



HAL
open science

Affine Combination of Diffusion Strategies over Networks

Danqi Jin, Jie Chen, Cédric Richard, Jingdong Chen, Ali H Sayed

► **To cite this version:**

Danqi Jin, Jie Chen, Cédric Richard, Jingdong Chen, Ali H Sayed. Affine Combination of Diffusion Strategies over Networks. *IEEE Transactions on Signal Processing*, 2020, 68, pp.2087-2104. 10.1109/TSP.2020.2975346 . hal-03347190

HAL Id: hal-03347190

<https://hal.science/hal-03347190>

Submitted on 17 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Affine Combination of Diffusion Strategies over Networks

Danqi Jin, *Student Member, IEEE*, Jie Chen, *Senior Member, IEEE*, Cédric Richard, *Senior Member, IEEE*,
Jingdong Chen, *Senior Member, IEEE*, Ali H. Sayed, *Fellow, IEEE*

Abstract

Diffusion adaptation is a powerful strategy for distributed estimation and learning over networks. Motivated by the concept of combining adaptive filters, this work proposes a combination framework that aggregates the operation of multiple diffusion strategies for enhanced performance. By assigning a combination coefficient to each node, and using an adaptation mechanism to minimize the network error, we obtain a combined diffusion strategy that benefits from the best characteristics of all component strategies simultaneously in terms of excess-mean-square error (EMSE). Analyses of the universality are provided to show the superior performance of affine combination scheme and to characterize its behavior in the mean and mean-square sense. Simulation results are presented to demonstrate the effectiveness of the proposed strategies, as well as the accuracy of theoretical findings.

Index Terms

Distributed optimization, diffusion strategy, affine combination, adaptive fusion strategy, stochastic performance.

I. INTRODUCTION

Distributed adaptive algorithms endow networks with the ability to estimate and track unknown parameters from streaming data in a collaborative manner. Typical existing techniques include consensus, incremental and diffusion strategies [1]–[6]. Diffusion strategies have been shown to have superior performance in adaptive scenarios where it is necessary to track drifts in the underlying models through constant step-size adaptation [7]. There are many variants of diffusion schemes, including diffusion LMS and its multitask counterparts [8]–[10], diffusion APA [11], diffusion RLS [12]. In this work, we focus on diffusion techniques due to their enhanced adaptation performance and wider stability ranges.

Diffusion algorithms typically consist of an adaptation step and a combination step. In the adaptation step, each agent updates its iterate by using a local gradient approximation. In the combination step, each agent collects intermediate estimates from its neighbors and fuses them with proper weights. The proper selection of the fusion weights is important for enhanced performance. However, finding an optimal setting for these weights is generally non-trivial. Several empirical strategies with fixed coefficients have been proposed, including averaging rule, Metropolis rule and relative-degree rule [5]. Adaptive strategies have also been derived by considering noise levels across the network [13], or relationships among agents [9]. Given that different strategies tend to deliver varying performance levels under different operating conditions, it is worth examining the possibility of combining strategies to extract the best performance possible. For example, it is useful to examine whether it

D. Jin, J. Chen and J. Chen are with Centre of Intelligent Acoustics and Immersive Communications at School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an, China (emails: danqijin@mail.nwpu.edu.cn, dr.jie.chen@ieee.org, jingdongchen@ieee.org). C. Richard is with Université Côte d'Azur, CNRS, France (email: cedric.richard@unice.fr). A. H. Sayed is with the School of Engineering, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland (email: ali.sayed@epfl.ch). The work of C. Richard was funded in part by ANR under grant ANR-19-CE48-0002.

is possible to combine two distributed strategies and obtain a new strategy whose performance is superior to its individual components. The question was answered in the affirmative for stand-alone adaptive filters in [14]–[16]. It is more challenging in the context of adaptive networks with a multitude of interacting agents over a graph. We will show nevertheless that this is still possible.

Combination strategies have been successfully used for classical adaptive filters [14], [17], multi-kernel learning [18], as well as modern deep neural network structures [19]. It is shown in some of these works that using convex combinations [14], [17] or affine combinations [20] of adaptive filters with diversity can lead to filters that combine the advantages of all component filters. Generally, such combination schemes are used to facilitate the selection of filter parameters, to increase robustness against an unknown environment, or to possibly enhance performance beyond the range of each component [15]. In this paper, we propose affine combination schemes for diffusion strategies. Each agent is designed to run several diffusion strategies in parallel, and to combine their estimates to generate the final estimates. Time-varying affine combination coefficients are set by minimizing the overall squared instantaneous error. Simulation results show that the proposed algorithms endow the networks with a significantly enhanced performance in the learning process. Some related works can be found in [21] and our previous work [22]. In [21], the authors use a useful convex combination scheme for combining two specific fusion strategies albeit without motivating it or formulating a driving optimization problem. Though similar to [21], in [22] a convex combination scheme is proposed for two diffusion strategies. However, both works do not examine the theoretical underpinnings of the algorithms. We should also note that the work proposed here is different from the useful formulation in [23]. This last reference introduces a diffusion scheme for networks with heterogeneous nodes, namely, nodes implementing different adaptive rules or differing in other aspects such as filter structure, length or step-size. In comparison, our work proposes affine combination schemes to combine different diffusion strategies, which can be either homogeneous or heterogeneous. In that sense, the results presented here are more general than [23].

The main contributions of this work are summarized as follows:

- We introduce a framework for the affine combination of two component diffusion algorithms over networks, and propose two methods for adjusting the combination weights. The framework can be extended to multiple component algorithms as explained in Appendix A.
- We establish the universality of the combined strategy, in a manner that extends a prior universality analysis for the combination of adaptive filters.
- We conduct a theoretical analysis of the performance of combined diffusion LMS strategies under some typical simplifying assumptions and approximations. Although the assumptions are not accurate in general, they are nevertheless typical in the context of studying adaptive systems and tend to lead to performance results that match well with practice for sufficiently small step-sizes with white measurement noise and white regressors [24].

Notation. Normal font x denotes scalars. Boldface small letters \boldsymbol{x} and capital letters \boldsymbol{X} denote column vectors and matrices, respectively. The superscript $(\cdot)^\top$ denotes the transpose operator. The inverse of a square matrix is denoted by $(\cdot)^{-1}$. The mathematical expectation is denoted by $\mathbb{E}\{\cdot\}$. The Gaussian distribution with mean μ and variance σ^2 is denoted by $\mathcal{N}(\mu, \sigma^2)$. The operators $\min\{\cdot\}$ and $\max\{\cdot\}$ return the minimal or maximal value of their arguments. The operator $\text{diag}\{\cdot\}$ extracts the diagonal elements of its matrix argument, or generates a diagonal matrix from its vector argument. The operator $|\cdot|$ returns

the absolute value of its argument. \mathbf{I}_N and $\mathbf{0}_N$ denote identity matrix and zero matrix of size $N \times N$, respectively. All-one vector of length N is denoted by $\mathbb{1}_N$. Symbol \mathcal{C}_i stands for cluster i , i.e., index set of nodes in the i -th cluster. \mathcal{N}_k denotes the neighbors of node k , including k .

II. NETWORK MODEL AND DIFFUSION LMS

A. Network Model

Consider a connected network consisting of N agents. The problem is to estimate an unknown parameter vector \mathbf{w}_k^* of length $L \times 1$ at each agent k . Agent k has access to temporal measurement sequences $\{d_{k,n}, \mathbf{x}_{k,n}\}$, where $d_{k,n}$ denotes a reference signal, and $\mathbf{x}_{k,n}$ is an $L \times 1$ regression vector with positive-definite covariance matrix. The data at agent k and time instant n are driven by the linear model:

$$d_{k,n} = \mathbf{x}_{k,n}^\top \mathbf{w}_k^* + z_{k,n}, \quad (1)$$

where $z_{k,n}$ is an additive noise. For ease of derivation, we introduce the following assumption **A1**¹ for $z_{k,n}$. Although rarely true in practice, **A1** simplifies the derivation of algorithm and theoretical analysis, thus it is widely adopted in the context of online learning and adaptation [5], [6].

A1: The additive noise $z_{k,n}$ is zero-mean, stationary, independent and identically distributed (i.i.d.) with variance $\sigma_{z,k}^2$, and independent of any other signal.

To determine the unknown parameter vector \mathbf{w}_k^* , we consider the following mean-square-error (MSE) cost at agent k :

$$J_k(\mathbf{w}) = \mathbb{E}\{|d_{k,n} - \mathbf{x}_{k,n}^\top \mathbf{w}|^2\}. \quad (2)$$

Observe in (1) that $J_k(\mathbf{w})$ is minimized at \mathbf{w}_k^* . For single-task problems, each agent in the network estimates the same parameter vector, while for multi-task problems, agents may estimate distinct parameter vectors [8].

B. Diffusion LMS Algorithm

The diffusion LMS algorithm is derived to minimize the following aggregate cost function:

$$J^{\text{glob}}(\mathbf{w}) = \sum_{k=1}^N J_k(\mathbf{w}) \quad (3)$$

in a cooperative manner. The general structure of the diffusion LMS algorithm consists of the following steps:

$$\phi_{k,n} = \sum_{\ell \in \mathcal{N}_k} a_{1,\ell k} \mathbf{w}_{\ell,n}, \quad (4)$$

$$\psi_{k,n+1} = \phi_{k,n} + \mu_k \sum_{\ell \in \mathcal{N}_k} c_{\ell k} \mathbf{x}_{\ell,n} (d_{\ell,n} - \mathbf{x}_{\ell,n}^\top \phi_{k,n}), \quad (5)$$

$$\mathbf{w}_{k,n+1} = \sum_{\ell \in \mathcal{N}_k} a_{2,\ell k} \psi_{\ell,n+1}, \quad (6)$$

where $\mathbf{w}_{k,n+1}$, $\psi_{k,n+1}$ and $\phi_{k,n}$ are estimates for the unknown parameter vectors \mathbf{w}_k^* obtained at different operating stages of the diffusion algorithm, with subscripts k and n denoting node index and time index, respectively, and $\mu_k > 0$ is the step-size

¹In this paper, we adopt the acronym “A” for “assumption”.

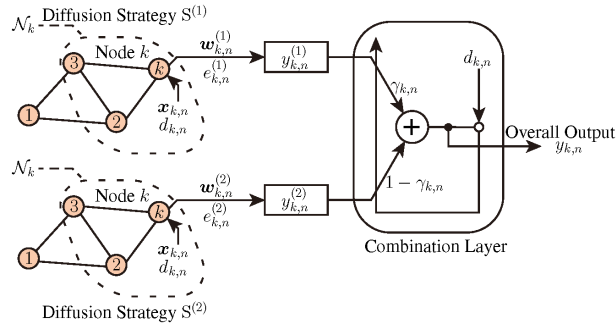


Fig. 1. Illustration of the combination framework for two diffusion strategies.

at node k , the nonnegative coefficients $a_{1,\ell k}$, $a_{2,\ell k}$ and $c_{\ell k}$ are the (ℓ, k) -th entries of two left stochastic matrices \mathbf{A}_1 , \mathbf{A}_2 and a right stochastic matrix \mathbf{C} , respectively, satisfying:

$$\mathbf{A}_1^\top \mathbf{1}_N = \mathbf{1}_N, \mathbf{A}_2^\top \mathbf{1}_N = \mathbf{1}_N, \mathbf{C} \mathbf{1}_N = \mathbf{1}_N, \quad (7)$$

$$a_{1,\ell k} = 0, a_{2,\ell k} = 0, c_{\ell k} = 0 \text{ if } \ell \notin \mathcal{N}_k. \quad (8)$$

Several adaptive strategies can be obtained as special cases of (4) to (6). Two popular strategies, namely, adapt-then-combine (ATC) and combine-then-adapt (CTA), can be achieved by setting $\mathbf{A}_1 = \mathbf{I}_N$ and $\mathbf{A}_2 = \mathbf{I}_N$ respectively.

III. GENERAL COMBINATION FRAMEWORK

A. Combination framework

Combining two diffusion strategies can be performed with two concurrent adaptive layers: a diffusion strategy layer and a combination layer. The diffusion strategy layer consists of a distributed network running two distinct diffusion strategies, say $S^{(1)}$ and $S^{(2)}$, individually and simultaneously, with associated quantities denoted by $\mathbf{A}_1^{(i)}$, $\mathbf{A}_2^{(i)}$, $\mathbf{C}^{(i)}$, $\mu^{(i)}$, where the superscript (i) denotes the i -th component diffusion strategy $S^{(i)}$ with $i = 1, 2$. The scheme is illustrated in Fig. 1. In $S^{(1)}$ and $S^{(2)}$, the agents have access to identical input and reference signals, and produce each an individual estimate of the optimal weight vector. We associate the combination coefficients $\gamma_{k,n}$ and $1 - \gamma_{k,n}$, to $S^{(1)}$ and $S^{(2)}$, respectively, at each agent k and time instant n . The goal of the combination layer is to learn which diffusion strategy performs better at each time instant, by adjusting $\gamma_{k,n}$ in order to optimize the overall network performance.

For each $S^{(i)}$, we define the filter output $y_{k,n}^{(i)}$, the estimation error $e_{k,n}^{(i)}$, and the a priori output estimation error $\tilde{e}_{k,n}^{(i)}$:

$$y_{k,n}^{(i)} \triangleq \mathbf{x}_{k,n}^\top \mathbf{w}_{k,n}^{(i)}, \quad (9)$$

$$e_{k,n}^{(i)} \triangleq d_{k,n} - \mathbf{x}_{k,n}^\top \mathbf{w}_{k,n}^{(i)}, \quad (10)$$

$$\tilde{e}_{k,n}^{(i)} \triangleq \mathbf{x}_{k,n}^\top (\mathbf{w}_k^* - \mathbf{w}_{k,n}^{(i)}). \quad (11)$$

By combining the estimates of two diffusion strategies $S^{(1)}$ and $S^{(2)}$ at each agent k with coefficients $\gamma_{k,n}$ and $1 - \gamma_{k,n}$, we

arrive at:

$$\mathbf{w}_{k,n} \triangleq \gamma_{k,n} \mathbf{w}_{k,n}^{(1)} + (1 - \gamma_{k,n}) \mathbf{w}_{k,n}^{(2)}, \quad (12)$$

$$y_{k,n} = \gamma_{k,n} y_{k,n}^{(1)} + (1 - \gamma_{k,n}) y_{k,n}^{(2)}, \quad (13)$$

$$e_{k,n} = \gamma_{k,n} e_{k,n}^{(1)} + (1 - \gamma_{k,n}) e_{k,n}^{(2)}, \quad (14)$$

$$\tilde{e}_{k,n} = \gamma_{k,n} \tilde{e}_{k,n}^{(1)} + (1 - \gamma_{k,n}) \tilde{e}_{k,n}^{(2)}. \quad (15)$$

The problem then becomes one of deriving a strategy for adjusting $\gamma_{k,n}$ based on the minimum mean-square-error (MMSE) criterion. Convex combination schemes require that $\gamma_{k,n} \in [0, 1]$. There is no such constraint in affine combination schemes. It is sufficient in this work to consider affine combination schemes to convey the main ideas.

