cdot\|_2$ denotes the l2-norm of a vector, and $E\{\cdot\}$ denotes the expectation of random variables. We use col$(\cdots)$ to denote an enlarged column vector structured by stacking its columns on top of each other, diag$(\cdots)$ to yield a diagonal matrix with its arguments, and Tr$(\cdot)$ to denote the trace of a matrix. $I_M$ is the identity matrix of size $M \times M$, $\otimes$ is the Kronecker product, and $I$ is the column vector of length $M$ with all entries being one. For symmetric matrices $X$ and $Y$, the notation $X \succeq Y$ stands for $X - Y \geq 0$, meaning that the matrix difference $X - Y$ is positive semi-definite.

II. Problem Formulation

Let us consider a diffusion network with $N$ nodes located at different positions in space, as shown in Fig. 1 where each node communicates only with its neighboring nodes by a link (single-hop communication). All nodes connected directly to node $k$ (including itself) are referred to as its neighborhood, denoted as $N_k$. At every time instant $i \geq 0$, every node has access to an $M \times 1$ input regressor vector $u_{k,i}$ and an output measurement $d_{k}(i)$, which are related as:

$$d_{k}(i) = u_{k,i}^T w^o + v_{k}(i),$$

where $w^o$ is a parameter vector of size $M \times 1$ to be estimated, and $v_{k}(i)$ is the additive noise at node $k$. The additive noises $v_{k}(i)$ and $v_{j}(j)$ are spatially and temporally independent for $k \neq l$ and $i \neq j$. Moreover, any $u_{k,i}$ is independent of any $v_{j}(j)$. The model (1) is used in many applications [3], [67]. The objective of the in-network processing is to estimate $w^o$, using the available data $\{u_{k,i}, d_{k}(i)\}$ collected at nodes.
this purpose, the global EWLS estimation problem is described as:

\[
\mathbf{w}_i = \arg \min_{\mathbf{w}} \left\{ \lambda^{i+1} \| \mathbf{w} \|^2_2 + \sum_{j=0}^{i} \lambda^{i-j} \sum_{k=1}^{N} (d_k(j) - \mathbf{u}_{k,j}^T \mathbf{w})^2 \right\},
\]

where \( \delta > 0 \) is a regularization constant, and \( 0 < \lambda \leq 1 \) is the forgetting factor. The dRLS algorithm solves (2) in a diffusion-based distributed manner. As already mentioned in the introduction, the noise \( v_k(i) \) may be non-Gaussian with impulsiveness so that the algorithms derived from (2), e.g., the dRLS algorithm, would exhibit poor convergence and even diverge. In general, when studying robust adaptive algorithms, both contaminated-Gaussian (CG) and \( \alpha \)-stable random processes are often used for modeling impulsive noise.

III. PROPOSED R-dRLS ALGORITHM

In this section, we derive the R-dRLS algorithm and propose a control method for endowing it with tracking capability. The diffusion strategy has two alternatives: the adapt-then-combine (ATC) and the combine-then-adapt (CTA). However, we focus only on the ATC policy, which performs first the adaptation step and then the combination step. This is based on the fact that the extension to CTA is straightforward by reversing the order of the adaptation and combination steps. In what follows, we neglect the notation ATC for brevity.

A. dRLS Algorithm

To conveniently develop the R-dRLS algorithm, we re-derive here the dRLS algorithm from the following method instead of directly solving (2).

In the adaptation step, every node \( k \), at time instant \( i \), finds an intermediate estimate \( \psi_{k,i} \) of \( \mathbf{w}^o \) by minimizing the individual local cost function:

\[
J_k(\psi_{k,i}) = \| \psi_{k,i} - w_{k,i-1} \|^2_{\mathbf{B}_{k,i}} + [d_k(i) - \mathbf{u}_{k,i}^T \psi_{k,i}]^2,
\]

with \( \mathbf{B}_{k,i} = \Phi_{k,i} - \mathbf{u}_{k,i} \mathbf{u}_{k,i}^T \), where

\[
\Phi_{k,i} \triangleq \lambda^{i+1} \delta \mathbf{I}_M + \sum_{j=0}^{i} \lambda^{i-j} \mathbf{u}_{k,j} \mathbf{u}_{k,j}^T
\]

\[
= \lambda \Phi_{k,i-1} + \mathbf{u}_{k,i} \mathbf{u}_{k,i}^T
\]

is the time-averaged correlation matrix for the input vector at node \( k \), and \( \mathbf{w}_{k,i-1} \) is an estimate of \( \mathbf{w}^o \) at node \( k \) at time instant \( i-1 \). Notice that the quadratic form \( \| \mathbf{x} \|^2_{\mathbf{B}_{k,i}} \) defines the Riemannian distance between vectors \( \psi_{k,i} \) and \( \mathbf{w}_{k,i-1} \), where \( \mathbf{B}_{k,i} \) is a Riemannian metric tensor characterizing that the distance properties are not uniform along the \( M \)-dimensional space.

Setting the derivative of \( J_k(\psi_{k,i}) \) with respect to \( \psi_{k,i} \) to zero, we obtain

\[
\psi_{k,i} = w_{k,i-1} + \mathbf{P}_{k,i} \mathbf{u}_{k,i} c_k(i),
\]

where

\[
c_k(i) = d_k(i) - \mathbf{u}_{k,i}^T \mathbf{w}_{k,i-1}
\]

stands for the output error at node \( k \), and

\[
\mathbf{P}_{k,i} \triangleq \Phi_{k,i}^{-1} = \frac{1}{\lambda} \left( \mathbf{P}_{k,i-1} - \mathbf{P}_{k,i-1} \mathbf{u}_{k,i} \mathbf{u}_{k,i}^T \mathbf{P}_{k,i-1} \right),
\]

with \( \mathbf{P}_{k,i} \) initialized as \( \mathbf{P}_{k,0} = \delta^{-1} \mathbf{I}_M \). The recursion (7) is the result of applying the matrix inversion lemma.

At the combination step, the intermediate estimates \( \psi_{m,i} \), \( m \in \mathcal{N}_k \) from the neighborhood of node \( k \) are linearly weighted, yielding a combined estimate \( \mathbf{w}_{k,i} \):

\[
\mathbf{w}_{k,i} = \sum_{m \in \mathcal{N}_k} c_{m,k} \psi_{m,i},
\]

where the combination coefficients \( \{ c_{m,k} \} \) are non-negative, and satisfy:

\[
\sum_{m \in \mathcal{N}_k} c_{m,k} = 1, \quad \text{and} \quad c_{m,k} = 0 \quad \text{if} \quad m \notin \mathcal{N}_k.
\]

Note that \( c_{m,k} \) is a weight that node \( k \) assigns to the intermediate estimate \( \psi_{m,i} \) received from its neighbor node \( m \). If one assumes \( \mathbf{w}_{k,i-1} = \Phi_{k,i-1}^{-1} \mathbf{z}_{k,i-1} \) in (3), where \( \mathbf{z}_{k,i} = \lambda \mathbf{z}_{k,i-1} + \mathbf{u}_{k,i} d_k(i) \), (3) is a standard RLS update for node \( k \). In summary, (3)-(8) formulate the dRLS algorithm. It is noteworthy that the term \( \mathbf{P}_{k,i} \mathbf{u}_{k,i} \) in (5) provides the decorrelating ability for colored inputs, thus speeding up the convergence.

Remark 1: In general, \( \{ c_{m,k} \} \) in (9) are determined by one of many static rules (e.g., the Metropolis rule that we adopt in this paper) which keeps them constant during the estimation. Considering that nodes may be working under different signal-to-noise ratios (SNRs), several adaptive rules have been proposed to optimize the algorithm behavior. However, these adaptive rules are severely polluted when impulsive noise samples appear, since the output errors at nodes directly participate in the adaptation of \( c_{m,k} \). Designing a robust adaptive rule is an alternative, but it is not the focus of this paper. In another approach, based on the detection of impulsive noise, Ahn et al. proposed a robust variable weighting coefficients algorithm which sets the weighting coefficients to zero at nodes disturbed by impulsive noise. Likewise, the RVWC scheme can be extended to dRLS in a straightforward way, resulting in the
RVWC-dRLS algorithm with robustness in impulsive noise[1], as can be seen in the simulations later on.