B. Affine combination schemes

The MSE at time instant n of the entire network at the output of the combination layer is defined by:

$$J_n^{\text{MSE}} = \frac{1}{2} \sum_{k=1}^N \mathbb{E}\{e_{k,n}^2\}. \quad (16)$$

We suggest to adjust $\gamma_{k,n}$ by minimizing (16). Using (14), setting the derivative of J_n^{MSE} with respect to $\gamma_{k,n}$ to zero, and using the relations:

$$e_{k,n}^{(i)} = \tilde{e}_{k,n}^{(i)} + z_{k,n} \quad (17)$$

$$\mathbb{E}\{z_{k,n}\} = 0, \quad (18)$$

we obtain the optimal value of $\gamma_{k,n}$ in the MMSE sense:

$$\gamma_{k,n}^* = \frac{\mathbb{E}\{(\tilde{e}_{k,n}^{(2)})^2\} - \mathbb{E}\{\tilde{e}_{k,n}^{(1)} \tilde{e}_{k,n}^{(2)}\}}{\mathbb{E}\{(\tilde{e}_{k,n}^{(1)})^2\} + \mathbb{E}\{(\tilde{e}_{k,n}^{(2)})^2\} - 2\mathbb{E}\{\tilde{e}_{k,n}^{(1)} \tilde{e}_{k,n}^{(2)}\}}. \quad (19)$$

However, it is not possible to evaluate $\gamma_{k,n}^*$ with (19) as it requires knowledge of the second-order moments of $\tilde{e}_{k,n}^{(i)}$. We address this problem by introducing adaptive strategies.

1) *Affine power-normalized scheme:* Using the gradient descent method to minimize (16), and approximating the expectation terms with their instantaneous values, yield the following affine power-normalized LMS iteration [25]:

$$\begin{aligned} \gamma_{k,n+1} &= \gamma_{k,n} - \frac{\nu_{\gamma_k}}{\varepsilon + p_{k,n}} \frac{\partial J_n^{\text{MSE}}}{\partial \gamma_{k,n}} \\ &\approx \gamma_{k,n} + \frac{\nu_{\gamma_k}}{\varepsilon + p_{k,n}} e_{k,n} \mathbf{x}_{k,n}^\top (\mathbf{w}_{k,n}^{(1)} - \mathbf{w}_{k,n}^{(2)}), \end{aligned} \quad (20)$$

where ε is a small positive parameter, ν_{γ_k} is a positive step-size, and $p_{k,n}$ is a low-pass filtered estimate of the power of $\mathbf{x}_{k,n}^\top (\mathbf{w}_{k,n}^{(1)} - \mathbf{w}_{k,n}^{(2)})$ given by:

$$p_{k,n} = \eta p_{k,n-1} + (1 - \eta) [\mathbf{x}_{k,n}^\top (\mathbf{w}_{k,n}^{(1)} - \mathbf{w}_{k,n}^{(2)})]^2, \quad (21)$$

with $0 < \eta < 1$ a temporal smoothing factor.

2) *Affine sign-regressor scheme*: Alternatively, we can adopt another normalization scheme for the step-size, leading to the affine sign-regressor LMS iteration [26]:

$$\begin{aligned}\gamma_{k,n+1} &= \gamma_{k,n} - \nu'_{\gamma_k} \frac{\partial J_n^{\text{MSE}}}{\partial \gamma_{k,n}} \\ &\approx \gamma_{k,n} + \nu_{\gamma_k} e_{k,n} \text{sgn}\{\mathbf{x}_{k,n}^\top (\mathbf{w}_{k,n}^{(1)} - \mathbf{w}_{k,n}^{(2)})\}\end{aligned}\quad (22)$$

where $\text{sgn}\{x\}$ is the sign function and $\nu'_{\gamma_k} = \frac{\nu_{\gamma_k}}{|\mathbf{x}_{k,n}^\top (\mathbf{w}_{k,n}^{(1)} - \mathbf{w}_{k,n}^{(2)})|}$.

IV. THEORETICAL ANALYSIS OF POWER-NORMALIZED SCHEME

A. Universality at steady state

We first illustrate the universality of the power-normalized scheme at steady state. In other words, we show that the algorithm results in a combined strategy tracks the best performance of each component strategy.

We start the analysis by defining several quantities to be used later. The EMSE at the output of the combination layer, and the EMSE of each component strategy, at node k and time instant n are defined as:

$$\begin{aligned}J_{\text{ex},k,n} &\triangleq \mathbb{E}\{(\tilde{e}_{k,n})^2\} \\ &= \mathbb{E}\{[\gamma_{k,n}\tilde{e}_{k,n}^{(1)} + (1 - \gamma_{k,n})\tilde{e}_{k,n}^{(2)}]^2\},\end{aligned}\quad (23)$$

$$J_{\text{ex},k,n}^{(i)} \triangleq \mathbb{E}\{(\tilde{e}_{k,n}^{(i)})^2\}.\quad (24)$$

Correspondingly, the EMSE of the whole network, at the output of the combination layer and for each component strategy, are defined by:

$$J_{\text{ex,net},n} \triangleq \sum_{k=1}^N J_{\text{ex},k,n} \quad (25)$$

$$J_{\text{ex,net},n}^{(i)} \triangleq \sum_{k=1}^N J_{\text{ex},k,n}^{(i)} \quad (26)$$

respectively. By taking the limit as $n \rightarrow \infty$, we obtain the corresponding values at steady-state: $J_{\text{ex},k,\infty}$, $J_{\text{ex},k,\infty}^{(i)}$, $J_{\text{ex,net},\infty}$ and $J_{\text{ex,net},\infty}^{(i)}$. In addition, by resorting to (11) and (15), we have:

$$\mathbf{x}_{k,n}^\top (\mathbf{w}_{k,n}^{(1)} - \mathbf{w}_{k,n}^{(2)}) = \tilde{e}_{k,n}^{(2)} - \tilde{e}_{k,n}^{(1)} \quad (27)$$

and

$$e_{k,n} = \gamma_{k,n}\tilde{e}_{k,n}^{(1)} + (1 - \gamma_{k,n})\tilde{e}_{k,n}^{(2)} + z_{k,n}.\quad (28)$$

Now we introduce some approximations to be used later to simplify the theoretical analysis:

Ap₁²: At steady state, the combination coefficient $\gamma_{k,n}$ is statistically independent of $\tilde{e}_{k,n}^{(i)}$ and $p_{k,n}$.

Ap₂: For a sufficiently large temporal smoothing factor η , $p_{k,n}$ is statistically independent of $\mathbf{x}_{k,n}^\top \mathbf{w}_{k,n}^{(i)}$, that is, of $\tilde{e}_{k,n}^{(i)}$.

As indicated in [14], approximation **Ap₁** is reasonable when adopting a decaying step-size ν_{γ_k} , and **Ap₂** is justified when using a large temporal smoothing factor η . With approximations **Ap₁** and **Ap₂**, we obtain the following results.

²In this paper, we adopt the acronym “**Ap**” for “approximation”.

Universality Analysis Result 1: Assume data model (1), assumption **A1** and approximations **Ap₁**, **Ap₂** hold. Then for any initial conditions with step-size ν_{γ_k} ensuring the stability of power-normalized scheme, the distributed diffusion network with (20) is universal at steady state, which means that the EMSE of the diffusion network after combination cannot be worse than that of the best component strategies, with

$$J_{\text{ex,net},\infty} \leq \min \{J_{\text{ex,net},\infty}^{(1)}, J_{\text{ex,net},\infty}^{(2)}\}. \quad (29)$$

Proof: See Appendix B. ■

B. Mean weight and mean-square behaviors analyses

We shall now examine the mean and mean-square error behavior of the power-normalized scheme on the basis of the two diffusion LMS strategies referred to in Section II-B. Collecting the quantities from the network into block vectors, we have:

$$\mathbf{w}^* \triangleq \text{col}\{\mathbf{w}_1^*, \dots, \mathbf{w}_N^*\}, \quad (30)$$

$$\mathbf{w}_n \triangleq \text{col}\{\mathbf{w}_{1,n}, \dots, \mathbf{w}_{N,n}\}, \quad (31)$$

$$\mathbf{w}_n^{(i)} \triangleq \text{col}\{\mathbf{w}_{1,n}^{(i)}, \dots, \mathbf{w}_{N,n}^{(i)}\}, \quad (32)$$

where \mathbf{w}^* is the block optimum weight vector, \mathbf{w}_n and $\mathbf{w}_n^{(i)}$ are the block weight estimates of the combination layer and component diffusion strategies $S^{(i)}$, respectively. Using (31), (32), and from (12) we arrive at:

$$\mathbf{w}_n = \mathbf{\Gamma}_n \mathbf{w}_n^{(1)} + (\mathbf{I}_{NL} - \mathbf{\Gamma}_n) \mathbf{w}_n^{(2)}, \quad (33)$$

where $\mathbf{\Gamma}_n$ is a diagonal weighting matrix defined by

$$\mathbf{\Gamma}_n \triangleq \text{diag}\{\gamma_{1,n}, \dots, \gamma_{N,n}\} \otimes \mathbf{I}_L, \quad (34)$$

with symbol \otimes denoting Kronecker product. The weight error vectors at node k for the component diffusion strategies $S^{(i)}$ and for the combination layer, as well as those of the entire network, are defined by:

$$\mathbf{v}_{k,n}^{(i)} \triangleq \mathbf{w}_{k,n}^{(i)} - \mathbf{w}_k^*, \quad (35)$$

$$\mathbf{v}_{k,n} \triangleq \mathbf{w}_{k,n} - \mathbf{w}_k^*, \quad (36)$$

$$\mathbf{v}_n^{(i)} \triangleq \text{col}\{\mathbf{v}_{1,n}^{(i)}, \dots, \mathbf{v}_{N,n}^{(i)}\}, \quad (37)$$

$$\mathbf{v}_n \triangleq \text{col}\{\mathbf{v}_{1,n}, \dots, \mathbf{v}_{N,n}\}. \quad (38)$$

Using (30)–(38), vector \mathbf{v}_n is given by:

$$\mathbf{v}_n = \mathbf{w}_n - \mathbf{w}^* = \mathbf{\Gamma}_n \mathbf{v}_n^{(1)} + (\mathbf{I}_{NL} - \mathbf{\Gamma}_n) \mathbf{v}_n^{(2)}. \quad (39)$$

To make the theoretical analysis tractable, we introduce the following assumption **A2** and approximation **Ap₃**:

A2 (Independent Regressors): The regression vector $\mathbf{x}_{k,n}$, generated from a zero-mean random process, is temporally stationary, white (over n) and spatially independent (over k) with covariance matrix $\mathbf{R}_{x,k} = \mathbb{E}\{\mathbf{x}_{k,n} \mathbf{x}_{k,n}^\top\} > 0$.

Ap₃: At each time instant n , $\gamma_{k,n}$ is statistically independent of $\mathbf{w}_{k,n}^{(i)}$ for $i = 1, 2$.

Although not true in general, assumption **A2** is usually adopted to simplify the derivation without constraining the conclusions. Besides, there are several results in the literature showing that performance results obtained under **A2** match well with actual

performance when the step-sizes $\mu_k^{(i)}$ of the component diffusion strategies are sufficiently small [24], [27]. \mathbf{Ap}_3 makes the theoretical analysis tractable. Though actually not true, it does not notably affect the theoretical results, as illustrated by the simulation results.

1) Mean weight behavior analysis

For the combination layer, taking expectation of (39) and using approximation \mathbf{Ap}_3 , we arrive at:

$$\mathbb{E}\{\mathbf{v}_{n+1}\} = \mathbb{E}\{\mathbf{\Gamma}_{n+1}\} \mathbb{E}\{\mathbf{v}_{n+1}^{(1)}\} + \mathbb{E}\{\mathbf{I}_{NL} - \mathbf{\Gamma}_{n+1}\} \mathbb{E}\{\mathbf{v}_{n+1}^{(2)}\}. \quad (40)$$

We now need to evaluate $\mathbb{E}\{\mathbf{v}_{n+1}^{(i)}\}$. Under assumptions $\mathbf{A1}$ – $\mathbf{A2}$, and following the derivation from [9], we have:

$$\mathbf{v}_{n+1}^{(i)} = \mathbf{B}_n^{(i)} \mathbf{v}_n^{(i)} + \mathbf{g}_n^{(i)} - \mathbf{r}_n^{(i)}, \quad (41)$$

$$\mathbb{E}\{\mathbf{v}_{n+1}^{(i)}\} = \overline{\mathbf{B}}^{(i)} \mathbb{E}\{\mathbf{v}_n^{(i)}\} - \overline{\mathbf{r}}^{(i)}, \quad (42)$$

with

$$\mathbf{B}_n^{(i)} = \mathcal{A}_2^{(i)\top} (\mathbf{I}_{NL} - \mathbf{U}^{(i)} \mathbf{H}_n^{(i)}) \mathcal{A}_1^{(i)\top}, \quad (43)$$

$$\overline{\mathbf{B}}^{(i)} = \mathcal{A}_2^{(i)\top} (\mathbf{I}_{NL} - \mathbf{U}^{(i)} \overline{\mathbf{H}}^{(i)}) \mathcal{A}_1^{(i)\top}, \quad (44)$$

$$\mathcal{A}_j^{(i)} = \mathbf{A}_j^{(i)} \otimes \mathbf{I}_L, \quad \forall j = 1, 2, \quad (45)$$

$$\mathbf{U}^{(i)} = \text{diag}\{\mu_1^{(i)}, \dots, \mu_N^{(i)}\} \otimes \mathbf{I}_L, \quad (46)$$

$$\mathbf{H}_n^{(i)} = \text{diag}\left\{\sum_{\ell \in \mathcal{N}_k} c_{\ell k}^{(i)} \mathbf{x}_{\ell, n} \mathbf{x}_{\ell, n}^\top\right\}_{k=1}^N, \quad (47)$$

$$\overline{\mathbf{H}}^{(i)} = \text{diag}\{\mathbf{R}_1^{(i)}, \dots, \mathbf{R}_N^{(i)}\}, \quad (48)$$

$$\mathbf{R}_k^{(i)} \triangleq \sum_{\ell \in \mathcal{N}_k} c_{\ell k}^{(i)} \mathbf{R}_{x, \ell}, \quad (49)$$

$$\mathbf{g}_n^{(i)} = \mathcal{A}_2^{(i)\top} \mathbf{U}^{(i)} \mathbf{p}_{z_x, n}^{(i)}, \quad (50)$$

$$\mathbf{p}_{z_x, n}^{(i)} = \text{col}\left\{\sum_{\ell \in \mathcal{N}_k} c_{\ell k}^{(i)} \mathbf{x}_{\ell, n} z_{\ell, n}\right\}_{k=1}^N, \quad (51)$$

$$\mathbf{h}_{u, n}^{(i)} = \text{col}\left\{\sum_{\ell \in \mathcal{N}_k} c_{\ell k}^{(i)} \mathbf{x}_{\ell, n} \mathbf{x}_{\ell, n}^\top (\mathbf{w}_k^* - \mathbf{w}_\ell^*)\right\}_{k=1}^N, \quad (52)$$

$$\begin{aligned} \mathbf{r}_n^{(i)} &\triangleq \underbrace{\mathcal{A}_2^{(i)\top} \mathbf{U}^{(i)} \mathbf{h}_{u, n}^{(i)}}_{\mathbf{r}_{\mathbf{u}, n}^{(i)}} - \\ &\underbrace{\left[\mathcal{A}_2^{(i)\top} (\mathbf{I}_{NL} - \mathbf{U}^{(i)} \mathbf{H}_n^{(i)}) (\mathcal{A}_1^{(i)\top} - \mathbf{I}_{NL}) + (\mathcal{A}_2^{(i)\top} - \mathbf{I}_{NL})\right] \mathbf{w}^*}_{\mathbf{r}_{\mathbf{w}, n}^{(i)}} \end{aligned} \quad (53)$$

$$\overline{\mathbf{r}}^{(i)} \triangleq \mathbb{E}\{\mathbf{r}_n^{(i)}\} = \overline{\mathbf{r}}_{\mathbf{u}}^{(i)} - \overline{\mathbf{r}}_{\mathbf{w}}^{(i)}, \quad (54)$$

where the symbols with an overhead bar denote the expectation of the corresponding quantities with subscript n .