B. R-dRLS Algorithm

An impulsive noise sample at time instant $i$ might lead the dRLS algorithm to diverge via $e_k(i)$ in (5) due to its large amplitude and the propagation of its effect. This degradation effect can last for many iterations. To endow the algorithm with robustness in impulsive noise scenarios, we propose to minimize (5) under the following constraint:

$$\|\psi_{k,i} - w_{k,i-1}\|^2_k \leq \xi_k(i-1),$$

where $\xi_k(i-1)$ is a positive bound. A similar constraint appeared in an adaptive filter for a single agent scenario [59], but when generalizing to the distributed version with multiple agents, the constraint could be imposed on the adaptation at all the nodes. This constraint represents that the energy (squared norm) of the update at every node $k$ from $w_{k,i-1}$ to $\psi_{k,i}$ always does not exceed the amount $\xi_k(i-1)$ regardless of the type of noise (possibly, impulsive noise), thereby guaranteeing the robustness of the algorithm. In doing so, if (5) satisfies (10), i.e.,

$$\|g_{k,i}\|_2|e_k(i)| \leq \sqrt{\xi_k(i-1)},$$

where $g_{k,i} \triangleq P_k,i u_{k,i}$ represents the Kalman gain vector, then (5) is a solution of the above constrained minimization problem. Conversely, if (10) is not satisfied (usually in the case of appearance of impulsive noise), i.e., $\|g_{k,i}\|_2|e_k(i)| > \sqrt{\xi_k(i-1)}$, we propose to replace the update (5) by its normalized form to satisfy the equality in (10), which is described by

$$\psi_{k,i} = w_{k,i-1} + \frac{\sqrt{\xi_k(i-1)}}{|g_{k,i}|_2} g_{k,i} \text{sign}(e_k(i)), \tag{12}$$

where $\text{sign}(\cdot)$ is the sign function. Thus, combining (5), (11) and (12), we obtain the adaptation step for each node $k$ as:

$$\psi_{k,i} = w_{k,i-1} + \min \left[ \frac{\sqrt{\xi_k(i-1)}}{|g_{k,i}|_2} |e_k(i)|, 1 \right] g_{k,i} e_k(i). \tag{13}$$

Evidently, the crucial problem is how to properly choose the bound $\xi_k(i)$ as it controls the robustness of the algorithm against impulsive noise and influences its dynamic behavior. To be more specific, we wish $\xi_k(i)$ to have larger values at the earlier adaptation stage to provide a fast initial convergence, while for enforcing good robustness against impulsive noise, its values cannot be too large. In addition, we also wish to obtain a small estimation error at steady-state, so $\xi_k(i)$ should be reduced to a small value. Based on these requirements, we consider the equality in (10) to propose a useful recursive method for adjusting $\xi_k(i)$, as described by

$$\xi_k(i) = \beta \xi_k(i-1) + (1 - \beta) \|\psi_{k,i} - w_{k,i-1}\|^2_k = \beta \xi_k(i-1) + (1 - \beta) \min [\|g_{k,i}\|_2^2 e_k^2(i), \xi_k(i-1)], \tag{14}$$

where $\beta$ is a memory factor with $0 < \beta < 1$. At every node $k$, $\xi_k(i)$ can be initialized by $\xi_k(0) = E_c \sigma^2_{w,k}(\xi_k(i-1), \xi_k(i-1))$, where $E_c$ is a positive integer, and $\sigma^2_{w,k}$ and $\sigma^2_{w,k}$ are powers of the output measurement $d_k(i)$ and the input regressor $u_{k,i}$, respectively. As one can see in (14), every node $k$ not only uses its own adaptive rule to update $\xi_k(i)$, but also exploits the side information $c_m(i)$ transmitted from its neighboring nodes by the diffusion cooperation. In doing so, the proposed R-dRLS algorithm is more effective at computing consistent estimates at all nodes, which will be observed in Section VI-A. Table III details the proposed R-dRLS algorithm together with the NC method.

**Remark 2:** As can be seen from (13), the operation mode of the proposed R-dRLS algorithm in the adaptation step can be as follows. At time instant $i$, if $\|g_{k,i}\|_2^2 e_k^2(i) \leq \xi_k(i-1)$, the classical RLS update is performed. If not, the squared norm of the RLS increment is limited to the amount $\xi_k(i-1)$.

| TABLE I |
| PROPOSED R-dRLS ALGORITHM ALLIED WITH THE NC METHOD |

| Parameters: $0 < \beta < 1, A, \delta$ and $E_c$ (R-dRLS part); $g$, $\tau$ and $t_0$ (NC part) |
| Initialization: $w_{k,0} = 0, P_{k,0} = \delta^{-1} I_M$ and $\xi_k(0) = E_c \sigma^2_{w,k}$ (R-dRLS part); $\Theta_{old,k} = \Theta_{new,k} = 0, \sigma^2_{u,k} = 0, V_i = \rho M$ and $V_g = 0.75V_i$ (NC part) |
| for iteration $i = 1, 2, 3, ...$ |
| for node $k = 1, 2, 3, \ldots, N$ |
| R-dRLS part: $e_k(i) = d_k(i) - w_{k,i}^T w_{k,i-1}$ $P_{k,i} = \frac{1}{\lambda} \left( P_{k,i-1} - \frac{P_{k,i-1} w_{k,i}^T w_{k,i-1}^T P_{k,i-1}}{\lambda + w_{k,i}^T P_{k,i-1} w_{k,i}} \right)$ $g_{k,i} = P_{k,i} u_{k,i}$ $\psi_{k,i} = w_{k,i-1} + \min \left[ \frac{\sqrt{\xi_k(i-1)}}{|g_{k,i}|_2} |e_k(i)|, 1 \right] g_{k,i} e_k(i)$ $w_{k,i} = \sum_{m \in \Lambda k} c_m \psi_{m,i}$ |
| NC part: $\Theta_{k,i} = \Theta_{new,k} - \Theta_{old,k}$ |
| Step 1: to compute $\Delta_k(i)$ if $i = nV_i$, $n = 0, 1, 2, ...$ $\alpha_{k,i} = \mathcal{R} \left( \left( \frac{1}{\sqrt{1 + 1}} \right)^{2i} \left( \frac{1}{\sqrt{1 + 1}} \right)^{2i} \ldots \left( \frac{1}{\sqrt{1 + 1}} \right)^{2i} \right)$ $\psi_{k,i} = w_{k,i} + \min \left[ \frac{\sqrt{\xi_k(i-1)}}{|g_{k,i}|_2} |e_k(i)|, 1 \right] g_{k,i} e_k(i)$ $\Delta_k(i) = \Theta_{new,k} - \Theta_{old,k} |
| Step 2: to reset $\xi_k(i)$ if $\Delta_k(i) > t_h$ $\xi_k(i) = \xi_k(0)$, $P_{k,i} = P_{k,0}$ else $\Theta_{new,k} > \Theta_{old,k}$ $\xi_k(i) = \xi_k(i-1) + (\Theta_{new,k} - \Theta_{old,k})$ $\Delta_k(i) = \Theta_{new,k} - \Theta_{old,k}$ $\xi_k(i) = \sum_{m \in \Lambda k} c_m \psi_{m,i}$ |
as in [12] for guaranteeing the robustness in impulsive noise. At the early iterations, the values of \( \xi_k(i-1) \) can be high compared to \( \|g_{k,i}\|^2 \xi_k(i) \) so that the algorithm will behave as the dRLS algorithm, providing a fast initial convergence. On the other hand, whenever an impulsive noise sample appears, due to its significant magnitude, the R-dRLS algorithm will work as an dRLS update multiplied by a very small scaling factor \( \frac{\sqrt{\xi_k(i-1)}}{\|g_{k,i}\|^2} \). It has been shown in [77], [78] that in the adaptation update term, the multiplication of a small scaling factor can reduce the negative influence of impulsive noise on the estimation. Thus, this also indirectly implies that the R-dRLS algorithm has robustness against impulsive noise. Moreover, the robustness is further maintained over the iterations, due to the decreasing property of \( \xi_k(i) \) given by (14). In addition to this, the diminishing \( \xi_k(i) \) also leads to a reduction in the steady-state error of the algorithm. To sum up, the R-dRLS algorithm can be considered as an improved dRLS algorithm with a variable ‘step-size’ scheme which has an automatic switch between 1 and \( \frac{\sqrt{\xi_k(i-1)}}{\|g_{k,i}\|^2} \), as can be observed in (13).

C. NC Method

As a consequence of the diminishing sequence \( \{\xi_k(i)\} \), the R-dRLS algorithm has poor ability of tracking (i.e., re-convergence of the algorithm) after \( w^o \) undergoes an abrupt change. In order to overcome this problem, inspired by the idea in [45] for the single-agent scenario, we propose here a diffusion-based NC method, as summarized in Table I The NC method is implemented in the following two steps.

Step 1: A variable \( \Delta_k(i) \) at node \( k \) is computed once for every \( V_i \) iterations, to judge whether the unknown vector changed or not. In this step, \( a^T_{k,i} = \mathcal{R} \left( \frac{\hat{e}^2(i)}{\|w_k\|}, \frac{\hat{e}^2(i-1)}{\|w_k\|}, \ldots, \frac{\hat{e}^2(i-V_i+1)}{\|w_k\|} \right) \) with \( \mathcal{R}(\cdot) \) denoting the ascending arrangement for its arguments. With \( e \in [1, \ldots, 1, 0, \ldots, 0]^T \) being a vector whose first \( V_1 \) – \( V_2 \) elements set to one, where \( V_2 \) is a positive integer with \( V_1 < V_2 \), the product \( a^T_{k,i}e \) can remove the effect of outliers (e.g., impulsive noise samples) when computing \( \Delta_k(i) \). We use a smooth filtering of \( a^T_{k,i}e \) to avoid large fluctuations in computing \( \Theta_{new,k} \) (see Table I), as follows:

\[
\sigma_{k,i}^2 \leftarrow \tau \sigma_{k,i}^2 + (1-\tau)a^T_{k,i}e,
\]

where \( \tau, 0 < \tau < 1 \), is a memory factor. Note that, every node \( k \) to compute \( \Theta_{new,k} \) also combines the information from its neighboring nodes based on a diffused cooperation; \( \Theta_{old,k} \) stores the value of \( \Theta_{new,k} \) at the last time instant.

From Step 1, one can see that using a larger \( V_i \), the algorithm has lower steady-state error but a higher delay in tracking. Moreover, for a large occurrence probability of impulsive noise, the value of \( V_1 \) should be increased to better discard the impulsive noise samples in the computation of \( \Delta_k(i) \). From our extensive simulations, we found out that for both \( V_1 \) and \( V_2 \), good choices are \( V_1 = \varphi M \) with \( 1 \leq \varphi \leq 3 \) and \( V_2 = 0.75V_1 \) [45].

Step 2: If \( \Delta_k(i) > t_h \), where \( t_h \) is a predefined threshold, it is decided that a change of \( w^o \) has occurred. Then, we reset \( \xi_k(i) \) to its initial value \( \xi_k(0) \) so that the R-dRLS algorithm can track this change rapidly. Meanwhile, \( P_{k,i} \) should also be re-initialized with \( P_{k,0} \).

It is worth noting that in this scheme the parameters \( \tau, \varphi \), and \( t_h \) are not affected by each other so that their choices are simplified.

IV. MEAN SQUARE PERFORMANCE ANALYSES

A. Steady-state Behavior

In this section, we discuss the steady-state behavior of the R-dRLS algorithm in impulsive noise. Assuming that the vector \( w^o \) is invariant, then we define the estimate deviation and intermediate estimate deviation vectors respectively as:

\[
\hat{w}_{k,i} \triangleq w^o - \hat{w}_{k,i},
\]

\[
\hat{\psi}_{k,i} \triangleq w^o - \psi_{k,i},
\]

Using these definitions and (14), it is easy to rearrange (13) and (8), respectively, as:

\[
\hat{w}_{k,i} - \beta E \{ \hat{w}_{k,i-1} \} \left( \frac{\zeta_k(i) - \beta \xi_k(i-1)}{\|g_{k,i}\|^2} \right) \|g_{k,i}\|^2 \|w^o\|^2 \approx \hat{w}_{k,i} - \beta E \{ \hat{w}_{k,i-1} \} \left( \frac{\zeta_k(i) - \beta \xi_k(i-1)}{\|g_{k,i}\|^2} \right) \|g_{k,i}\|^2 \|w^o\|^2.
\]

Equating the squared \( l_2 \)-norm of both sides of (17) and then taking the expectation, we obtain

\[
E \{ \|\hat{w}_{k,i} - \beta E \{ \hat{w}_{k,i-1} \} \|^2 \} \approx \beta E \{ \|w^o\|^2 \} \|g_{k,i}\|^2 \|w^o\|^2.
\]

Likewise treating (18) and applying Jensen’s inequality [79, p.77], we obtain

\[
E \{ \|\hat{\psi}_{k,i} - \beta E \{ \hat{\psi}_{k,i-1} \} \|^2 \} \leq \sum_{m \in \mathcal{N}_k} c_{m,k} E \{ \|\hat{\psi}_{m,i} \|^2 \}.
\]
Then, with this approximation, (19) is changed to
\[
E \left\{ \left\| \tilde{v}_{m,i} \right\|_2^2 \right\} = E \left\{ \left\| \tilde{w}_{k,i-1} \right\|_2^2 \right\} - 2 \sqrt{E \left\{ \zeta_k(i) \right\} - \beta E \left\{ \xi_k(i-1) \right\}} \times E \left\{ \frac{\tilde{w}_{k,i-1}^T g_{k,i}}{\|g_{k,i}\|_2} \text{sign}(e_k(i)) \right\} + \frac{E \left\{ \zeta_k(i) \right\} - \beta E \left\{ \xi_k(i-1) \right\}}{\beta - 1} \tag{22}
\]

To deal with the (a) term in (22), some assumptions are helpful.

**Assumption 1:** The input regressors \( u_{k,i} \) are zero-mean with covariance matrices \( R_k = E\{u_{k,i}u_{k,i}^T\} \) and spatially independent.

**Assumption 2:** The regressors \( \{u_{k,i}\} \) are independent of the estimates \( \{w_{m,i}\} \) for \( j \leq i \) and all \( k,m \), referred to as the independence assumption, which is known as useful in the analysis of adaptive algorithms [67] and distributed estimation algorithms [55, 80].

**Assumption 3:** There is an iteration number \( i_0 \) such that for all \( i > i_0 \), the time-averaged matrix \( \Phi_{k,i}^{-1} \) at every node \( k \) can be replaced by its expected value \( E\{\Phi_{k,i}\} \). This is an ergodicity assumption since \( 0 < \lambda < 1 \), and from (4) we have
\[
\lim_{i \to \infty} E \{ \Phi_{k,i} \} = R_k \frac{1}{1-\lambda} \triangleq \Phi_k. \tag{23}
\]

Correspondingly, we can also replace the random matrix \( \Phi_{k,i}^{-1} \) by \( \Phi_k^{-1} \triangleq E \{ \Phi_{k,i}^{-1} \} \) for a sufficiently large number of iterations \( i \). Note that such replacements are commonly used in the performance analysis of RLS-type algorithms, see [55, 67, 81, 56] and the references therein.

Applying assumption 3, we are able to represent the term (a) in (22) as:
\[
E \left\{ \frac{\tilde{w}_{k,i-1}^T g_{k,i}}{\|g_{k,i}\|_2} \text{sign}(e_k(i)) \right\} \approx E \left\{ \frac{\tilde{w}_{k,i-1}^T \Phi_k^{-1} u_{k,i}}{\sqrt{u_{k,i}^T \Phi_k^{-2} u_{k,i}}} \text{sign}(e_k(i)) \right\} \tag{24}
\]
\[
= E \left\{ \frac{\tilde{w}_{k,i-1}^T R_k^{-1} u_{k,i}}{\sqrt{u_{k,i}^T R_k^{-2} u_{k,i}}} \text{sign}(e_k(i)) \right\} \tag{25}
\]

In the light of assumption 1, if the dimension of \( w^o \) is large, i.e., \( M \gg 1 \), the fluctuation of the denominator term in (24) from one iteration to the next can be assumed to be small. So, we could make the following approximation (which is also verified in Appendix A):
\[
E \left\{ \frac{\tilde{w}_{k,i-1}^T R_k^{-1} u_{k,i}}{\sqrt{u_{k,i}^T R_k^{-2} u_{k,i}}} \text{sign}(e_k(i)) \right\} \approx \chi_k E \left\{ e_{a,k}^{-1}(i) \text{sign}(e_{a,k}(i) + v_k(i)) \right\}, \tag{26}
\]

where
\[
e_{a,k}(i) \triangleq \tilde{w}_{k,i-1}^T u_{k,i},
\]
\[
e_{a,k}^{-1}(i) \triangleq \tilde{w}_{k,i-1}^T R_k^{-1} u_{k,i},
\]
\[
\chi_k = E \left\{ \frac{1}{\sqrt{u_{k,i}^T R_k^{-2} u_{k,i}}} \right\}.
\]

Considering the presence of impulsive noise, we need the following assumptions to continue the analysis.

**Assumption 4:** At every node \( k \), the additive noise \( v_k(i) \) is drawn from a CG random process, \( v_k(i) = \theta_k(i) + \eta_k(i) \), where \( \theta_k(i) \) is the background noise assumed to be zero-mean white Gaussian with variance \( \sigma_{\theta,k}^2 \). The impulsive part \( \eta_k(i) \) is described as \( \eta_k(i) = b_k(i)g_k(i) \), where \( b_k(i) \) is drawn from a Bernoulli random process with the probability \( P[b_k(i) = 1] = p_r,k \), and \( g_k(i) \) is drawn from a white Gaussian random process with zero-mean and variance \( \sigma_{g,k}^2 = h^2 \sigma_{v,k}^2 \), \( h \gg 1 \). Usually, \( p_r,k \) is also called the appearance probability of an impulsive noise sample.

Then, the mean and variance of \( v_k(i) \) are zero and \( \sigma_{a,k}^2 = p_r,k(h + 1)\sigma_{\theta,k}^2 + (1 - p_r,k)\sigma_{g,k}^2 \) respectively. Note that, only when \( p_r,k = 0 \) or 1, \( v_k(i) \) is Gaussian; otherwise, \( v_k(i) \) is non-Gaussian. Also, \( v_k(i) \) conditioned on \( b_k(i) \) is Gaussian [68]. Although the \( \alpha \)-stable process is more appropriate for modeling impulsive noise in practice [44, 69, 70], one would not consider it in the algorithms’ analysis because its probability density function has no explicit form. Accordingly, the above assumption was used frequently for performance analysis of adaptive algorithms in impulsive noise environments, providing mathematical tractability [51, 45, 82, 68].