Stability Analysis Result 1: (Stability in the mean) Assume data model (1), $\mathbf{A1}$ – $\mathbf{A2}$ and \mathbf{Ap}_3 hold. Then, for any initial condition, the distributed network with power-normalized diffusion scheme (20) asymptotically converges in the mean if the

step-sizes are chosen to satisfy:

$$0 < \mu_k^{(i)} < \frac{2}{\lambda_{\max}\{\mathbf{R}_k^{(i)}\}}, \quad k = 1, \dots, N \text{ and } i = 1, 2, \quad (55)$$

where $\lambda_{\max}\{\cdot\}$ denotes the largest eigenvalue of its matrix argument, and if

$$0 < \nu_{\gamma_k} < 1 - \eta, \quad (56)$$

where η is the temporal smoothing factor used in (21). The asymptotic bias is given by:

$$\begin{aligned} \mathbb{E}\{\mathbf{v}_\infty\} &= -\mathbf{\Gamma}_\infty (\mathbf{I}_{NL} - \overline{\mathbf{B}}^{(1)})^{-1} \overline{\mathbf{r}}^{(1)} \\ &\quad - (\mathbf{I}_{NL} - \mathbf{\Gamma}_\infty) (\mathbf{I}_{NL} - \overline{\mathbf{B}}^{(2)})^{-1} \overline{\mathbf{r}}^{(2)}. \end{aligned} \quad (57)$$

Proof: See Appendix C. ■

2) Mean-square behavior analysis

To perform the mean-square behavior analysis, we introduce the following approximation.

Ap₄: At each time instant n , $\mathbf{\Gamma}_{n+1}$ is statistically independent of $\mathbf{B}_n^{(i)}$, $\mathbf{v}_n^{(i)}$, $\mathbf{g}_n^{(i)}$ and $\mathbf{r}_n^{(i)}$ in (41) for $i = 1, 2$.

Although this approximation is not true in general, it will make the analysis tractable.

We shall now evaluate the evolution of $\mathbb{E}\{\|\mathbf{v}_{n+1}\|_\Sigma^2\}$ over time, where Σ is an arbitrary positive semi-definite matrix, and $\|\mathbf{x}\|_\Sigma^2 \triangleq \mathbf{x}^\top \Sigma \mathbf{x}$. Using (39), we have:

$$\begin{aligned} \mathbb{E}\{\|\mathbf{v}_{n+1}\|_\Sigma^2\} &= 2\mathbb{E}\{\mathbf{v}_{n+1}^{(1)\top} \mathbf{\Gamma}_{n+1} \Sigma (\mathbf{I}_{NL} - \mathbf{\Gamma}_{n+1}) \mathbf{v}_{n+1}^{(2)}\} \\ &+ \mathbb{E}\{\|(\mathbf{I}_{NL} - \mathbf{\Gamma}_{n+1}) \mathbf{v}_{n+1}^{(2)}\|_\Sigma^2\} + \mathbb{E}\{\|\mathbf{\Gamma}_{n+1} \mathbf{v}_{n+1}^{(1)}\|_\Sigma^2\}. \end{aligned} \quad (58)$$

Let

$$\Sigma_{n+1}^{(1)} \triangleq \mathbb{E}\{\mathbf{\Gamma}_{n+1}^\top \Sigma \mathbf{\Gamma}_{n+1}\}, \quad (59)$$

$$\Sigma_{n+1}^{(2)} \triangleq \mathbb{E}\{(\mathbf{I}_{NL} - \mathbf{\Gamma}_{n+1})^\top \Sigma (\mathbf{I}_{NL} - \mathbf{\Gamma}_{n+1})\}, \quad (60)$$

$$\sigma_{n+1}^{(i)} = \text{vec}\{\Sigma_{n+1}^{(i)}\}, \quad (61)$$

where $\text{vec}\{\cdot\}$ operator stacks the columns of its matrix argument on top of each other. Under **Ap₃**, the last two terms on the RHS of (58) can be written in compact form as $\mathbb{E}\{\|\mathbf{v}_{n+1}^{(i)}\|_{\Sigma_{n+1}^{(i)}}^2\}$ with $i = 1, 2$. They can be evaluated compactly as:

$$\begin{aligned} \mathbb{E}\{\|\mathbf{v}_{n+1}^{(i)}\|_{\sigma_{n+1}^{(i)}}^2\} &= \mathbb{E}\{\|\mathbf{v}_n^{(i)}\|_{\mathbf{K}^{(i)} \sigma_{n+1}^{(i)}}^2\} + \\ &[\text{vec}\{\mathbf{G}^{(i)\top}\}]^\top \sigma_{n+1}^{(i)} + \mathbf{f}(\overline{\mathbf{r}}^{(i)}, \Sigma_{n+1}^{(i)}, \mathbb{E}\{\mathbf{v}_n^{(i)}\}), \end{aligned} \quad (62)$$

where $\|\cdot\|_{\Sigma_{n+1}^{(i)}}^2$ and $\|\cdot\|_{\sigma_{n+1}^{(i)}}^2$ are used interchangeably, with:

$$\mathbf{f}(\mathbf{r}_n^{(i)}, \Sigma_{n+1}^{(i)}, \mathbf{v}_n^{(i)}) \triangleq \|\mathbf{r}_n^{(i)}\|_{\Sigma_{n+1}^{(i)}}^2 - 2\mathbf{r}_n^{(i)\top} \Sigma_{n+1}^{(i)} \mathbf{B}_n^{(i)} \mathbf{v}_n^{(i)} \quad (63)$$

$$\mathbf{K}^{(i)} \approx \overline{\mathbf{B}}^{(i)\top} \otimes \overline{\mathbf{B}}^{(i)\top} \quad (64)$$

$$\mathbf{G}^{(i)} \triangleq \mathbb{E}\{\mathbf{g}_n^{(i)} \mathbf{g}_n^{(i)\top}\}. \quad (65)$$

The derivation of equation (62) is provided in Appendix D.

For the first term on RHS of (58), since it has a similar structure to (107), the analysis is omitted. Using (41), following the routine (107)–(112), and ignoring second-order terms in the step-size, we finally arrive at:

$$\begin{aligned} \mathbb{E}\{\mathbf{v}_{n+1}^{(1)\top} \boldsymbol{\Sigma}_{x,n+1} \mathbf{v}_{n+1}^{(2)}\} &= \\ \mathbb{E}\{\mathbf{v}_n^{(1)\top} \boldsymbol{\Sigma}_{xc,n+1} \mathbf{v}_n^{(2)}\} &+ [\text{vec}\{\mathbf{G}_x^\top\}]^\top \boldsymbol{\sigma}_{x,n+1} + \\ \mathbf{f}_x(\bar{\mathbf{r}}^{(1)}, \bar{\mathbf{r}}^{(2)}, \boldsymbol{\Sigma}_{x,n+1}, \mathbb{E}\{\mathbf{v}_n^{(1)}\}, \mathbb{E}\{\mathbf{v}_n^{(2)}\}, \bar{\mathbf{B}}^{(1)}, \bar{\mathbf{B}}^{(2)}), \end{aligned} \quad (66)$$

where

$$\begin{aligned} \mathbf{f}_x(\bar{\mathbf{r}}^{(1)}, \bar{\mathbf{r}}^{(2)}, \boldsymbol{\Sigma}_{x,n+1}, \mathbb{E}\{\mathbf{v}_n^{(1)}\}, \mathbb{E}\{\mathbf{v}_n^{(2)}\}, \bar{\mathbf{B}}^{(1)}, \bar{\mathbf{B}}^{(2)}) \\ \triangleq \bar{\mathbf{r}}^{(1)\top} \boldsymbol{\Sigma}_{x,n+1} \bar{\mathbf{r}}^{(2)} - \mathbb{E}\{\mathbf{v}_n^{(1)\top}\} \bar{\mathbf{B}}^{(1)\top} \boldsymbol{\Sigma}_{x,n+1} \bar{\mathbf{r}}^{(2)} \\ - \bar{\mathbf{r}}^{(1)\top} \boldsymbol{\Sigma}_{x,n+1} \bar{\mathbf{B}}^{(2)} \mathbb{E}\{\mathbf{v}_n^{(2)}\} \end{aligned} \quad (67)$$

with

$$\boldsymbol{\Sigma}_{x,n+1} \triangleq \mathbb{E}\{\boldsymbol{\Gamma}_{n+1} \boldsymbol{\Sigma} (\mathbf{I}_{NL} - \boldsymbol{\Gamma}_{n+1})\}, \quad (68)$$

$$\boldsymbol{\Sigma}_{xc,n+1} \triangleq \text{vec}^{-1}\{\mathbf{K}_x \boldsymbol{\sigma}_{x,n+1}\}, \quad (69)$$

$$\boldsymbol{\sigma}_{x,n+1} \triangleq \text{vec}\{\boldsymbol{\Sigma}_{x,n+1}\}, \quad (70)$$

$$\mathbf{K}_x \approx \bar{\mathbf{B}}^{(2)\top} \otimes \bar{\mathbf{B}}^{(1)\top}, \quad (71)$$

$$\mathbf{G}_x \triangleq \mathbb{E}\{\mathbf{g}_n^{(2)} \mathbf{g}_n^{(1)\top}\}, \quad (72)$$

and $\text{vec}^{-1}\{\cdot\}$ is the inverse vectorization operator.

Finally, using (62) for $i = 1, 2$ and (66) in (58), we obtain the explicit expression of the weighted mean-square behavior of the power-normalized diffusion scheme.

Stability Analysis Result 2: (Mean-square Stability) Assume data model (1) and **A1–A2**, **Ap₃–Ap₄** hold. Assume further that the step-sizes $\mu_k^{(i)}$ for $i = 1, 2$ are sufficiently small such that condition (55) is satisfied and approximations (64), (71), (112) are justified by ignoring higher powers of the step-size. Furthermore, assume that the step-size ν_{γ_k} at the combination layer satisfies the condition:

$$0 < \nu_{\gamma_k} < \frac{1 - \eta}{3} \quad (73)$$

to ensure mean-square stability of the power-normalized scheme. Then, for any initial conditions, the distributed network with doubly stochastic matrices $\mathbf{A}_1^{(i)}, \mathbf{A}_2^{(i)}$, of which both columns and rows add up to one, and scheme (20) is mean-square stable if $\rho(\mathbf{I}_{NL} - \mathbf{U}^{(i)} \bar{\mathbf{H}}^{(i)}) < 1$, which is further guaranteed for sufficiently small step-sizes that also satisfy condition (55).

Proof: See Appendix E. ■

Remark 1. (*Transient MSD*): We shall adopt the mean-square-deviation (MSD) learning curve of the entire network, defined by $\xi_{n+1} \triangleq \mathbb{E}\{\|\mathbf{v}_{n+1}\|_{\frac{1}{N} \mathbf{I}_{NL}}^2\}$, as a metric to evaluate the performance of diffusion networks. We observe that although the weighting matrix $\boldsymbol{\Sigma}$ of $\mathbb{E}\{\|\mathbf{v}_{n+1}\|_{\boldsymbol{\Sigma}}^2\}$ in (58) may be constant over time, such as choosing $\boldsymbol{\Sigma} = \frac{1}{N} \mathbf{I}_{NL}$ in evaluating the MSD, matrices $\boldsymbol{\Sigma}_{n+1}^{(i)}$ and $\boldsymbol{\Sigma}_{x,n+1}$ become time-variant since $\boldsymbol{\Gamma}_{n+1}$ varies over time. Thus, it is almost impossible to derive a recursion to relate $\mathbb{E}\{\|\mathbf{v}_{n+1}\|_{\boldsymbol{\Sigma}}^2\}$ and $\mathbb{E}\{\|\mathbf{v}_n\|_{\boldsymbol{\Sigma}}^2\}$ directly. To evaluate $\mathbb{E}\{\|\mathbf{v}_{n+1}\|_{\boldsymbol{\Sigma}}^2\}$, we must evaluate (62) and (66), in

which $\mathbb{E}\{\|\mathbf{v}_n^{(i)}\|_{\mathbf{K}^{(i)}\boldsymbol{\sigma}_{n+1}^{(i)}}^2\}$ and $\mathbb{E}\{\mathbf{v}_n^{(1)\top}\boldsymbol{\Sigma}_{\text{xc},n+1}\mathbf{v}_n^{(2)}\}$ are calculated iteratively as shown by expressions (121) and (123) in Appendix F, while the remaining terms can be evaluated directly.

Remark 2. (Steady-state MSD): For sufficiently small step-sizes satisfying conditions (55) and (73) to ensure the mean and mean-square stabilities of the power-normalized diffusion scheme, we can obtain the explicit expression of the steady-state MSD as shown by expression (132) in Appendix G.

Proof: See Appendix G. ■

C. Mean and mean-square behaviors of $\gamma_{k,n}$

In order to evaluate the mean and mean-square behavior of the power-normalized diffusion scheme, we must determine the mean and mean-square behavior of $\boldsymbol{\Gamma}_n$ at the combination layer, which are obtained by evaluating those of $\gamma_{k,n}$ since $\boldsymbol{\Gamma}_n$ is diagonal. To make the analysis tractable, we introduce the following approximations.

Ap₅: The combination coefficient $\gamma_{k,n}$ varies slowly enough so that the correlation between $\gamma_{k,n}$ and $\tilde{e}_{k,n}^{(m)}\tilde{e}_{k,n}^{(n)}$ for $m, n = 1, 2$ can be ignored.