Furthermore, as pointed out in [83], when \( M \gg 1 \), then by using the central limit theorem, it can be assumed that \( e_{a,k}(i) \) and \( \chi_{a,k}^{-1}(i) \) are zero mean Gaussian variables for any constant matrix \( \Sigma \). Then, we can employ the following Lemma.

**Lemma:** Let \( e_a \) and \( u \) be jointly Gaussian zero-mean random variables. Let \( e = e_a + v \), where \( v \) is a zero-mean CG random variable with variance \( \sigma_v^2 = p_v(h + 1)\sigma_{\theta}^2 + (1 - p_v)\sigma_{g}^2 \), and \( v \) is independent of \( e_a \) and \( u \). If \( e_1 = e_a + \omega_1 \) and \( e_2 = e_a + \omega_2 \), where \( \omega_1 \) and \( \omega_2 \) are zero-mean Gaussian random variables with variances \( \sigma_{\omega_1}^2 = (h + 1)\sigma_{\theta}^2 \) and \( \sigma_{\omega_2}^2 = \sigma_{g}^2 \), and are independent of \( u \) and \( e_a \), then
\[
E\{\text{sign}(e)u\} = p_r E\{\text{sign}(e_1)u\} + (1 - p_r) E\{\text{sign}(e_2)u\}. \tag{27}
\]

Such a Lemma has been commonly used in the past for analyzing the sign-based algorithms [82, 51]. Based on Price’s theorem in [83], Lemma and assumption 2, we can establish the following equation:
\[
E \left\{ \frac{e_{a,k}^{-1}(i) \text{sign}(e_{a,k}(i) + v_k(i))}{\sqrt{\chi_{k,i}^T R_k^{-2} \tilde{w}_{k,i-1}}} \right\}, \tag{27}
\]
where
\[ \varpi_{k,i} = \sqrt{\frac{2}{\pi}} \left( \frac{p_{r,k}}{\sqrt{E[e_{a,k}^2(i)] + (h + 1)\sigma_0^2}} + \frac{1 - p_{r,k}}{\sqrt{E[e_{a,k}^2(i)] + \sigma_0^2}} \right) \neq 0, \]
and the notation \( E\{s|q\} \) accounts for the expectation of \( s \) conditioned on \( q \). Subsequently, the right-hand term in equality (25) becomes
\[
E \left\{ e_{a,k}^{-1}(i) \text{sign}(e_{a,k}(i) + v_k(i)) \right\} = E \left\{ E \left\{ e_{a,k}^{-1}(i) \text{sign}(e_{a,k}(i) + v_k(i)) \mid \bar{w}_{k,i-1} \right\} \right\} = \varpi_{k,i} E \left\{ \|\bar{w}_{k,i-1}\|_2^2 \right\},
\]
where the equality (a) is the result of using (26) under assumption 2.

Substituting (25) and (29) into (22), it is rearranged as
\[
E \left\{ \|\bar{\psi}_{k,i}\|_2^2 \right\} = E \left\{ \|\bar{w}_{k,i-1}\|_2^2 \right\} - 2\chi_k \varpi_{k,i} \sqrt{E \left\{ \zeta_k(i) - \beta E[\xi_k(i-1)] \right\} \left\{ \|\bar{w}_{k,i-1}\|_2^2 \right\}} + E \left\{ \zeta_k(i) - \beta E[\xi_k(i-1)] \right\}.
\]

Next, we introduce the following network global vectors:
\[
\mathcal{X}_i \triangleq \text{col} \left\{ \|\bar{\psi}_{i,1}\|_2^2, \ldots, \|\bar{\psi}_{i,N}\|_2^2 \right\},
\]
\[
\mathcal{W}_i \triangleq \text{col} \left\{ \|\bar{\psi}_{i,1}\|_2^2, \ldots, \|\bar{\psi}_{i,N}\|_2^2 \right\},
\]
and the network global matrices
\[
\Lambda_i \triangleq \text{diag} \left\{ \chi_{1,i}, \ldots, \chi_{N,i} \right\},
\]
\[
\Omega_i \triangleq \text{diag} \left\{ \sqrt{E \left\{ \zeta_1(i) - \beta E[\xi_1(i-1)] \right\} \left\{ \|\bar{\psi}_{i,1}\|_2^2 \right\}}, \ldots, \sqrt{E \left\{ \zeta_N(i) - \beta E[\xi_N(i-1)] \right\} \left\{ \|\bar{\psi}_{i,N}\|_2^2 \right\}} \right\}.
\]

Also, we define the matrix \( C \) to collect the combination coefficients, i.e., \( C_{m,k} = c_{m,k} \). Following (31) and (32), we can formulate (20) and (30) for all nodes as follows:
\[
\mathcal{W}_i \leq C^T \mathcal{X}_i = C^T \left[ \mathcal{W}_{i-1} - 2\Lambda_i \Omega_i \mathcal{W}_{i-1} + \Omega_i^2 1_M \right].
\]

Taking the \( \infty \)-norm for both sides of (33) leads to
\[
\|\mathcal{W}_i\|_\infty \leq \|C^T \left( \mathcal{W}_{i-1} - 2\Lambda_i \Omega_i \mathcal{W}_{i-1} + \Omega_i^2 1_M \right)\|_\infty \leq \|C^T\|_\infty \|\mathcal{W}_{i-1} - 2\Lambda_i \Omega_i \mathcal{W}_{i-1} + \Omega_i^2 1_M\|_\infty \leq \|\mathcal{W}_{i-1} - 2\Lambda_i \Omega_i \mathcal{W}_{i-1} + \Omega_i^2 1_M\|_\infty \]

where the equality (a) uses the fact that \( \|C^T\|_\infty = 1 \) in that the summation of each column of \( C \) is 1. Since \( \Lambda_i \) and \( \Omega_i \) are diagonal matrices with positive entries, (34) can be equivalently expressed as
\[
E \left\{ \|\bar{w}_{k,i}\|_2^2 \right\} \leq E \left\{ \|\bar{w}_{k,i-1}\|_2^2 \right\}
\]
\[
- 2\chi_k \varpi_{k,i} \sqrt{E \left\{ \zeta_k(i) - \beta E[\xi_k(i-1)] \right\} \left\{ \|\bar{w}_{k,i-1}\|_2^2 \right\}} + E \left\{ \zeta_k(i) - \beta E[\xi_k(i-1)] \right\} \frac{1}{1 - \beta}.
\]

for \( k = 1, \ldots, N \). When the algorithm reaches the steady-state, i.e., \( E \left\{ \|\bar{w}_{k,i}\|_2^2 \right\} = E \left\{ \|\bar{w}_{k,i-1}\|_2^2 \right\} \) as \( i \to \infty \), from (35) we will get:
\[
2\chi_k \varpi_{k,i} \lim_{i \to \infty} E \left\{ \|\bar{w}_{k,i-1}\|_2^2 \right\} \leq \sqrt{\lim_{i \to \infty} E \left\{ \zeta_k(i) - \beta E[\xi_k(i-1)] \right\} \left\{ \|\bar{w}_{k,i-1}\|_2^2 \right\}} \frac{1}{1 - \beta}.
\]

In view of the result that \( E \{ \zeta_k(i) \} \) and \( E \{ \xi_k(i) \} \) converge approximately to 0 as \( i \to \infty \) (see Appendix B) as well as \( \chi_k \neq 0 \) and \( \varpi_{k,i} \neq 0 \), thus, from (36) we are able to deduce that
\[
E \{ \|\bar{w}_k(\infty)\|_2^2 \} \approx 0, \quad \text{for} \quad k = 1, \ldots, N.
\]

As a result, (37) illustrates that based on given assumptions, the R-dRLS algorithm can converge to the true parameter vector in the mean-square sense after enough iterations even in impulsive noise environments.

**B. Analysis of Evolution Behavior**

The result (37) is qualitative so that it does not predict the steady-state performance of the algorithm, due mainly to the use of the upper bound relation (20). In this subsection, we will establish a recursive model to describe the evolution behavior of the algorithm in impulsive noise. We start by defining the following network vectors collected from all nodes:
\[
\bar{\psi}_i \triangleq \text{col} \left\{ \bar{\psi}_{i,1}, \ldots, \bar{\psi}_{i,N} \right\},
\]
\[
\bar{w}_i \triangleq \text{col} \left\{ \bar{w}_{i,1}, \ldots, \bar{w}_{i,N} \right\},
\]
\[
\Xi_i \triangleq \text{col} \left\{ \Xi_{1,i}, \ldots, \Xi_{N,i} \right\},
\]
where
\[
\Xi_{k,i} = \sqrt{E \{ \zeta_k(i) - \beta E[\xi_k(i-1)] \}} \frac{g_{k,i}}{\|g_{k,i}\|_2} \text{sign}(e_k(i))
\]
for nodes \( k = 1, \ldots, N \). By these defined vectors, we can associate (17) with (18) at all the nodes:
\[
\bar{w}_i = C^T \left[ \bar{w}_{i-1} - \Xi_i \right],
\]
where \( C = C \odot I_M \). Post-multiplying (40) by its transpose and taking the expectation, we have
\[
W_i = C^T \left[ W_{i-1} - E \{ \bar{w}_{i-1} \Xi_i^T \} \right] C.
\]
where $W_i \triangleq E\{\hat{w}_i \hat{w}_i^T\}$ denotes the covariance matrix of the deviation vector $\hat{w}_i$, and its $k$-th diagonal block of size $M \times M$, i.e., $W_{k,i} \triangleq E\{\hat{w}_{k,i} \hat{w}_{k,i}^T\}$, represents the covariance matrix of the deviation vector $\hat{w}_{k,i}$ at node $k$.

To evaluate terms I-III in (41), in addition to the spatially independence in assumption 1, we also require the input regressors $u_{k,i}$ to be statistically independent in time, which is also often used in analysis of distributed estimation algorithms [3, 2]. Therefore, performing similar manipulations as in Section IV-A on the expectations under assumptions 2-4, Lemma and Price’s theorem, we can compute these three terms. Specifically, the term I in (41) becomes

$$E\{\hat{w}_{i-1} \Xi^T_{i-1}\} = E\{E\{\hat{w}_{i-1} \Xi^T_{i-1} | \hat{w}_{i-1}\}\} = W_{i-1}[(\Lambda_i \Omega_i) \otimes I_M],$$

where we rewrite $\omega_{k,i}$ contained in $\Lambda_i$ as

$$\omega_{k,i} = \sqrt{\frac{2}{\pi}} \left\{ \frac{p_{r,k}}{\sqrt{\text{Tr}\{W_{k,i-1} R_i\} + (\bar{h} + 1)\sigma_0^2}} + \frac{1 - p_{r,k}}{\sqrt{\text{Tr}\{W_{k,i-1} R_i\} + \sigma_0^2}} \right\}$$

The term II in (41) is the transpose of (43). For any $k$ and $m$ belonging to the set $\{1, ..., N\}$, we define the $(m,k)$-th $M \times M$ matrix $E\{\Xi_i, \Xi_i^T\}$ as follows:

$$E\{\Xi_i, \Xi_i^T\}_{m,k} = E\{\Xi_{m,i} \Xi_{k,i}^T\}.$$  (44)

When $k = m$, (44) represents the k-th diagonal block of $E\{\Xi_i, \Xi_i^T\}$, which is described as

$$E\{\Xi_i, \Xi_i^T\}_{k,k} = \Omega_{k,i}^2 E\left( \frac{R_{k,i}^{-1} u_{k,i} u_{k,i}^T R_{k,i}^{-1}}{\sqrt{u_{k,i}^T R_{k,i}^{-2} u_{k,i}}} \right).$$  (45)

where $\Omega_{k,i}$ is the k-th element of $\Omega_i$. When $k \neq m$, the off-diagonal blocks will be simplified as

$$E\{\Xi_i, \Xi_i^T\}_{m,k} = E\{\Xi_{m,i} \Xi_{k,i}^T\}$$

$$= \chi_{m,k} \omega_{m,i} \omega_{k,i} \Xi_{m,i}^T \Xi_{k,i} + \Omega_{m,k} \Xi_{m,i}^T \Xi_{k,i},$$

From (45) and (46), we obtain the third term III in (41):

$$E\{\Xi_i, \Xi_i^T\} = [(\Lambda_i \Omega_i) \otimes I_M] [W_{i-1} - \tilde{W}_{i-1}] + \tilde{R},$$  (47)

where

$$\tilde{W}_{i-1} = \text{diag}[W_{1,i-1}, ..., W_{N,i-1}],$$

$$\tilde{R} = \text{diag} \left\{ \Omega_{1,i}^2 E\left( \frac{R_{1,i}^{-1} u_{1,i} u_{1,i}^T R_{1,i}^{-1}}{\sqrt{u_{1,i}^T R_{1,i}^{-2} u_{1,i}}} \right), ..., \right\},$$

$$\Omega_{N,i}^2 E\left( \frac{R_{N,i}^{-1} u_{N,i} u_{N,i}^T R_{N,i}^{-1}}{\sqrt{u_{N,i}^T R_{N,i}^{-2} u_{N,i}}} \right) \right\}. $$  (48)

By substituting (42) and (47) into (41), we obtain the recursive expression for $W_i$:

$$W_i = C_i \{W_{i-1} - \tilde{W}_{i-1}[(\Lambda_i \Omega_i) \otimes I_M] - [(\Lambda_i \Omega_i) \otimes I_M]^T W_{i-1} - ([\Lambda_i \Omega_i] \otimes I_M) \times \quad (49)[W_{k-1} - \tilde{W}_{k-1}] [(\Lambda_i \Omega_i) \otimes I_M] + \tilde{R} \} C_i.$$  

The mean square deviation (MSD) at node $k$ is defined as MSD$_k(i) \triangleq \text{Tr}\{W_k(i)\}$, and the network MSD over all the nodes is defined as MSDnet(i) = $\frac{1}{N} \sum_{k=1}^{N} \text{MSD}_k(i) = \frac{1}{N} \text{Tr}\{W_i\}$. Equation (49) models the MSD evolution behavior of the algorithm. It needs to be mentioned that to implement the model (49), $E\{\xi_k\}$ and $E\{\varsigma_k\}$ defined in $\Omega_i$ still need to be evaluated further. However, as shown in [12], $\xi_k$ and $\varsigma_k$ between interconnected nodes are affected by each other and there is a comparison operation, so it is difficult to provide an evolution expression for them. In this paper, we suggest that $E\{\xi_k\}$ and $E\{\varsigma_k\}$ are obtained by the ensemble average using simulations. Consequently, although (49) is a semi-analytic result, it can also be used to evaluate the convergence of the proposed algorithm.

V. DCD-BASED ALGORITHMS

In this section, we review the DCD-dRLS algorithm from [64], and then develop a robust DCD-dRLS algorithm.

A. The Original DCD-dRLS Algorithm

Since the dRLS algorithm involves the matrix operation of size $M \times M$ in the computations of $g_{k,i}$ and $\tilde{R}$ at every node, it requires a computational complexity that scales as a quadratic function of $M$ in terms of additions and multiplications per iteration $i$. To reduce the complexity, the adaptation step of the DCD-dRLS algorithm is described as [65]:

$$\psi_{k,i} = w_{k,i-1} + \Delta w_{k,i},$$  (50)

where the increment $\Delta w_{k,i}$ is obtained by solving the normal equation:

$$\Phi_{k,i} \Delta w_{k,i} = b_{k,i},$$  (51)

$$b_{k,i} = \lambda r_{k,i-1} + e_{k}(i) u_{k,i},$$  (52)

$r_{k,i-1}$ defines the residual vector at node $k$ at time instant $i-1$:

$$r_{k,i-1} = b_{k,i-1} - \Phi_{k,i-1} \Delta w_{k,i-1}.$$  (53)

For reducing the complexity of computing $\Delta w_{k,i}$ and $r_{k,i}$, the DCD method presented in Table II is used; see [63], [62], [64] for details. In Table II, $[r_{l,i}]$ is the $l$-th entry of a vector $r_{k,i}$, and $\{\Phi_{k,i}\}_{l,l}$ and $\{\Phi_{k,i}\}_{l,l}$ are the $(l,l)$-th entry and the $l$-th column of $\Phi_{k,i}$, respectively.

The accuracy and complexity of the DCD method are dependent on three parameters: $H$, $M_b$, and $N_u$. In general, $H$ is chosen as a power-of-two number; $M_b$ is the number of bits being enough for a fixed-point representation of $\Delta \hat{w}_{k,i}$ within an amplitude range $[-H, H]$; and $N_u$ defines a maximum number of elements in $\Delta \hat{w}_{k,i}$ that can be updated at a time instant. The DCD method only requires $2N_uM + M_b$ additions at most at each time instant with no multiplication [62]. Also, a larger $N_u$ makes the solution $\Delta \hat{w}_{k,i}$ closer to the optimal solution $\Delta w_{k,i}$ in (41), but increases the number of additions. It follows that if $N_u < M$, the DCD-based algorithm implements a selective partial update [65].

Similar to the dRLS algorithm, however, the DCD-dRLS algorithm will also encounter the performance deterioration when impulsive noise happens.
B. Proposed DCD-R-dRLS Algorithm

To achieve robustness against impulsive noise, we present here the DCD-R-dRLS algorithm.