Ap₆: For a large enough temporal smoothing factor η , $p_{k,n}$ is statistically independent of $\gamma_{k,n}$.

Ap₇: The a priori errors $\tilde{e}_{k,n}^{(1)}$ and $\tilde{e}_{k,n}^{(2)}$ are jointly Gaussian with zero-mean, which implies [28]:

$$\mathbb{E}\{(\tilde{e}_{k,n}^{(i)})^4\} = 3(J_{\text{ex},k,n}^{(i)})^2, \quad \forall i = 1, 2, \quad (74)$$

$$\mathbb{E}\{(\tilde{e}_{k,n}^{(1)})^3(\tilde{e}_{k,n}^{(2)})^1\} = 3J_{\text{ex},k,n}^{(1)}J_{\text{ex},k,n}^{(1,2)}, \quad (75)$$

$$\mathbb{E}\{(\tilde{e}_{k,n}^{(1)})^1(\tilde{e}_{k,n}^{(2)})^3\} = 3J_{\text{ex},k,n}^{(1,2)}J_{\text{ex},k,n}^{(2)}, \quad (76)$$

$$\mathbb{E}\{(\tilde{e}_{k,n}^{(1)})^2(\tilde{e}_{k,n}^{(2)})^2\} = 2(J_{\text{ex},k,n}^{(1,2)})^2 + J_{\text{ex},k,n}^{(1)}J_{\text{ex},k,n}^{(2)}. \quad (77)$$

Approximation **Ap₅** is commonly adopted in transient analysis of affine combinations of two adaptive filters [26], which also coincides with simulation result that $\gamma_{k,n}$ converges slowly compared to the variations of input signal $\mathbf{x}_{k,n}$, thus to the variations of a priori errors $\tilde{e}_{k,n}^{(i)}$. Although not true in general, **Ap₆** makes the analysis tractable. Although may be violated in general, assumption **Ap₇** is frequently adopted to facilitate the transient analysis of adaptive filters [6], [26], [29]–[33]. It becomes more reasonable for small step-sizes and long filters [6].

Starting from (20), and following the derivation in Appendix H, we arrive at the following stability analysis results.

Stability Analysis Result 3: (Stability in the Mean) Assume data model (1), assumption **A1** and approximations **Ap₂**, **Ap₅**, **Ap₆** hold. Then for any initial conditions, the power-normalized scheme (20) asymptotically converges in the mean if the step-sizes ν_{γ_k} are chosen to satisfy condition (56). Besides, the mean behavior of $\gamma_{k,n}$ is evaluated as (136) in Appendix H, with steady-state value $\mathbb{E}\{\gamma_{k,\infty}\}$ given by equation (95).

Proof: See Appendix H. ■

Stability Analysis Result 4: (Mean-square Stability) Assume data model (1), assumption **A1** and approximations **Ap₂**, **Ap₅–Ap₇** hold. Then for any initial conditions, the power-normalized scheme (20) is mean-square stable if the step-sizes ν_{γ_k} are chosen to satisfy condition (73). Besides, the transient mean-square behavior of $\gamma_{k,n}$ is evaluated by (141) in Appendix H, with steady-state value $\mathbb{E}\{\gamma_{k,\infty}^2\}$ given by equation (78).

Proof: See Appendix H. ■

$$\mathbb{E}\{\gamma_{k,\infty}^2\} = \frac{\bar{\nu}_\infty (J_{\text{ex},k,\infty}^{(2)} + \sigma_{z,k}^2) (\Delta J_{k,\infty}^{(1)} + \Delta J_{k,\infty}^{(2)}) + 2\bar{\nu}_\infty (\Delta J_{k,\infty}^{(2)})^2 + 2\mathbb{E}\{\gamma_{k,\infty}\} [\Delta J_{k,\infty}^{(2)} - 3\bar{\nu}_\infty \Delta J_{k,\infty}^{(2)} (\Delta J_{k,\infty}^{(1)} + \Delta J_{k,\infty}^{(2)})]}{2(\Delta J_{k,\infty}^{(1)} + \Delta J_{k,\infty}^{(2)}) - 3\bar{\nu}_\infty (\Delta J_{k,\infty}^{(1)} + \Delta J_{k,\infty}^{(2)})^2} \quad (78)$$

V. THEORETICAL ANALYSIS OF SIGN-REGRESSOR SCHEME

By following the same routine as in Section IV, we conduct theoretical analysis for the sign-regressor diffusion scheme.

A. Universality at steady state

Taking expectation of (22), and using (27), (28), we obtain:

$$\begin{aligned} \mathbb{E}\{\gamma_{k,n+1}\} &= \mathbb{E}\{\gamma_{k,n}\} + \nu_{\gamma_k} \mathbb{E}\{\tilde{e}_{k,n}^{(2)} \text{sgn}\{\tilde{e}_{k,n}^{(2)} - \tilde{e}_{k,n}^{(1)}\} \\ &\quad - \gamma_{k,n} |\tilde{e}_{k,n}^{(2)} - \tilde{e}_{k,n}^{(1)}|\}. \end{aligned} \quad (79)$$

From (79), the stationary point of $\gamma_{k,n}$ is reached if

$$\mathbb{E}\{\tilde{e}_{k,n}^{(2)} \text{sgn}\{\tilde{e}_{k,n}^{(2)} - \tilde{e}_{k,n}^{(1)}\} - \gamma_{k,n} |\tilde{e}_{k,n}^{(2)} - \tilde{e}_{k,n}^{(1)}|\} = 0. \quad (80)$$

Using the independence approximation **Ap**₁, (80) gives a closed-form solution of $\mathbb{E}\{\gamma_{k,n}\}$ at steady-state as:

$$\mathbb{E}\{\gamma_{k,\infty}\} = \frac{\mathbb{E}\{\tilde{e}_{k,\infty}^{(2)} \text{sgn}\{\tilde{e}_{k,\infty}^{(2)} - \tilde{e}_{k,\infty}^{(1)}\}\}}{\mathbb{E}\{|\tilde{e}_{k,\infty}^{(2)} - \tilde{e}_{k,\infty}^{(1)}|\}}. \quad (81)$$

We further resort to the joint Gaussian approximation **Ap**₇, which leads to approximations (148) and (149) further ahead in Section V-C. Using (148) and (149), (81) simplifies to (95). With (95), the proof of universality for the sign-regressor diffusion scheme at steady state is identical to that for power-normalized diffusion scheme.

B. Mean weight and mean-square behaviors analyses

Since there is almost no difference in the mean weight and mean-square behaviors of the sign-regressor and power-normalized schemes, we only provide the main conclusions here. Under assumptions **A1–A2**, approximations **Ap**_{3–Ap}₄ and following the same routine as that in Section IV-B, the mean and mean-square behaviors of the weight error vector \mathbf{v}_{n+1} at combination layer of the sign-regressor diffusion scheme are given by (40) and (58), respectively.

Stability Analysis Result 5: (Stability in the Mean) Assume data model (1), **A1–A2** and **Ap**₃ hold. Then for any initial conditions, the distributed network with sign-regressor diffusion scheme (22) asymptotically converges in the mean if the step-sizes of network are chosen to satisfy (55), and if the step-sizes at the combination layer are chosen to satisfy

$$0 < \nu_{\gamma_k} < \sqrt{\frac{\pi}{2 \cdot \max_n \{\Delta J_{k,n}^{(1)} + \Delta J_{k,n}^{(2)}\}}}. \quad (82)$$

The asymptotic bias is given by (57).

Stability Analysis Result 6: (Mean-square Stability) Assume data model (1) and **A1–A2**, **Ap**_{3–Ap}₄ hold. Assume further that the step-sizes $\mu_k^{(i)}$ for $i = 1, 2$ are sufficiently small such that condition (55) is satisfied and approximations (64), (71), (112) hold. Furthermore, assume that the step-sizes ν_{γ_k} at the combination layer satisfy condition

$$0 < \nu_{\gamma_k} < \sqrt{\frac{2}{\pi \max_n \{\Delta J_{k,n}^{(1)} + \Delta J_{k,n}^{(2)}\}}}. \quad (83)$$

$$\mathbb{E}\{\gamma_{k,\infty}^2\} = \frac{\nu_{\gamma_k} (J_{\text{ex},k,\infty}^{(2)} + \sigma_{z,k}^2) + 2\mathbb{E}\{\gamma_{k,\infty}\} \left[\Delta J_{k,\infty}^{(2)} \sqrt{\frac{2}{\pi(\Delta J_{k,\infty}^{(1)} + \Delta J_{k,\infty}^{(2)})}} - \nu_{\gamma_k} \Delta J_{k,\infty}^{(2)} \right]}{\sqrt{8(\Delta J_{k,\infty}^{(1)} + \Delta J_{k,\infty}^{(2)})/\pi} - \nu_{\gamma_k}(\Delta J_{k,\infty}^{(1)} + \Delta J_{k,\infty}^{(2)})} \quad (84)$$

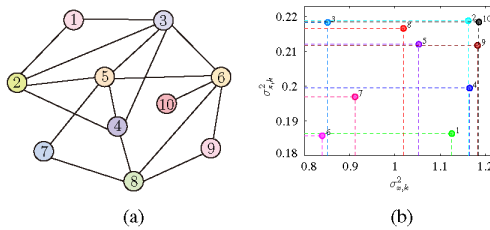


Fig. 2. Network topology and associated input variances and noise variances. (a) Network topology; (b) Agent input and noise variances.

to ensure the mean-square stability of the sign-regressor diffusion scheme. Then for any initial conditions, the distributed network with doubly stochastic matrices $\mathbf{A}_1^{(i)}, \mathbf{A}_2^{(i)}$ and scheme (22) is mean-square stable if $\rho(\mathbf{I}_{NL} - \mathbf{U}^{(i)} \overline{\mathbf{H}}^{(i)}) < 1$, which is further guaranteed by sufficiently small step-sizes that also satisfy condition (55).

C. Mean and mean-square behaviors of $\gamma_{k,n}$

Starting from (22), and following the derivation in Appendix I, we arrive at the following stability analysis result.

Stability Analysis Result 7: (Stability in the Mean) Assume data model (1) and **A1**, **Ap₅**, **Ap₇** hold. Then for any initial conditions, the sign-regressor scheme (22) asymptotically converges in the mean if the step-sizes ν_{γ_k} are chosen to satisfy condition (82). Besides, the mean behavior of $\gamma_{k,n}$ is evaluated by (150) in Appendix I, with steady-state value $\mathbb{E}\{\gamma_{k,\infty}\}$ given by equation (95).

Proof: See Appendix I. ■

Stability Analysis Result 8: (Mean-square Stability) Assume data model (1) and **A1**, **Ap₅**, **Ap₇** hold. Then for any initial conditions, the sign-regressor scheme (22) is mean-square stable if the step-sizes ν_{γ_k} are chosen to satisfy condition (83). Besides, the transient mean-square behavior of $\gamma_{k,n}$ is evaluated by (154) in Appendix I, with steady-state value $\mathbb{E}\{\gamma_{k,\infty}^2\}$ given by equation (84).

Proof: See Appendix I. ■

VI. SIMULATION RESULTS

In this section, we present simulation results to illustrate the proposed combination schemes and theoretical results. All simulated curves were averaged over 100 Monte Carlo runs.

A. Affine combination schemes validation

Consider a non-stationary system identification scenario with \mathbf{w}_k^* varying over time. The distributed network consisted of $N = 10$ nodes with connection topology depicted in Fig. 2(a). The regressors were generated from a multivariate Gaussian distribution with zero-mean and covariance matrix $\mathbf{R}_{x,k} = \sigma_{x,k}^2 \mathbf{I}_{50}$. The noise signals were generated from Gaussian distribution $\mathcal{N}(0, \sigma_{z,k}^2)$. Variances $\sigma_{x,k}^2$ and $\sigma_{z,k}^2$ at each agent were generated randomly as depicted in Fig. 2(b).

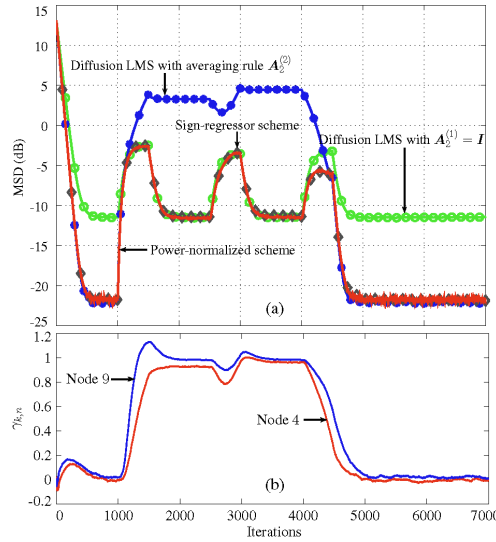


Fig. 3. Simulation results with static fusion matrices. (a) Network MSD learning curves; (b) Evolution of combination coefficient $\gamma_{k,n}$ for power-normalized scheme.

1) *Affine combination of diffusion LMS strategies*: We first considered two ATC diffusion LMS strategies as component strategy. Without loss of generality, matrices \mathbf{C} and $\mathbf{A}_1^{(i)}$ were set to the identity matrix. As a result, vector $\phi_{k,n}$ coincides with $\mathbf{w}_{k,n}$ in equations (4) and (5). We considered two groups of combination matrices $\mathbf{A}_2^{(i)}$: static combination matrices with $\mathbf{A}_2^{(1)} = \mathbf{I}_N$ and averaging rule [5] for $\mathbf{A}_2^{(2)}$, and adaptive combination matrices for $\mathbf{A}_2^{(1)}$, $\mathbf{A}_2^{(2)}$ given in [9] and [13], with (ℓ, k) -th entries at time instant $n + 1$ given by:

$$a_{2,\ell k}^{(1)}(n+1) = \frac{\|\boldsymbol{\psi}_{k,n+1} + \mu_k \mathbf{q}_{k,n} - \boldsymbol{\psi}_{\ell,n+1}\|^{-2}}{\sum_{j \in \mathcal{N}_k} \|\boldsymbol{\psi}_{k,n+1} + \mu_k \mathbf{q}_{k,n} - \boldsymbol{\psi}_{j,n+1}\|^{-2}} \quad (85)$$

$$a_{2,\ell k}^{(2)}(n+1) = \frac{\zeta_{\ell k,n+1}^{-2}}{\sum_{j \in \mathcal{N}_k} \zeta_{jk,n+1}^{-2}} \quad (86)$$

for $\ell \in \mathcal{N}_k$. We use the index $(n + 1)$ in (85) and (86) to highlight the time-variant nature of adaptive combination matrices, and quantities $\mathbf{q}_{k,n}$ and $\zeta_{\ell k,n+1}$ are evaluated as:

$$\mathbf{q}_{k,n} = [d_{k,n} - \mathbf{x}_{k,n}^\top \boldsymbol{\psi}_{k,n+1}] \cdot \mathbf{x}_{k,n} \quad (87)$$

$$\zeta_{\ell k,n+1}^2 = (1 - \tau_k) \cdot \zeta_{\ell k,n}^2 + \tau_k \cdot \|\boldsymbol{\psi}_{\ell,n+1} - \mathbf{w}_{k,n}\|^2 \quad (88)$$

where $0 < \tau_k \ll 1$ are forgetting factors. The evolution of the coefficient vectors \mathbf{w}_k^* was divided into four stationary stages and three transient episodes. For stationary stages, the vectors \mathbf{w}_k^* were generated randomly from a standard Gaussian distribution. During stationary stages, we set \mathbf{w}_k^* at each agent so that, from time instant $n = 1$ to 1000 and $n = 4501$ to 7000, the whole network tracked the same target, while from instant $n = 1501$ to 2500 and $n = 3001$ to 4000, the network started to track 2 and 3 targets, respectively. The transient episodes were designed by using linear interpolation over 500 time instants by $a + \frac{b-a}{500}(n - n_c)$, with a and b denoting an element of the weight of the previous and next stationary stages respectively, and n_c denoting the starting instant of the current transient episode — see [9]. Besides, we set $\varepsilon = 0.05$ and $\eta = 0.95$ for power-normalized diffusion scheme, with $\nu_{\gamma_k} = 0.01$ for static combination matrices and $\nu_{\gamma_k} = 0.04$ for adaptive combination matrices. For sign-regressor diffusion, we set ν_{γ_k} to 0.015 and 0.03 for static and adaptive combination matrices, respectively.