**Step 1:** At every node $k$, we firstly use the DCD method to solve the normal equation (51) with (4) and (52), yielding a solution $\Delta \hat{w}_{k,i}^{(1)}$ and residual vector $r_{k,i}^{(1)}$. In the presence of impulsive noise, we also impose a constraint similar to that in (10):

$$
\|\Delta \hat{w}_{k,i}^{(1)}\|^2 \leq \xi_k(i-1). \tag{54}
$$

**Step 2:** If $\|\Delta \hat{w}_{k,i}^{(1)}\|^2 \leq \xi_k(i-1)$, we set $\Delta \hat{w}_{k,i} = \Delta \hat{w}_{k,i}^{(1)}$ and $r_{k,i} = r_{k,i}^{(1)}$ and then perform the update (50). Otherwise, we need to recalculate $b_{k,i}$ in (51) as:

$$
b_{k,i} = \lambda r_{k,i-1} + \frac{\sqrt{\xi_k(i-1)}}{\|\Delta \hat{w}_{k,i}^{(1)}\|_2} u_{k,i}. \tag{55}
$$

Subsequently, based on the DCD method, we obtain the solution $\Delta \hat{w}_{k,i}^{(2)}$ and the residual vector $r_{k,i}^{(2)}$ from the normal equation (51) under (4) and (55), thereby performing the update (50) with the increment

$$
\Delta \hat{w}_{k,i} = \frac{\sqrt{\xi_k(i-1)}}{\|\Delta \hat{w}_{k,i}^{(2)}\|_2} \Delta \hat{w}_{k,i}^{(2)}, \tag{56}
$$

and $r_{k,i} = r_{k,i}^{(2)}$.

**Step 3:** The combination step (12) is performed.

**Step 4:** The bound parameter $\xi_k(i)$ in the DCD-R-dRLS algorithm is updated according to

$$
\xi_k(i) = \beta \xi_k(i-1) + (1-\beta) \|\Delta \hat{w}_{k,i}\|_2^2, \tag{57}
$$

Table II summarizes the DCD-R-dRLS algorithm.

**Remark 3:** An impulsive noise sample appearing at time instant $i$ would yield a mismatch solution $\Delta \hat{w}_{k,i}^{(1)}$ so that $\|\Delta \hat{w}_{k,i}^{(1)}\|^2 > \xi_k(i-1)$. In this case, the scaling factor $\frac{\sqrt{\xi_k(i-1)}}{\|\Delta \hat{w}_{k,i}^{(1)}\|_2}$ in (55) is small enough to eliminate impulsive noise hidden in $e_k(i)$. A similar scaling factor $\frac{\sqrt{\xi_k(i-1)}}{\|\Delta \hat{w}_{k,i}^{(2)}\|_2}$ in (56) is to make the increment satisfy the constraint (54). Consequently, the DCD-R-dRLS algorithm improves the robustness to impulsive noise relative to the DCD-R-dRLS algorithm.

Moreover, the decreasing sequence $\{\xi_k(i)\}$ shown in (57) further guarantees the robustness. It is worth noting that due to $\|g_{k,i}\|_2 e_k(i) \approx \|\Delta \hat{w}_{k,i}\|_2$, the DCD-R-dRLS algorithm is a DCD-based variant of the R-dRLS algorithm. Unlike the R-dRLS algorithm, based on the NC method we re-initialize $\Phi_{k,0}$ with $\Phi_{k,0}$ to endow the DCD-R-dRLS algorithm with the tracking capability when $w^0$ suddenly changes.

**Remark 4:** Let $C_{d,cd}$ denote the only required number of additions for the DCD algorithm, with $C_{d,cd} \leq 2N_uM + M_b$.

In Table IV, we provide the computational complexity of the existing dLMS, dRLS, DCD-dRLS, and both proposed R-dRLS and DCD-R-dRLS algorithms at node $k$ per time instant $i$, where $n_k$ denotes the cardinality of $\mathcal{N}_k$. For shift structured input regressor at node $k$ (54), i.e., $u_{k,i} = [u_k(i), u_k(i-1), \ldots, u_k(i- M + 1)]^T$, where $u_k(i)$ is an input sample at time instant $i$, implementing $\Phi_{k,i}$ in (4) is very simplified. In this situation, by copying the upper-left $(M - 1) \times (M - 1)$ block of $\Phi_{k,i-1}$ leads to the lower-right $(M - 1) \times (M - 1)$ block of $\Phi_{k,i}$. The remaining part of $\Phi_{k,i}$ that needs to be updated is the first row and first column. Owing to symmetry of $\Phi_{k,i}$, only calculating the first column is sufficient, which is formulated as:

$$
[\Phi_{k,i}];1 = \lambda [\Phi_{k,i-1}];1 + u_k(i) u_{k,i}.
$$

Note that, in the DCD-R-dRLS algorithm, $\kappa = 1$ represents the case $\|\Delta \hat{w}_{k,i}\|^2 > \xi_k(i-1)$ at time instant $i$ (which leads to the maximum complexity), otherwise $\kappa = 0$. The comparisons required in the R-dRLS and DCD-R-dRLS algorithms are counted as additions.

Consider an example with $n_k = 10$, $M_b = 32$ and $\kappa = 1$, Fig. 2 depicts the number of operations of some diffusion algorithms in terms of multiplications and additions at node $k$ at each time instant versus $M$. It is clear that the computational complexity of the dLMS algorithm, with the order of $O(M)$, is much lower than that of the dRLS algorithm. As expected, since $N_u < M$, compared with the standard dRLS and R-dRLS algorithms, their DCD versions obtain about 50% reduction in both multiplications and additions for the case of...
general input regressors. However, for shift structured input regressors, the computational cost is drastically reduced from the order $O(M^2)$ to $O(M)$, which is more pronounced in scenarios with large $M$. Moreover, the multiplications required in the DCD-based algorithms are not dependent of $N_u$. On the other hand, in contrast with the existing dRLS and DCD-dRLS algorithms, the additional complexities in the proposed R-dRLS and DCD-R-dRLS algorithms resulted from the computations of the scaling factor and the bound parameter are small. In addition to the complexity, for both proposed algorithms, each node $k$ increases communication cost of $n_k - 1$ numbers for transmitting $\zeta_k$ to its neighbors.

The Metropolis rule $[73]$ used for computing the combination coefficients $\{c_{m,k}\}$ in combination step is expressed as:

$$c_{m,k} = \begin{cases} 
1 / \max(n_m, n_k), & \text{if } m \in \mathcal{N}_k, \ m \neq k \\
1 - \sum_{m \neq k} c_{m,k}, & \text{if } m = k \\
0, & \text{otherwise.}
\end{cases}$$

A. Distributed Parameter Estimation

The vector $w^o$ to be estimated has a length of $M = 16$ and a unit norm; it is generated randomly from a zero-mean uniform distribution. The input regressor $u_{k,i}$ has a shift structure, where $u_k(i)$ is colored and generated by a second-order autoregressive system:

$$u_k(i) = 1.6u_k(i - 1) - 0.81u_k(i - 2) + \epsilon_k(i),$$

where $\epsilon_k(i)$ is a zero-mean white Gaussian process with variance $\sigma^2_{\epsilon,k}$. The background noise $\theta_k(i)$ is zero-mean white Gaussian noise with variance $\sigma^2_{\theta,k}$. Variances $\sigma^2_{\epsilon,k}$ and $\sigma^2_{\theta,k}$ are shown in Fig. 3(b) and (c), respectively, for all the nodes. We employ the network MSD to assess the performance of algorithms. All results are the average over 200 independent trials.

Example 1: Except for the background noise $\theta_k(i)$, a cluster of impulses with length 200 is also added to corrupt $d_k(i)$ at iteration $i = 5001$ \footnote{Such a scenario is similar to double-talk in echo cancellation.} The cluster is drawn from a zero-mean white Gaussian process, but with a large variance $1000\sigma^2_{y_k}$ to generate impulsive samples, where $\sigma^2_{y_k}$ denotes the power of $y_k(i) = u_k^T, w^o$. Fig. 4 compares the performance of the proposed R-dRLS algorithm with that of the dRLS and both LTVFF-dRLS and LCTVFF-dRLS algorithms presented in [81]. The parameters of the algorithms are set to make a comparable convergence rate. The regularization constant for all RLS-type algorithms is chosen as $\delta = 0.01$. It is clear to

Remark 5: From the DCD-R-dRLS algorithm, we can directly obtain its special form for a single-agent scenario, referred it to as the DCD-R-RLS algorithm. In other words, the DCD-R-RLS algorithm is the DCD implementation of the algorithm presented in [59].

VI. SIMULATION RESULTS

Simulation examples are presented for a diffusion network with $N = 20$ nodes on distributed parameter estimation and distributed spectrum estimation. The network topology adopted for all simulations is shown in Fig. 3(a), unless otherwise specified. Herein, we do not consider the measurement sharing in the adaptation step for all diffusion algorithms.

Fig. 2. Complexity of the algorithms versus the length of the target vector at node $k$. (a) multiplications and (b) additions.

Fig. 3. (a) topology of network with 20 nodes, and profiles of (b) $\sigma^2_{\epsilon,k}$ and (c) $\sigma^2_{\theta,k}$ per node $k$.
see, for a small forgetting factor $\lambda = 0.98$, the conventional dRLS algorithm converges faster but has a higher estimation error; conversely, by increasing the forgetting factor, it has a lower estimation error but its convergence rate becomes slower. In particular, using a large forgetting factor $\lambda = 0.998$, the dRLS will need more time to converge again after a cluster of impulses enforces the algorithm to diverge. Due to the use of variable forgetting factor schemes, both LTVFF-dRLS and LCTVFF-dRLS algorithms solve this performance trade-off to a certain extent. As stated in Remark 2, the R-dRLS algorithm also overcomes this performance trade-off since it employs a variable ‘step-size’ factor in the adaptation step. Besides, unlike the dRLS, LTVFF-dRLS and LCTVFF-dRLS algorithms, even though a cluster of impulses does not happen until the algorithms reach the steady-state, the R-dRLS algorithm does not undergo divergence. This is because the R-dRLS algorithm can judge by (11) whether impulses occur or not and perform corresponding updates.

A fair comparison of RLS-type algorithms, we choose the same forgetting factor $\lambda=0.985$ and regularization constant $\delta=0.01$, except $\delta=0.5$ in the dRLP and RVWC-dRLS algorithms.

Fig. [5] checks the validity of the semi-analytic result (49), where we plot $E\{\xi_k(i)\}$ at node 1 (having similar results at other nodes). To take into account the assumption on input regressors $u_{k,i}$ in analysis, here its entries are generated from a white Gaussian process $c_k(i)$. To compute (49), we use the same impulsive noise parameters: $p_{r,k} = 0.01$ or $0.05$, and $\sigma_{g,k}^2 = 10000\sigma_{g,k}^2$ at all the nodes. As one can see, the theoretical results have good fit with the simulated results. Moreover, $E\{\xi_k(i)\}$ obtained by the ensemble average of simulations is a decreasing function of the iteration $i$, which further supports the theory in Appendix B.

**Example 2:** The additive noise $v_k(i)$ is a CG process given in assumption 4. At every node $k$, we set $p_{r,k}$ as a random number in the range of $[0.001, 0.05]$ and $\sigma_{g,k}^2 = 10000\sigma_{g,k}^2$. For
In Fig. 7 we compare the performance of the dRLS, dSE-LMS, dLMP, RVWC-dRLS, and dRLP algorithms with that of the proposed R-dRLS with NC algorithm. Note that, the R-dRLS (no cooperation) is that each node performs a standalone adaptive algorithm presented in [59]. As expected, the dRLS algorithm has a poor performance in the presence of impulsive noise, while other algorithms are robust. Among these robust algorithms, the convergence of dSE-LMS and dLMP algorithms is slow. Thanks to the decorrelation property of dRLS, the RVWC-dRLS, dRLP, and R-dRLS with NC algorithms obtain fast convergence. In particular, the proposed R-dRLS with NC algorithm has also a large reduction in the steady-state MSD. This is due mainly to the fact that its updated energy described by (10) and (14) diminishes with iterations.

In Fig. 6 we show the effect of the NC method. Parameter setting of algorithms: $\beta=0.97$ and $E_c=1$ (R-dRLS); $\varrho=3$ and $t_{th}=15$ (NC).

Example 3: The additive noise $v_k(i)$ here is generated by the $\alpha$-stable process, also called the $\alpha$-stable noise. Its characteristic function is given by $\varphi(t) = \exp(-\gamma |t|^\alpha)$ [72], [69], where the characteristic exponent $\alpha \in (0, 2]$ describes the impulsiveness of the noise (smaller $\alpha$ leads to more impulsive noise samples) and $\gamma > 0$ represents the dispersion level of the noise. In particular, when $\alpha = 2$, it reduces to the Gaussian noise. It is rare to find $\alpha$-stable noise with $\alpha < 1$ in practice [72], [69]. In this example, thus we set $\alpha = 1.2$ and $\gamma = 2/15$. The learning performance of the algorithms is shown in Fig. 8. Fig. 9 shows the node-wise steady-state MSD of the robust algorithms (i.e., excluding the dRLS), by averaging MSD values from iteration 2 400 to 2500. As can be seen from Figs. 8 and 9, the proposed R-dRLS algorithm outperforms the known robust diffusion algorithms in terms of convergence rate, steady-state accuracy and tracking capability. As shown in Fig. 7 to Fig. 9 due to the cooperation of interconnected nodes, the R-dRLS algorithm improves the estimation performance compared with its non-cooperative counterpart.

Fig. 7. Network MSD curves of the algorithms in impulsive noise with BG distribution. Parameter setting of algorithms (with notations from references) is as follows: $\mu_k=0.015$ (dSE-LMS); $\mu_k=0.015$ and $p=1.3$ (dLMP); $p=1.3$ (dRLP); $\beta=0.97$ and $E_c=1$ (R-dRLS); $\varrho=3$, $\tau=0.96$ and $t_{th}=15$ (NC). According to Remark 1, parameters of RVWC-dRLS are chosen as $L=16$, $\alpha=2.58$ and $\lambda=0.97$ (see [12] for detailed design of RVWC).

We also perform the simulations for the network in Fig. 10 with less connections among nodes. Fig. 11 shows the node-wise steady-state MSD of the algorithms.

Fig. 8. Network MSD curves of algorithms. $[\alpha$-stable noise]. Parameters in some of algorithms are tuned as follows: $p = 1.18$ (dLMP and dRLP); $\varrho=2$ and $t_{th}=5$ (NC).

Fig. 9. Node-wise steady-state MSD of the algorithms. $[\alpha$-stable noise].
wise steady-state MSD of those algorithms in Fig. 9. By comparing these two figures, it is seen that the proposed R-dRLS algorithm is more likely to reach the same estimates at all nodes.

![Fig. 10. Topology of a less connected network.](image)

**Example 4:** Comparison of DCD-algorithms. Figs. 12 and 13 compare the DCD-R-dRLS algorithm using different $N_u$ values with its standard version in CG-noise and $\alpha$-noise scenarios. The DCD parameters are $H = 4$ and $M_b = 16$. It is seen that, the proposed DCD-R-dRLS algorithm is also robust to impulsive noises, and approaches the R-dRLS performance with increase in $N_u$. In this example, the DCD-R-dRLS algorithm with $N_u = 4 (< M)$ has a good approximation to the R-dRLS algorithm, while the complexity of the former is significantly lower than that of the latter. Moreover, many simulations have been carried out in different impulsive noise scenarios by prolonging the iteration $i$ to a larger number than the one in Fig. 5 e.g., $5 \times 10^5$, using MATLAB R2013A on a Intel(R) Core(TM) i5-4590 CPU @ 3.30 GHz processor. We did not observe any numerical instability during the simulations for both proposed R-dRLS and DCD-dRLS algorithms.

3Here the curves of both the R-dRLS and DCD-dRLS algorithms are omitted due to their divergence performance in impulsive noise.

**B. Application: Distributed Spectrum Estimation**

We have also tested the proposed algorithms’ performance in an application of distributed spectrum estimation in CR, in which the objective is to estimate the spectrum of a transmitted signal source $s$ in the network with $N$ nodes [7], [8], [81]. We use $\phi_s(f) = \sum_{m=1}^{M} q_m(f) w_{m}^o = q^T(f) w^o$ to denote the power spectral density (PSD) of the signal $s$ at frequency $f$, where $q(f) = [q_1(f), ..., q_M(f)]^T$ is a vector consisting of basis functions evaluated at normalized frequency $f$, and $w^o = [w_1^o, ..., w_M^o]^T$ stands for the power that transmits the signal $s$ over each of $M$ basis functions and needs to be estimated. Such basis expansion can accurately model the spectrum of the signal $s$ for large enough $M$. Considering $H_{k}(f, i)$ is the transfer function of the channel between the station emitting the signal $s$ and receiver node $k$ at time instant $i$, the PSD of the received signal at node $k$ can be expressed as

$$\phi_{k, r}(f) = |H_{k}(f, i)|^2 \phi_s(f) + \sigma_{r, k}^2$$

$$= q^T_{k, i}(f) w^o + \sigma^2_{r, k},$$

(58)
where \( q_{k,i}(f) = |H_k(f,i)|q(f), \) and \( \sigma^2_{r,k} \) is the received noise power at node \( k. \)

At time instant \( i, \) each node \( k \) observes the received PSD expressed in (58) over \( N_c \) frequency samples \( f_i = f_{\min} : (f_{\max} - f_{\min})/N_c : f_{\max} \) for \( i = 1, ..., N_c; \) accordingly, the output measurements of node \( k \) obey the following relation:

\[
d_k(i) = q_{k,i}(f_i)w^o + \sigma^2_{r,k} + v_k(i),
\]

(59)

where \( v_k(i) \) denotes the observation noise at frequency \( f_i. \) The noise power \( \sigma^2_{r,k} \) can be estimated with high accuracy before the spectrum estimation, using, for example, an energy estimator over an idle band, and then subtracted from (59) [7]. Then, by collecting the output measurements over \( N_c \) frequencies, we obtain a data model at every node \( k \) for distributed spectrum estimation:

\[
d_k(i) = Q_{k,i}w^o + v_k(i),
\]

(60)

where \( Q_{k,i} = [q_{k,i}(f_1), ..., q_{k,i}(f_{N_c})]^T, \) \( d_k(i) = [d_k(i), ..., d_k(i)]^T, \) and \( v_k(i) = [v_k(i), ..., v_k(i)]^T. \)

Based on this model, we estimate the unknown spectrum \( w^o \) of the signal \( s \) using different diffusion algorithms over the network given in Fig. 3(a). In the simulation [8], [81], we use \( M = 50 \) nonoverlapping rectangular basis functions with amplitude equal to one to model the PSD of the signal \( s. \) The nodes scan \( N_c \) frequencies in the normalized frequency axis between 0 and 1. We assume that \( w^o \) has only 8 non-zero elements, meaning that the unknown spectrum is transmitted over 8 basis functions, and the power transmitted over each basis function is set to 0.7. The observation noise \( v_k(i) \) is an \( \alpha \)-stable process as in the previous Example 3 [4].