The results are plotted in Figs. 3 and 4 for static and adaptive fusion matrices, respectively. In Fig. 3(a), as expected, the

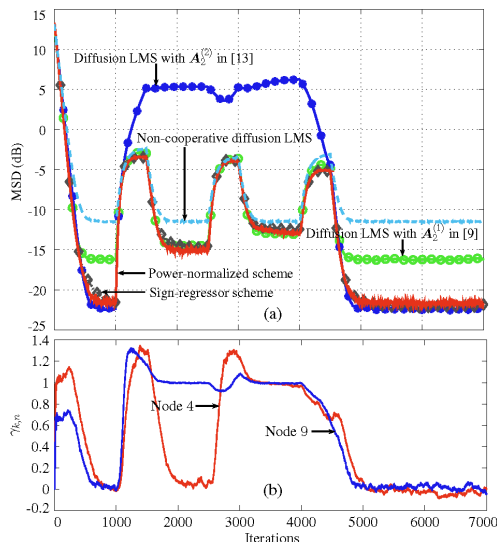


Fig. 4. Simulation results with adaptive fusion matrices.

power-normalized and sign-regressor diffusion schemes led to a MSD learning curve approaching the best of each component strategies at the different stages. This coincides with the theoretical results that these schemes are universal at steady state. The evolutions of affine combination coefficients in Fig. 3(b) ensure the effectiveness of proposed schemes.

The results for adaptive fusion matrices plotted in Fig. 4 lead to similar conclusions as Fig. 3. Interestingly, from time instant $n = 1901$ to 2500, the combined result of the power-normalized diffusion scheme outperformed each individual one, driven by combination coefficients $\gamma_{k,n}$ such as at node 4 in Fig. 4(b). This confirms the theoretical result in (29) that shows that the combined strategy can outperform each component strategy in certain situations. The power-normalized diffusion performs better than the sign-regressor diffusion due to its faster convergence rate. All of the results in Fig. 3 and Fig. 4 illustrate the effectiveness of the proposed schemes.

2) *Affine combination of other strategies*: Consider the diffusion strategies for clustered multi-task networks proposed in [8] and [34]. The former uses squared ℓ_2 -norm co-regularizer to promote cooperation within clusters, while the latter uses an ℓ_1 -norm co-regularizer. The simulation setting is similar to that of Section VI-A1, except that the 10 nodes were grouped into 3 clusters, to track three groups of different but related targets. For stationary stages, the coefficient vectors $w_{C_i}^*$ were generated as $w_{C_i}^* = w_o + \delta_{C_i} w_{C_i}$, with w_{C_i} drawn from standard Gaussian distribution. When δ_{C_i} for $i = 1, 2, 3$ are the same or similar, [34] with ℓ_1 norm co-regularizer works better, otherwise the method in [8] is better. The regularization strength of the two co-regularizers were both set to 0.1, and a uniform $A_2^{(i)}$ was used such that $a_{2,\ell_k}^{(i)} = |\mathcal{N}_k \cap \mathcal{C}(k)|^{-1}$. For the power-normalized diffusion scheme, we set ν_{γ_k} to 0.01, ε to 0.05, and η to 0.95. For sign-regressor diffusion, we set ν_{γ_k} to 0.1.

The results are plotted in Fig. 5 and Fig. 6. Both power-normalized and sign-regressor diffusion not only led to the best of each component strategies at different stages, they also outperformed each component strategies at some instants.

3) *Influence of parameters*: Since there are several parameters in the power-normalized and sign-regressor schemes, such as step-size ν_{γ_k} , temporal smoothing factor η and parameter ε , we examine their influence on the performance. Based on various experiments, we find that the performance of the power-normalized scheme is not sensitive to η and to small-valued ε . We therefore suggest setting η to a typical value of 0.95 and ε to 0.05. We also examine the influence of the step-size ν_{γ_k} . The simulation settings are identical to those used in the first experiment, and we use static combination matrices.

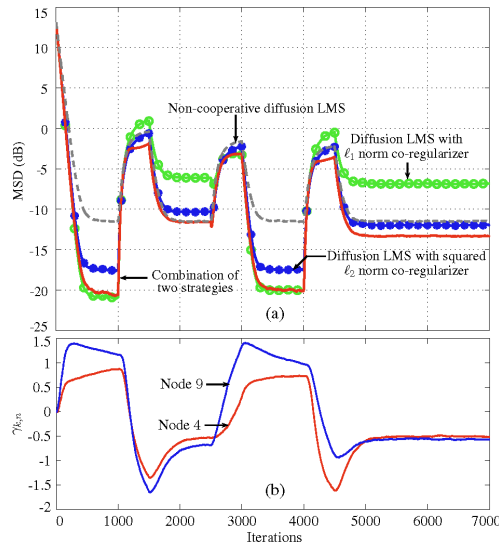


Fig. 5. Simulation results of the power-normalized scheme with the two diffusion strategies in [8] and [34]. (a) Network MSD learning curves; (b) Evolution of the affine combination coefficients $\gamma_{k,n}$.

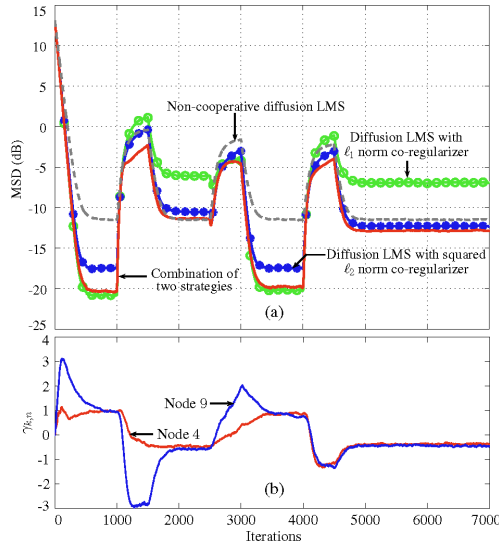


Fig. 6. Simulation results of the sign-regressor scheme with the two diffusion strategies in [8] and [34].

The results are plotted in Figs. 7 and 8. For both power-normalized and sign-regressor diffusion, small ν_{γ_k} lead to weak ability in tracking the best component and slow convergence toward the best component at steady state, such as $\nu_{\gamma_k} = 0.001$, while a large ν_{γ_k} results in biases from the best component at steady state, though a large step-size ensures good tracking for the best component. The step-size parameter ν_{γ_k} needs to be fine-tuned to ensure good tracking ability and a lower bias from the best component at steady state. In this simulation setting, the values of ν_{γ_k} are set to 0.01 and 0.015 for power-normalized and sign-regressor schemes, respectively.

B. Theoretical models

To illustrate the theoretical results as well as challenge the assumptions and approximations adopted in the theoretical analysis, we considered three networks with different connectivity parameters as described in Table I. Net1 consisted of 10 nodes with the network topology given in Fig. 2(a). Net2 was a more complicated network consisting of 20 nodes. Net3

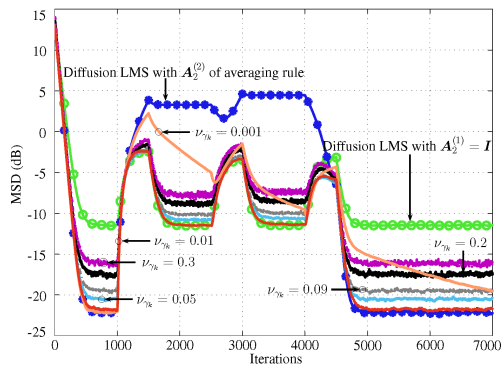


Fig. 7. Simulation results of the power-normalized scheme with different step-sizes ν_{γ_k} .

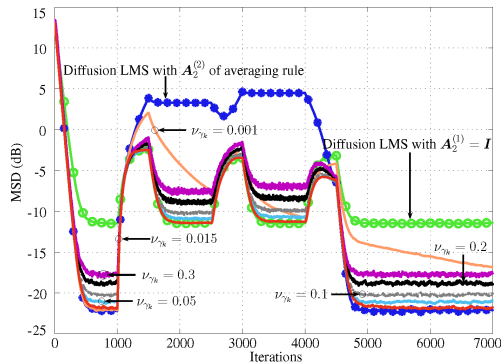


Fig. 8. Simulation results of the sign-regressor scheme with different step-sizes ν_{γ_k} .

was generated by dividing 20 nodes into seven fully connected clusters, with 3 nodes in each of the first six clusters and 2 nodes in the last cluster. These seven clusters were connected in chain, with a single edge connecting adjacent clusters: agent 3 (in cluster 1) was connected with agent 4 (in cluster 2), and agent 6 (in cluster 2) was connected with agent 7 (in cluster 3), and so on until agent 18 (in cluster 6) was connected to agent 19 (in cluster 7).

TABLE I

NETWORK STATISTICS FOR THEORETICAL MODELS VALIDATION. \mathbf{L} IS THE LAPLACIAN MATRIX ASSOCIATED WITH THE GRAPH (NETWORK), $\lambda_2(\mathbf{L})$ IS THE ALGEBRAIC CONNECTIVITY [35] OF GRAPH, SIZE IS THE NUMBER OF NODES, DENSITY IS THE NUMBER OF NON-ZERO ENTRIES OF THE ADJACENCY MATRIX OF GRAPH, AND DIAMETER IS THE MAXIMUM DISTANCE BETWEEN ANY TWO NODES [36].

Network	Size	Density	$\lambda_2(\mathbf{L})$	Diameter
Net1	10	44%	0.7962	3
Net2	20	38%	0.9549	3
Net3	20	17.25%	0.0439	13

The unknown coefficient vectors to be estimated were of length $L = 2$. We first considered regressors drawn from a zero-mean Gaussian distribution with covariance matrix $\mathbf{R}_{x,k} = \sigma_{x,k}^2 \mathbf{I}_L$. Next we considered colored regressors $\mathbf{x}_{k,n} = [x_{k,n} \ x_{k,n-1}]^\top$ generated from a first-order AR model: $x_{k,n} = 0.5x_{k,n-1} + \sqrt{0.75\sigma_{x,k}^2} w_{k,n}$. The input signal $w_{k,n}$ was i.i.d and drawn from a zero-mean Gaussian distribution, with variance $\sigma_w^2 = 1$, so that:

$$\mathbf{R}_{x,k} = \sigma_{x,k}^2 \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}. \quad (89)$$

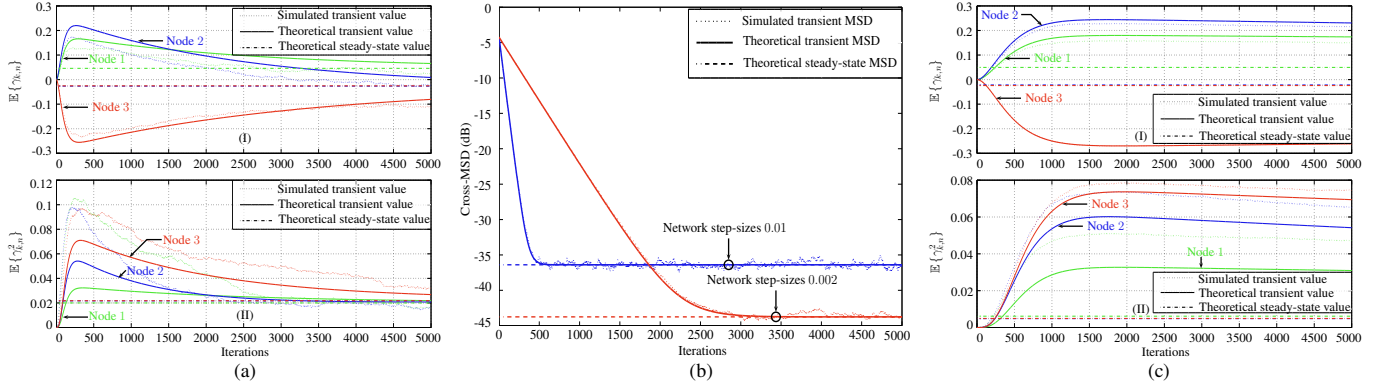


Fig. 9. Illustration of simulation results (model vs. Monte Carlo) for the power-normalized scheme in Net1 and SNR1. Transient and steady-state values of $\mathbb{E}\{\gamma_{k,n}\}$ derived in (136) and (95) (top), as well as these of $\mathbb{E}\{\gamma_{k,n}^2\}$ derived in (141) and (78) (bottom) for network step-size 0.01 (a) and 0.002 (c); (b) Transient and steady-state cross-MSDs derived in (123) and (130).

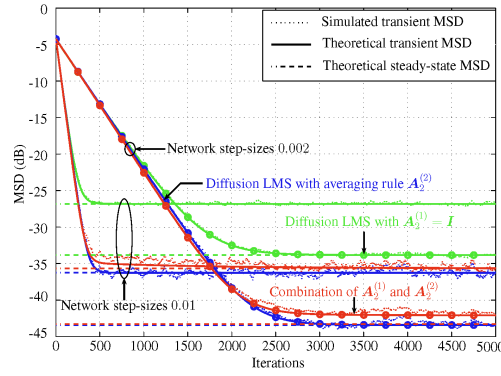


Fig. 10. Network performance with the power-normalized scheme (model vs. Monte Carlo) with two different network step-sizes in Net1 and SNR1. For each step-size, the results plotted with a same color stand for same diffusion strategy. For each component diffusion strategy, theoretical transient MSD and theoretical steady-state MSD are derived in (121) and (127). The combination results are derived in (58) and (132).

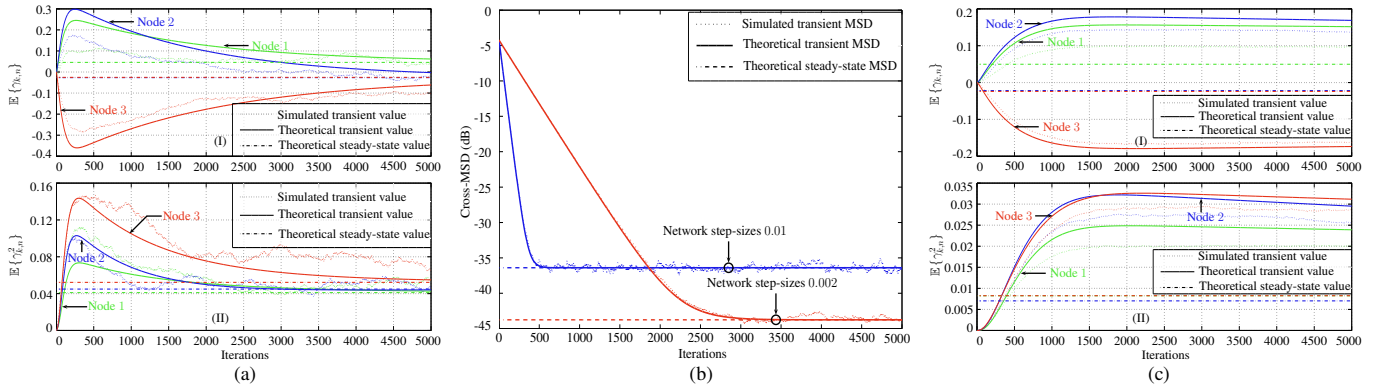


Fig. 11. Illustration of simulation results (model vs. Monte Carlo) of the sign-regressor scheme in Net1 and SNR1. Transient and steady-state values of $\mathbb{E}\{\gamma_{k,n}\}$ derived in (150) and (95) (top), as well as $\mathbb{E}\{\gamma_{k,n}^2\}$ derived in (154) and (78) (bottom) for network step-size 0.01 (a) and 0.002 (c); (b) Transient and steady-state cross-MSDs derived in (123) and (130).

The noise signals were generated from Gaussian distributions $\mathcal{N}(0, \sigma_{z,k}^2)$. Variances $\sigma_{x,k}^2$ and $\sigma_{z,k}^2$ at each agent were generated randomly. By varying $\sigma_{z,k}^2$, we changed the signal-to-noise ratio (SNR) [37] to three levels as described in Table II. Both combination schemes were run with network step-sizes being set to $\{0.01, 0.002\}$, and with ν_{γ_k} correspondingly being

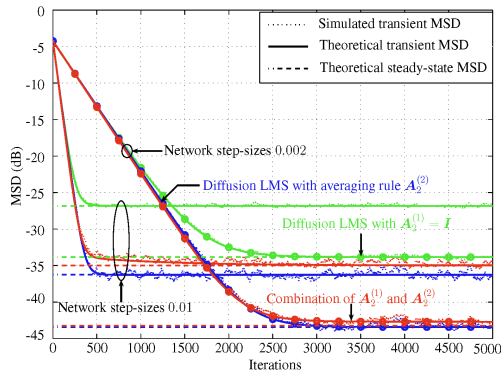


Fig. 12. Network performance with the sign-regressor scheme in Net1 and SNR1.

set to $\{0.01, 0.002\}$ and $\{0.015, 0.001\}$ for power-normalized diffusion and sign-regressor diffusion, respectively. Besides, we set ε to 0.05 and η to 0.95 for power-normalized diffusion. We first validated the theoretical results related to the mean and mean-square behaviors of $\gamma_{k,n}$, the theoretical MSD of each component strategy, as well as cross-MSD of the whole network defined by $\text{MSD}_{\text{cross}} \triangleq \frac{1}{N} \mathbb{E}\{\mathbf{v}_n^{(1)\top} \mathbf{v}_n^{(2)}\}$. Then we evaluated the theoretical MSD behavior of the combined strategy.

TABLE II

THREE SNR LEVELS IN DECIBEL (dB) FOR THEORETICAL MODELS VALIDATION. SINCE SNRS VARY FROM NODE TO NODE, WE ENUMERATE THE MAXIMUM, MINIMUM AND MEAN VALUES.

SNR Level	Maximum	Minimum	Mean
SNR1	3.5724	1.6673	2.7946
SNR2	-9.4379	-11.343	-10.2157
SNR3	-18.9803	-20.8855	-19.7581

The results for white Gaussian inputs with Net1 and SNR1 are plotted in Fig. 9–Fig. 12. In both Fig. 9 and Fig. 11, we observe that the simulated transient values and theoretical transient values accurately matched, especially for small step-sizes, since the approximations adopted in theoretical analyses are more reasonable for small step-sizes. Further, we observe that a larger ν_{γ_k} results in a faster convergence rate, which coincides with equations (138) and (151) in theoretical analyses. In Fig. 10 and Fig. 12, besides the accurate matching of the simulated and theoretical MSD learning curves, the almost superposition of theoretical steady-state MSDs for combined strategy and that of the best component validates the conclusion again that the combination schemes are universal at steady state. Further, the gaps of these two theoretical values becomes smaller as the step-sizes decrease, since the analyses of the universality are based on the approximations that are more valid for small step-sizes.

The results of the power-normalized diffusion for white Gaussian inputs with Net2 and Net3 in SNR1 are plotted in Fig. 13 and Fig. 14, respectively. And the results with Net1 in SNR2 and SNR3 are plotted in Fig. 15 and Fig. 16, respectively. Together with Fig. 10, all these validate the accuracy of theoretical analyses under different SNR conditions and network connectivity parameters.

The results of the power-normalized diffusion for correlated input with Net1 and SNR1 are plotted in Fig. 17. Though assumption **A2** is violated, the superimposition of simulated and theoretical curves validates the accuracy of the theoretical analyses for sufficiently small step-sizes with moderately correlated regressors.

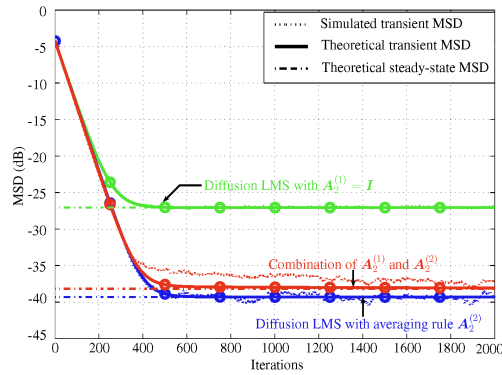


Fig. 13. Network performance with the power-normalized scheme in Net2 and SNR1, and the network step-sizes are 0.01.

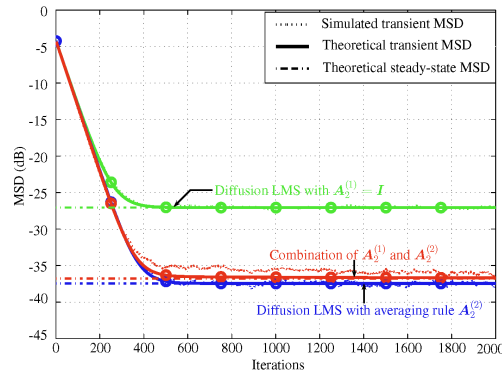


Fig. 14. Network performance with the power-normalized scheme in Net3 and SNR1, and the network step-sizes are 0.01

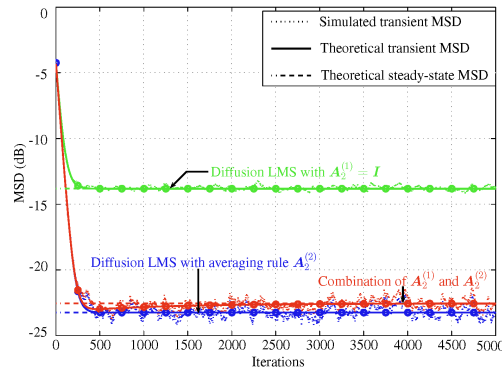


Fig. 15. Network performance with the power-normalized scheme in Net1 and SNR2, and the network step-sizes are 0.01.

VII. CONCLUSION AND PERSPECTIVES

Combining diffusion strategies enables a network to reach better performance. In this paper, we proposed two schemes for affine combination of two diffusion strategies. By using the proposed combination schemes, we obtained a combined diffusion strategy ensuring the advantages of both component strategies simultaneously, sometimes even better, in terms of EMSE. We conducted theoretical analyses in the mean and mean-square sense, and analyzed the universality of each approach. Simulation results illustrate the interesting properties of affine combination schemes, as well as the accuracy of the theoretical results. Several open problems still have to be addressed. For instance, it would be interesting to design combination schemes and conduct theoretical analysis for colored measurement noise and correlated regressors. Some works focusing on adaptive networks

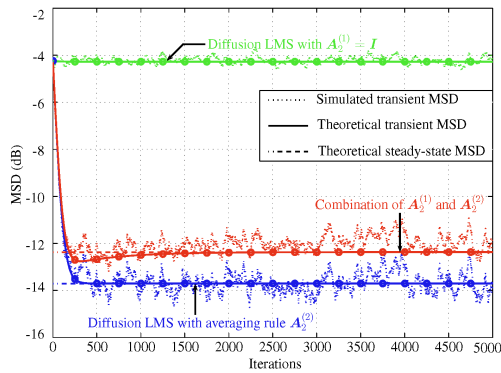


Fig. 16. Network performance with the power-normalized scheme in Net1 and SNR3, and the network step-sizes are 0.01.

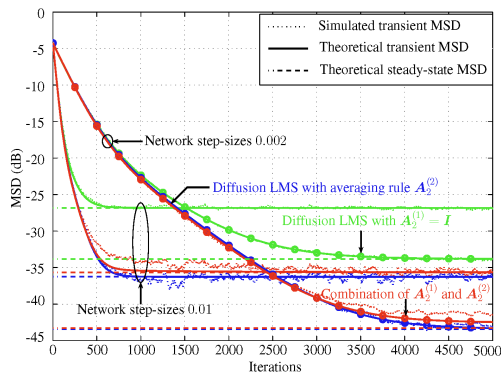


Fig. 17. Network performance with the power-normalized scheme using correlated inputs in Net1 and SNR1.

with colored measurement noise include [38]–[40]. It would also be interesting to explore other combination frameworks and schemes.

APPENDIX A

COMBINATION SCHEME FOR MULTIPLE STRATEGIES

The general scheme follows the description from Section III except that we now have M component diffusion strategies. We introduce M affine combination coefficients $\gamma_{k,n}^{(1)}, \gamma_{k,n}^{(2)}, \dots, \gamma_{k,n}^{(M)}$ for node k at time instant n , satisfying the constraint $\sum_{i=1}^M \gamma_{k,n}^{(i)} = 1$. By combining the estimates of M component strategies at each agent k , we obtain the overall system coefficients $w_{k,n}$ and estimation error $e_{k,n}$ at combination layer, with $e_{k,n} = \sum_{i=1}^M \gamma_{k,n}^{(i)} e_{k,n}^{(i)}$. In order to keep $\gamma_{k,n}^{(i)}$ satisfying the sum-to-one constraint, we calculate $\gamma_{k,n}^{(i)}$ via the mapping:

$$\gamma_{k,n}^{(i)} = \frac{\alpha_{k,n}^{(i)} + \delta}{\sum_{j=1}^M \alpha_{k,n}^{(j)} + M\delta}, \quad (90)$$

where $\alpha_{k,n}^{(i)}$ are newly introduced auxiliary parameters, and δ is a small positive constant to avoid zero-division. We shall update $\alpha_{k,n}^{(i)}$ instead of updating $\gamma_{k,n}^{(i)}$ directly. By minimizing the MSE (16) over the entire network, and using stochastic gradient approximation, we obtain:

$$\alpha_{k,n+1}^{(i)} \approx \alpha_{k,n}^{(i)} + \nu_{\alpha_k} e_{k,n} \text{sgn} \left\{ \frac{e_{k,n} - e_{k,n}^{(i)}}{\sum_{j=1}^M \alpha_{k,n}^{(j)} + M\delta} \right\}. \quad (91)$$

APPENDIX B

PROOF OF UNIVERSALITY ANALYSIS RESULT 1

Starting from the update equation of $\gamma_{k,n}$ in (20), we know that the stationary point of $\gamma_{k,n}$ is reached for:

$$\mathbb{E}\left\{\frac{\nu\gamma_k}{\varepsilon + p_{k,n}} e_{k,n} \mathbf{x}_{k,n}^\top (\mathbf{w}_{k,n}^{(1)} - \mathbf{w}_{k,n}^{(2)})\right\} = 0. \quad (92)$$

Using (27) and (28), we obtain:

$$\begin{aligned} & \mathbb{E}\left\{\frac{\nu\gamma_k}{\varepsilon + p_{k,n}} \gamma_{k,n} [\tilde{e}_{k,n}^{(1)} \tilde{e}_{k,n}^{(2)} - (\tilde{e}_{k,n}^{(1)})^2]\right\} + \\ & \frac{\nu\gamma_k}{\varepsilon + p_{k,n}} (1 - \gamma_{k,n}) [(\tilde{e}_{k,n}^{(2)})^2 - \tilde{e}_{k,n}^{(1)} \tilde{e}_{k,n}^{(2)}] \Big\} = 0, \end{aligned} \quad (93)$$

where we used the zero-mean property of $z_{k,n}$ under assumption **A1**. Though we cannot obtain a closed-form solution for $\mathbb{E}\{\gamma_{k,n}\}$ at each time instant, by resorting to approximations **Ap₁**, **Ap₂** and introducing the quantity $\Delta J_{k,n}^{(i)} \triangleq J_{\text{ex},k,n}^{(i)} - J_{\text{ex},k,n}^{(1,2)}$, with $J_{\text{ex},k,n}^{(1,2)}$ denoting the cross-EMSE defined by $J_{\text{ex},k,n}^{(1,2)} \triangleq \mathbb{E}\{\tilde{e}_{k,n}^{(1)} \tilde{e}_{k,n}^{(2)}\}$, equation (93) at steady-state becomes:

$$-\bar{\gamma}_{k,\infty} \Delta J_{k,\infty}^{(1)} + (1 - \bar{\gamma}_{k,\infty}) \Delta J_{k,\infty}^{(2)} = 0 \quad (94)$$

with $\bar{\gamma}_{k,\infty} \triangleq \mathbb{E}\{\gamma_{k,\infty}\}$. This leads to a closed-form expression for the stationary point at steady-state:

$$\bar{\gamma}_{k,\infty} = \frac{\Delta J_{k,\infty}^{(2)}}{\Delta J_{k,\infty}^{(1)} + \Delta J_{k,\infty}^{(2)}}. \quad (95)$$