In Fig. 14, we compare the network MSD performance of different algorithms considered for the distributed spectrum estimation. As depicted, the dRLS algorithm can not identify the spectrum coefficients \( w^o \) due to its divergence in an \( \alpha \)-stable noise environment. In comparison with the dSE-LMS, dLMP, R-dRLS, and dRVL algorithms, the proposed R-dRLS and DCD-R-dRLS (with \( N_c = 4 \)) algorithms still obtain better estimation performance. We also notice from this figure that the DCD-R-dRLS algorithm with lower computational complexity approaches the R-dRLS performance. In Fig. 15, we also select the robust dRLS-type algorithms to show their performance in terms of PSD at node 1. From the results, the proposed R-dRLS and DCD-R-dRLS algorithms have lower side lobes in the PSD curves than those of the other two algorithms, thus fitting much better the true spectrum.

VII. CONCLUSION

In this paper, we have derived a new dRLS algorithm which is robust in impulsive noise, based on the minimization of a local RLS cost function with a time-dependent constraint on the squared norm of the intermediate estimate update. Following the diffusion strategy, the constraint is dynamically adjusted with the help of side information from the neighboring nodes. We also analyze the convergence of the proposed algorithm in the mean square sense under impulsive noise. Then, its

DCD version was developed to reduce the computational complexity. Moreover, to adapt the proposed algorithms to an abrupt change of the unknown parameter vector, a non-stationary control approach has also been designed. Simulation results have verified that the proposed algorithms perform better than the known algorithms in impulsive noise scenarios.

Appendix A

Verification of (25)

From Fig. 16, one can see that the left side of (25) has a good agreement with the right side of that. This reveals that the simplification from (24) to (25) is reasonable.

4 Other basis functions are also possible, e.g., raised cosines, or Gaussian bells.

5 Similar results at other nodes have not been shown here because of the page limitation.
APPENDIX B

CONVERGENCE OF $E\{\xi_k(i)\}$ TO 0

It is evident from (14) that $\xi_k(i)$ as a function of $i$ is non-increasing in adaptation process, with positive values. So, the limit of $E\{\xi_k(i)\}$ at $i \to \infty$ is existent. Applying the expectation operator to (14), we obtain

$$E\{\xi_k(i)\} = \beta E\{\xi_k(i-1)\} + (1-\beta)E\left\{\min\left(\|g_{k,i}\|_2^2, \xi_k(i-1)\right)\right\},$$

(B.1)

$$E\{\xi_k(i)\} = \sum_{m \in \mathcal{N}_k} c_{m,k} E\{\zeta_m(i)\},$$

(B.2)

Again using the assumption that the variance of $\xi_k(i)$ is small enough since $\beta$ closes to 1, we are able to make the approximation,

$$\begin{align*}
E\{\xi_k(i-1)\} & \approx E\{\xi_k(i-1)\} + E\{\xi_k(i-1)\} P_{k,i}[m_k > E\{\xi_k(i-1)\}] \\
+ E\{\xi_k(i-1)\} P_{k,i}[m_k > E\{\xi_k(i-1)\}] & \leq \sum_{m \in \mathcal{N}_k} c_{m,k} E\{\zeta_m(i)\},
\end{align*}$$

(B.3)

where $m_k = \|g_{k,i}\|_2^2$ means that both $m_k$ and $\|g_{k,i}\|_2^2$ have the same distribution, $P_{k,i}[\cdot]$ denotes the probability of event in the argument, and $F_{k,i}(m_k)$ denotes the distribution function of $m_k$ at time instant $i$.

Let us define the network global vectors as follows:

$$\xi(i) \triangleq \text{col}\{\xi_1(i), ..., \xi_N(i)\},$$

$$\zeta(i) \triangleq \text{col}\{\zeta_1(i), ..., \zeta_N(i)\}.$$  

(B.4)

Therefore, according to (B.3) and (B.4), we reformulate (B.1) and (B.2) as follows:

$$E\{\xi(i)\} = C^T \left[ \beta E\{\xi(i-1)\} + (1-\beta)(H_i E\{\xi(i-1)\} + m(i)) \right],$$

(B.5)

where

$$H_i = \text{diag}\{P_{1,i}[m_1 > E\{\xi_1(i-1)\}], ..., P_{N,i}[m_N > E\{\xi_N(i-1)\}]\},$$

(B.6)

and

$$m_i = \text{col}\left\{ \int_0^{E\{\xi_1(i-1)\}} m_1 dF_{1,i}(m_1), ..., \int_0^{E\{\xi_N(i-1)\}} m_N dF_{N,i}(m_N) \right\}.$$  

(B.7)

Taking the $\infty$-norm of both sides of (B.5) and recalling $\| C^T \|_\infty = 1$, it is found the following inequality:

$$\| E\{\xi(i)\} \|_\infty \leq \| \beta E\{\xi(i-1)\} \|_\infty + (1-\beta)\|H_i E\{\xi(i-1)\} + m(i)\|_\infty.$$  

(B.8)

Based on the diagonal definition in (B.6), we deduce an equivalent form from (B.8), i.e., for $k = 1, ..., N$,

$$E\{\xi_k(i)\} \leq \beta E\{\xi_k(i-1)\} + (1-\beta)E\{\xi_k(i-1)\} P_{k,i}[m_k > E\{\xi_k(i-1)\}] + E\{\xi_k(i-1)\} \int_0^{m_k} m_k dF_{k,i}(m_k).$$

(B.9)

It is supposed that there is a limit for $F_{k,i}(m_k)$ when $i \to \infty$, (B.9) further reduces to

$$E\{\xi_k(\infty)\} \cdot P_{k,\infty}[m_k \leq E\{\xi_k(\infty)\}] \leq E\{\xi_k(\infty)\} \int_0^{m_k} m_k dF_{k,\infty}(m_k).$$

(B.10)

In (B.10), the relation $E\{\xi_k(i)\} = E\{\xi_k(i-1)\}$ as $i \to \infty$ is also used. Hence, we consider the equality case in (B.10), i.e.,

$$E\{\xi_k(\infty)\} \cdot P_{k,\infty}[m_k \leq E\{\xi_k(\infty)\}] \leq E\{\xi_k(\infty)\} \int_0^{m_k} m_k dF_{k,\infty}(m_k).$$

(B.11)

It is shown in Appendix A in [45], for a similar equation (B.11), its solution is $E\{\xi_k(\infty)\} = 0$. Since (B.11) is an upper bound of (B.10), we can conclude that $E\{\xi_k(i)\}$ given by (14) would also converge to zero. Moreover, as its intermediate quantity, $E\{\zeta_m(i)\}$ also converges to zero.

REFERENCES

[1] Y. Yu, H. Zhao, R. C. de Lamare, and Y. Zakharov, “Robust diffusion recursive least squares estimation with side information for networked agents,” in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), April 2018, pp. 4099–4103.

[2] A. H. Sayed, “Adaptive networks,” Proceedings of the IEEE, vol. 102, no. 4, pp. 460–497, 2014.

[3] A. H. Sayed, “Adaptation, learning, and optimization over networks,” Foundations and Trends in Machine Learning, vol. 7, no. 4–5, pp. 311–801, 2014.

[4] S.-Y. Tu and A. H. Sayed, “Mobile adaptive networks,” IEEE Journal of Selected Topics in Signal Processing, vol. 5, no. 4, pp. 649–664, 2011.

[5] A. H. Sayed, S.-Y. Tu, J. Chen, X. Zhao, and Z. J. Towfic, “Diffusion strategies for adaptation and learning over networks: an examination of distributed strategies and network behavior,” IEEE Signal Processing Magazine, vol. 30, no. 3, pp. 155–171, 2013.

[6] S. Kanna, D. H. Dini, Y. Xia, S. Hui, and D. P. Mandic, “Distributed widely linear kalman filtering for frequency estimation in power networks,” IEEE Transactions on Signal and Information Processing over Networks, vol. 1, no. 1, pp. 45–57, 2015.