We can now establish the universality of the power-normalized scheme (20) at steady-state by using (95). Under approximation **Ap₁**, and assuming further that the variance of $\gamma_{k,\infty}$ is small enough so that the approximation $\mathbb{E}\{\gamma_{k,\infty}^2\} \approx \bar{\gamma}_{k,\infty}^2$ holds, the EMSE at node k after combination at steady-state can be written as:

$$\begin{aligned} J_{\text{ex},k,\infty} & \approx \bar{\gamma}_{k,\infty}^2 J_{\text{ex},k,\infty}^{(1)} + (1 - \bar{\gamma}_{k,\infty})^2 J_{\text{ex},k,\infty}^{(2)} \\ & + 2\bar{\gamma}_{k,\infty}(1 - \bar{\gamma}_{k,\infty}) J_{\text{ex},k,\infty}^{(1,2)}. \end{aligned} \quad (96)$$

By substituting $\bar{\gamma}_{k,\infty}$ in (95) into (96), and after some algebraic manipulations, we arrive at:

$$J_{\text{ex},k,\infty} = J_{\text{ex},k,\infty}^{(1,2)} + \frac{\Delta J_{k,\infty}^{(1)} \Delta J_{k,\infty}^{(2)}}{\Delta J_{k,\infty}^{(1)} + \Delta J_{k,\infty}^{(2)}}. \quad (97)$$

From Cauchy-Schwartz inequality, we know that $J_{\text{ex},k,\infty}^{(1,2)}$ cannot be simultaneously larger than $J_{\text{ex},k,\infty}^{(1)}$ and $J_{\text{ex},k,\infty}^{(2)}$. Therefore, to evaluate (97), we split the problem into three cases according to the relations between $J_{\text{ex},k,\infty}^{(1,2)}$ and $J_{\text{ex},k,\infty}^{(i)}$:

- Case 1: $J_{\text{ex},k,\infty}^{(i)} \geq J_{\text{ex},k,\infty}^{(1,2)}$ for $i = 1, 2$, which means $\Delta J_{k,\infty}^{(1)} \geq 0$ and $\Delta J_{k,\infty}^{(2)} \geq 0$ at the same time. Since $\frac{\Delta J_{k,\infty}^{(1)} \Delta J_{k,\infty}^{(2)}}{\Delta J_{k,\infty}^{(1)} + \Delta J_{k,\infty}^{(2)}} \leq \Delta J_{k,\infty}^{(1)}$ and $\frac{\Delta J_{k,\infty}^{(1)} \Delta J_{k,\infty}^{(2)}}{\Delta J_{k,\infty}^{(1)} + \Delta J_{k,\infty}^{(2)}} \leq \Delta J_{k,\infty}^{(2)}$, we conclude from (97) that:

$$J_{\text{ex},k,\infty} \leq J_{\text{ex},k,\infty}^{(1)} \quad \text{and} \quad J_{\text{ex},k,\infty} \leq J_{\text{ex},k,\infty}^{(2)}. \quad (98)$$

- Case 2: $J_{\text{ex},k,\infty}^{(1)} < J_{\text{ex},k,\infty}^{(1,2)} < J_{\text{ex},k,\infty}^{(2)}$, which corresponds to $\Delta J_{k,\infty}^{(1)} < 0$ and $\Delta J_{k,\infty}^{(2)} > 0$. Then (97) leads to:

$$J_{\text{ex},k,\infty} = J_{\text{ex},k,\infty}^{(1)} - \frac{(\Delta J_{k,\infty}^{(1)})^2}{\Delta J_{k,\infty}^{(1)} + \Delta J_{k,\infty}^{(2)}}. \quad (99)$$

Since $\Delta J_{k,\infty}^{(1)} + \Delta J_{k,\infty}^{(2)} > 0$, we have $J_{\text{ex},k,\infty} < J_{\text{ex},k,\infty}^{(1)}$. As a result, $J_{\text{ex},k,\infty} < \min\{J_{\text{ex},k,\infty}^{(1)}, J_{\text{ex},k,\infty}^{(2)}\}$.

- Case 3: $J_{\text{ex},k,\infty}^{(1)} > J_{\text{ex},k,\infty}^{(1,2)} > J_{\text{ex},k,\infty}^{(2)}$. Due to the symmetry of (97) for $\Delta J_{k,\infty}^{(1)}$ and $\Delta J_{k,\infty}^{(2)}$, we obtain the same conclusion as Case 2.

Therefore, at steady-state when the combination coefficient (95) is applied, the resulting EMSE after combination cannot be worse than that of the best component, that is,

$$J_{\text{ex},k,\infty} \leq \min \{J_{\text{ex},k,\infty}^{(1)}, J_{\text{ex},k,\infty}^{(2)}\}. \quad (100)$$

We also conclude that equation (29) holds. It means that the EMSE of the diffusion network after combination cannot be worse than that of the best component strategies, leading to the universality property of the power-normalized scheme at steady state.

APPENDIX C

PROOF OF STABILITY ANALYSIS RESULT 1

$\mathbb{E}\{\mathbf{v}_{n+1}\}$ converges as $n \rightarrow \infty$ if, and only if, both terms on the right-hand side (RHS) of (40) converge to finite values. Using (42) and iterating the first term from $\mathbb{E}\{\mathbf{v}_0^{(1)}\}$, we obtain:

$$\begin{aligned} \mathbb{E}\{\mathbf{\Gamma}_{n+1}\} \mathbb{E}\{\mathbf{v}_{n+1}^{(1)}\} &= \mathbb{E}\{\mathbf{\Gamma}_{n+1}\} (\overline{\mathbf{B}}^{(1)})^{n+1} \mathbb{E}\{\mathbf{v}_0^{(1)}\} \\ &\quad - \sum_{j=0}^n \mathbb{E}\{\mathbf{\Gamma}_{n+1}\} (\overline{\mathbf{B}}^{(1)})^j \overline{\mathbf{r}}^{(1)}. \end{aligned} \quad (101)$$

To prove convergence of (101), it is sufficient to prove that the sequence $s_{1,m} \triangleq \left[\mathbb{E}\{\mathbf{\Gamma}_{n+1}\} (\overline{\mathbf{B}}^{(1)})^{n+1} \mathbb{E}\{\mathbf{v}_0^{(1)}\} \right]_m$ and series $\sum_{j=0}^n \left[\mathbb{E}\{\mathbf{\Gamma}_{n+1}\} (\overline{\mathbf{B}}^{(1)})^j \overline{\mathbf{r}}^{(1)} \right]_m$ converge for $m = 1, \dots, NL$. It is known that a sequence is convergent if it is upper bounded and lower bounded by two sequences with the same limit [41]. A series is absolutely convergent if each of its terms is bounded by a term of an absolutely convergent series [42], [43]. Define $s_{2,m} \triangleq \left[\mathbb{E}\{\mathbf{\Gamma}_{n+1}\} (\overline{\mathbf{B}}^{(1)})^j \overline{\mathbf{r}}^{(1)} \right]_m$. Since the largest absolute value of the entries in a block vector is smaller than or equal to the block maximum norm of this vector, we have:

$$\begin{aligned} |s_{1,m}| &\leq \|\mathbb{E}\{\mathbf{\Gamma}_{n+1}\}\|_{b,\infty} \cdot \|\overline{\mathbf{B}}^{(1)}\|_{b,\infty}^{n+1} \cdot \|\mathbb{E}\{\mathbf{v}_0^{(1)}\}\|_{b,\infty} \\ &\stackrel{(44)}{\leq} \|\mathbb{E}\{\mathbf{\Gamma}_{n+1}\}\|_{b,\infty} \cdot \|\mathbf{M}^{(1)}\|_{b,\infty}^{n+1} \cdot \|\mathbb{E}\{\mathbf{v}_0^{(1)}\}\|_{b,\infty} \\ &= \rho(\mathbb{E}\{\mathbf{\Gamma}_{n+1}\}) \cdot [\rho(\mathbf{M}^{(1)})]^{n+1} \|\mathbb{E}\{\mathbf{v}_0^{(1)}\}\|_{b,\infty}, \end{aligned} \quad (102)$$

$$\begin{aligned} |s_{2,m}| &\leq \|\mathbb{E}\{\mathbf{\Gamma}_{n+1}\}\|_{b,\infty} \cdot \|\overline{\mathbf{B}}^{(1)}\|_{b,\infty}^j \cdot \|\overline{\mathbf{r}}^{(1)}\|_{b,\infty} \\ &\stackrel{(44)}{\leq} \rho(\mathbb{E}\{\mathbf{\Gamma}_{n+1}\}) \cdot [\rho(\mathbf{M}^{(1)})]^j \cdot \|\overline{\mathbf{r}}^{(1)}\|_{b,\infty}, \end{aligned} \quad (103)$$

where

$$\mathbf{M}^{(1)} \triangleq \mathbf{I}_{NL} - \mathbf{U}^{(1)} \overline{\mathbf{H}}^{(1)}, \quad (104)$$

$\|\cdot\|_{b,\infty}$ is the block maximum norm [5], and $\rho(\cdot)$ is the spectral radius of a given matrix. Since $\mathbb{E}\{\mathbf{v}_0^{(1)}\}$ and $\overline{\mathbf{r}}^{(1)}$ are bounded, it is sufficient to require that $\rho(\mathbf{M}^{(1)}) < 1$ and $\rho(\mathbb{E}\{\mathbf{\Gamma}_{n+1}\})$ is uniformly bounded. It is similar for the second term on the RHS of (40). As discussed later in Section IV-C, condition (56) ensures the convergence of $\mathbb{E}\{\mathbf{\Gamma}_{n+1}\}$ and $\mathbb{E}\{\mathbf{I}_{NL} - \mathbf{\Gamma}_{n+1}\}$. Given that ν_{γ_k} satisfies condition (56), the convergence of (40) requires only $\rho(\mathbf{I}_{NL} - \mathbf{U}^{(i)} \overline{\mathbf{H}}^{(i)}) < 1$, which is ensured by step-sizes satisfying (55).

For the asymptotic bias, from (40) and (42), we obtain:

$$\mathbb{E}\{\mathbf{v}_\infty\} = \mathbb{E}\{\mathbf{\Gamma}_\infty\} \mathbb{E}\{\mathbf{v}_\infty^{(1)}\} + (\mathbf{I}_{NL} - \mathbb{E}\{\mathbf{\Gamma}_\infty\}) \mathbb{E}\{\mathbf{v}_\infty^{(2)}\} \quad (105)$$

$$\mathbb{E}\{\mathbf{v}_\infty^{(i)}\} = -(\mathbf{I}_{NL} - \overline{\mathbf{B}}^{(i)})^{-1} \overline{\mathbf{r}}^{(i)}, \quad (106)$$

which leads to expression (57).

APPENDIX D

DERIVATION OF EQUATION (62)

Using (41), (63), (65), and approximation **Ap**₄, we have:

$$\begin{aligned} \mathbb{E}\left\{\|\mathbf{v}_{n+1}^{(i)}\|_{\Sigma_{n+1}^{(i)}}^2\right\} &= \mathbb{E}\left\{\|\mathbf{v}_n^{(i)}\|_{\Sigma_{B,n+1}^{(i)}}^2\right\} + \text{trace}\left\{\Sigma_{n+1}^{(i)} \mathbf{G}^{(i)}\right\} \\ &\quad + \mathbb{E}\left\{\mathbf{f}(\mathbf{r}_n^{(i)}, \Sigma_{n+1}^{(i)}, \mathbf{v}_n^{(i)})\right\} \end{aligned} \quad (107)$$

with

$$\Sigma_{B,n+1}^{(i)} \triangleq \mathbb{E}\left\{\mathbf{B}_n^{(i)\top} \Sigma_{n+1}^{(i)} \mathbf{B}_n^{(i)}\right\}. \quad (108)$$

In the derivation of expression (107), we used relation $\mathbb{E}\{\mathbf{g}_n^{(i)}\} = 0$, which is derived from (50) and (51) directly by using the zero-mean property of $z_{\ell,n}$.

Let $\boldsymbol{\sigma}_{B,n+1}^{(i)} = \text{vec}\{\Sigma_{B,n+1}^{(i)}\}$. Under approximation **Ap**₄, and using the properties of $\text{vec}\{\cdot\}$ operator, we have:

$$\boldsymbol{\sigma}_{B,n+1}^{(i)} = \mathbf{K}^{(i)} \boldsymbol{\sigma}_{n+1}^{(i)}, \quad (109)$$

where

$$\mathbf{K}^{(i)} \triangleq \mathbb{E}\left\{\mathbf{B}_n^{(i)\top} \otimes \mathbf{B}_n^{(i)\top}\right\} \approx \overline{\mathbf{B}}^{(i)\top} \otimes \overline{\mathbf{B}}^{(i)\top} \quad (110)$$

for sufficiently small step-sizes. We have:

$$\text{trace}\left\{\Sigma_{n+1}^{(i)} \mathbf{G}^{(i)}\right\} = \left[\text{vec}\left\{\mathbf{G}^{(i)\top}\right\}\right]^\top \boldsymbol{\sigma}_{n+1}^{(i)}. \quad (111)$$

For sufficiently small step-sizes, we have:

$$\mathbb{E}\left\{\mathbf{f}(\mathbf{r}_n^{(i)}, \Sigma_{n+1}^{(i)}, \mathbf{v}_n^{(i)})\right\} \approx \mathbf{f}(\overline{\mathbf{r}}^{(i)}, \Sigma_{n+1}^{(i)}, \mathbb{E}\{\mathbf{v}_n^{(i)}\}). \quad (112)$$

Substituting (111) and (112) into (107) yields expression (62).

APPENDIX E

PROOF OF STABILITY ANALYSIS RESULT 2

The convergence of $\mathbb{E}\{\|\mathbf{v}_{n+1}\|_{\Sigma}^2\}$ requires the convergence of the terms on the RHS of (58). For the last term, using (62) and iterating from $n = 0$, we find that

$$\begin{aligned} \mathbb{E}\left\{\|\mathbf{v}_{n+1}^{(1)}\|_{\sigma_{n+1}^{(1)}}^2\right\} &= \|\mathbf{v}_0^{(1)}\|_{(\mathbf{K}^{(1)})^{n+1} \sigma_{n+1}^{(1)}}^2 \\ &\quad + \left[\text{vec}\left\{\mathbf{G}^{(1)\top}\right\}\right]^\top \sum_{t=0}^n (\mathbf{K}^{(1)})^t \boldsymbol{\sigma}_{n+1}^{(1)} \\ &\quad + \sum_{t=0}^n \mathbf{f}(\overline{\mathbf{r}}^{(1)}, (\mathbf{K}^{(1)})^t \sigma_{n+1}^{(1)}, \mathbb{E}\{\mathbf{v}_{n-t}^{(1)}\}) \end{aligned} \quad (113)$$

for $n \geq 0$ with initial condition $\mathbf{v}_0^{(1)} = \mathbf{w}_0^{(1)} - \mathbf{w}^*$. Since

$$\boldsymbol{\sigma}_{n+1}^{(1)} = \mathbb{E}\left\{\boldsymbol{\Gamma}_{n+1}^\top \otimes \boldsymbol{\Gamma}_{n+1}^\top\right\} \boldsymbol{\sigma} \quad (114)$$

where $\boldsymbol{\sigma} \triangleq \text{vec}\{\boldsymbol{\Sigma}\}$, the convergence of (113) requires that the sequence $s_{3,m} \triangleq [(\mathbf{K}^{(1)})^{n+1} \mathbb{E}\{\boldsymbol{\Gamma}_{n+1}^\top \otimes \boldsymbol{\Gamma}_{n+1}^\top\} \boldsymbol{\sigma}]_m$ and series $\sum_{t=0}^n [(\mathbf{K}^{(1)})^t \mathbb{E}\{\boldsymbol{\Gamma}_{n+1}^\top \otimes \boldsymbol{\Gamma}_{n+1}^\top\} \boldsymbol{\sigma}]_m$ be convergent for $m = 1, \dots, (NL)^2$. Define $s_{4,m} \triangleq [(\mathbf{K}^{(1)})^t \mathbb{E}\{\boldsymbol{\Gamma}_{n+1}^\top \otimes \boldsymbol{\Gamma}_{n+1}^\top\} \boldsymbol{\sigma}]_m$. Similar to (102) and (103), we have:

$$\begin{aligned} |s_{3,m}| &\leq \|\mathbf{K}^{(1)}\|_{b,\infty}^{n+1} \cdot \|\mathbb{E}\{\boldsymbol{\Gamma}_{n+1}^\top \otimes \boldsymbol{\Gamma}_{n+1}^\top\}\|_{b,\infty} \|\boldsymbol{\sigma}\|_{b,\infty} \\ &= \|\mathbf{K}^{(1)}\|_{b,\infty}^{n+1} \cdot \rho(\mathbb{E}\{\boldsymbol{\Gamma}_{n+1}^\top \otimes \boldsymbol{\Gamma}_{n+1}^\top\}) \|\boldsymbol{\sigma}\|_{b,\infty}, \end{aligned} \quad (115)$$

$$\begin{aligned} |s_{4,m}| &\leq \|\mathbf{K}^{(1)}\|_{b,\infty}^t \cdot \|\mathbb{E}\{\boldsymbol{\Gamma}_{n+1}^\top \otimes \boldsymbol{\Gamma}_{n+1}^\top\}\|_{b,\infty} \|\boldsymbol{\sigma}\|_{b,\infty} \\ &= \|\mathbf{K}^{(1)}\|_{b,\infty}^t \cdot \rho(\mathbb{E}\{\boldsymbol{\Gamma}_{n+1}^\top \otimes \boldsymbol{\Gamma}_{n+1}^\top\}) \cdot \|\boldsymbol{\sigma}\|_{b,\infty}. \end{aligned} \quad (116)$$

According to assumption **A2**, we have:

$$\mathbb{E}\{\gamma_{k,n+1} \gamma_{\ell,n+1}\} = 0, \quad \forall k \neq \ell. \quad (117)$$

Thus, $\rho(\mathbb{E}\{\boldsymbol{\Gamma}_{n+1}^\top \otimes \boldsymbol{\Gamma}_{n+1}^\top\})$ is bounded as long as $\mathbb{E}\{\boldsymbol{\Gamma}_{n+1} \boldsymbol{\Gamma}_{n+1}^\top\}$ is convergent. We further have:

$$\begin{aligned} \|\mathbf{K}^{(1)}\|_{b,\infty} &\stackrel{(64)}{=} \|\overline{\mathbf{B}}^{(1)\top} \otimes \overline{\mathbf{B}}^{(1)\top}\|_{b,\infty} \\ &\stackrel{(44)}{\leq} \|\mathbf{A}_1^{(1)} \otimes \mathbf{A}_1^{(1)}\|_{b,\infty} \|\mathbf{M}^{(1)\top} \otimes \mathbf{M}^{(1)\top}\|_{b,\infty} \\ &\quad \|\mathbf{A}_2^{(1)} \otimes \mathbf{A}_2^{(1)}\|_{b,\infty}. \end{aligned} \quad (118)$$

Since $\|\boldsymbol{\sigma}\|_{b,\infty}$ is bounded,

$$\begin{aligned} \|\mathbf{M}^{(1)\top} \otimes \mathbf{M}^{(1)\top}\|_{b,\infty} &\stackrel{(104)}{=} \rho(\mathbf{M}^{(1)\top} \otimes \mathbf{M}^{(1)\top}) \\ &= [\rho(\mathbf{M}^{(1)})]^2 \end{aligned} \quad (119)$$

and

$$\begin{aligned} \|\mathbf{A}_j^{(1)} \otimes \mathbf{A}_j^{(1)}\|_{b,\infty} &= \|\mathbf{A}_j^{(1)} \otimes \mathbf{A}_j^{(1)}\|_\infty \\ &= \max_{\ell,p} \sum_{k=1}^N a_{j,\ell k}^{(1)} \sum_{q=1}^N a_{j,pq}^{(1)} = 1, \end{aligned} \quad (120)$$

where the notation $\|\cdot\|_\infty$ denotes the maximum absolute row sum of its argument, the convergence of (115) and (116) requires that $\rho(\mathbf{M}^{(1)}) < 1$ and the convergence of $\mathbb{E}\{\boldsymbol{\Gamma}_{n+1} \boldsymbol{\Gamma}_{n+1}^\top\}$. It is similar for the first two terms on the RHS of (58). As derived later in Section IV-C, condition (73) guarantees the stabilities of $\mathbb{E}\{\boldsymbol{\Gamma}_{n+1}\}$ and $\mathbb{E}\{\boldsymbol{\Gamma}_{n+1} \boldsymbol{\Gamma}_{n+1}^\top\}$. Then, for any given weighting matrix $\boldsymbol{\Sigma}$, condition (55) ensures the mean-square stability of the power-normalized diffusion scheme.

APPENDIX F

RECURSIONS FOR EVALUATING TRANSIENT MSD

Recursion for evaluating $\mathbb{E}\{\|\mathbf{v}_n^{(i)}\|_{\mathbf{K}^{(i)} \boldsymbol{\sigma}_{n+1}^{(i)}}^2\}$ is given in [9] as:

$$\begin{aligned} \xi_{n+1}^{(i)} &= \xi_n^{(i)} + [(\text{vec}\{\mathbf{G}^{(i)\top}\})^\top (\mathbf{K}^{(i)})^n \boldsymbol{\sigma}^{(i)} \\ &\quad + \|\overline{\mathbf{r}}^{(i)}\|_{(\mathbf{K}^{(i)})^n \boldsymbol{\sigma}^{(i)}}^2 - \|\mathbf{v}_0^{(i)}\|_{(\mathbf{I} - \mathbf{K}^{(i)}) (\mathbf{K}^{(i)})^n \boldsymbol{\sigma}^{(i)}}^2 \\ &\quad - 2(\boldsymbol{\Lambda}_n^{(i)} + (\overline{\mathbf{B}}^{(i)} \mathbb{E}\{\mathbf{v}_n^{(i)}\})^\top \otimes \overline{\mathbf{r}}^{(i)\top}) \boldsymbol{\sigma}^{(i)}], \end{aligned} \quad (121)$$

and

$$\boldsymbol{\Lambda}_{n+1}^{(i)} = \boldsymbol{\Lambda}_n^{(i)} \mathbf{K}^{(i)} + ((\overline{\mathbf{B}}^{(i)} \mathbb{E}\{\mathbf{v}_n^{(i)}\})^\top \otimes \overline{\mathbf{r}}^{(i)\top}) (\mathbf{K}^{(i)} - \mathbf{I}) \quad (122)$$

with $\mathbf{\Lambda}_0^{(i)} = \mathbf{0}_{1 \times (NL)^2}$, $\xi_{n+1}^{(i)} = \mathbb{E}\{\|\mathbf{v}_{n+1}^{(i)}\|_{\mathbf{K}^{(i)}\boldsymbol{\sigma}_{n+1}^{(i)}}^2\}$, $\xi_0^{(i)} = \|\mathbf{v}_0^{(i)}\|_{\mathbf{K}^{(i)}\boldsymbol{\sigma}_{n+1}^{(i)}}^2$, $\boldsymbol{\sigma}^{(i)} = \mathbf{K}^{(i)}\boldsymbol{\sigma}_{n+1}^{(i)}$ for $i = 1, 2$. Besides, [44] gives an alternative expression for (121), which greatly saves computations. Following the same routine, $\mathbb{E}\{\mathbf{v}_n^{(1)\top} \boldsymbol{\Sigma}_{\text{xc},n+1} \mathbf{v}_n^{(2)}\}$ can be evaluated as follows:

$$\begin{aligned} \xi_{x,n+1} &= \xi_{x,n} + (\text{vec}\{\mathbf{G}_x^\top\})^\top (\mathbf{K}_x)^n \boldsymbol{\sigma}_x + (\boldsymbol{\Pi}_n^{(1)} + \boldsymbol{\Pi}_n^{(2)}) \boldsymbol{\sigma}_x \\ &\quad - [(\overline{\mathbf{B}}^{(2)} \mathbb{E}\{\mathbf{v}_n^{(2)}\})^\top \otimes \bar{\mathbf{r}}^{(1)\top} + \bar{\mathbf{r}}^{(2)\top} \otimes (\overline{\mathbf{B}}^{(1)} \mathbb{E}\{\mathbf{v}_n^{(1)}\})^\top] \boldsymbol{\sigma}_x \\ &\quad - \mathbf{v}_0^{(1)\top} \text{vec}^{-1}\{(\mathbf{I} - \mathbf{K}_x)((\mathbf{K}_x)^n \boldsymbol{\sigma}_x)\} \mathbf{v}_0^{(2)} \\ &\quad + \bar{\mathbf{r}}^{(1)\top} \text{vec}^{-1}\{(\mathbf{K}_x)^n \boldsymbol{\sigma}_x\} \bar{\mathbf{r}}^{(2)}, \end{aligned} \quad (123)$$

and

$$\boldsymbol{\Pi}_{n+1}^{(1)} = \boldsymbol{\Pi}_n^{(1)} \mathbf{K}_x + [\bar{\mathbf{r}}^{(2)\top} \otimes (\overline{\mathbf{B}}^{(1)} \mathbb{E}\{\mathbf{v}_n^{(1)}\})^\top] (\mathbf{I} - \mathbf{K}_x), \quad (124)$$

$$\boldsymbol{\Pi}_{n+1}^{(2)} = \boldsymbol{\Pi}_n^{(2)} \mathbf{K}_x + [(\overline{\mathbf{B}}^{(2)} \mathbb{E}\{\mathbf{v}_n^{(2)}\})^\top \otimes \bar{\mathbf{r}}^{(1)\top}] (\mathbf{I} - \mathbf{K}_x) \quad (125)$$

with $\boldsymbol{\Pi}_0^{(i)} = \mathbf{0}_{1 \times (NL)^2}$, $\boldsymbol{\sigma}_x = \mathbf{K}_x \boldsymbol{\sigma}_{x,n+1}$, $\xi_{x,n+1} = \mathbb{E}\{\mathbf{v}_{n+1}^{(1)\top} \boldsymbol{\Sigma}_{\text{xc},n+1} \mathbf{v}_{n+1}^{(2)}\}$, $\xi_{x,0} = \mathbb{E}\{\mathbf{v}_0^{(1)\top} \boldsymbol{\Sigma}_{\text{xc},n+1} \mathbf{v}_0^{(2)}\}$.

APPENDIX G

PROOF OF REMARK 2

The steady-state MSD is given by the limit $\text{MSD}^{\text{steady}} = \lim_{n \rightarrow \infty} \frac{1}{N} \mathbb{E}\{\|\mathbf{v}_{n+1}\|^2\}$. By resorting to expression (58), we obtain:

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{N} \mathbb{E}\{\|\mathbf{v}_{n+1}\|^2\} &= \frac{1}{N} \mathbb{E}\{\|\mathbf{v}_\infty^{(1)}\|_{\boldsymbol{\Gamma}_\infty}^2\} \\ &\quad + \frac{1}{N} \mathbb{E}\{\|\mathbf{v}_\infty^{(2)}\|_{(\mathbf{I}_{NL} - \boldsymbol{\Gamma}_\infty)^\top (\mathbf{I}_{NL} - \boldsymbol{\Gamma}_\infty)}^2\} \\ &\quad + \frac{2}{N} \mathbb{E}\{\mathbf{v}_\infty^{(1)\top} \boldsymbol{\Gamma}_\infty (\mathbf{I}_{NL} - \boldsymbol{\Gamma}_\infty) \mathbf{v}_\infty^{(2)}\}, \end{aligned} \quad (126)$$

which is the summation of three steady-state values. For the first two terms, recursing (62) with $n \rightarrow \infty$ yields:

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E}\left\{\|\mathbf{v}_{n+1}^{(i)}\|_{(\mathbf{I}_{(NL)^2} - \mathbf{K}^{(i)})\boldsymbol{\sigma}_{n+1}^{(i)}}\right\} &= \\ &= [\text{vec}\{\mathbf{G}^{(i)\top}\}]^\top \boldsymbol{\sigma}_\infty^{(i)} + \mathbf{f}(\bar{\mathbf{r}}^{(i)}, \boldsymbol{\Sigma}_\infty^{(i)}, \mathbb{E}\{\mathbf{v}_\infty^{(i)}\}). \end{aligned} \quad (127)$$

In order to use (127) in (126), we select $\boldsymbol{\sigma}_\infty^{(1)}$ and $\boldsymbol{\sigma}_\infty^{(2)}$ to satisfy:

$$(\mathbf{I}_{(NL)^2} - \mathbf{K}^{(1)})\boldsymbol{\sigma}_\infty^{(1)} = \frac{1}{N} \text{vec}\{\mathbb{E}\{\boldsymbol{\Gamma}_\infty^\top \boldsymbol{\Gamma}_\infty\}\}, \quad (128)$$

$$\begin{aligned} (\mathbf{I}_{(NL)^2} - \mathbf{K}^{(2)})\boldsymbol{\sigma}_\infty^{(2)} &= \\ &= \frac{1}{N} \text{vec}\{\mathbb{E}\{(\mathbf{I}_{NL} - \boldsymbol{\Gamma}_\infty)^\top (\mathbf{I}_{NL} - \boldsymbol{\Gamma}_\infty)\}\}. \end{aligned} \quad (129)$$

Similarly for the last term of (126), recursing (66) with $n \rightarrow \infty$ yields:

$$\begin{aligned} \mathbb{E}\{\mathbf{v}_\infty^{(1)\top} \text{vec}^{-1}\{(\mathbf{I}_{(NL)^2} - \mathbf{K}_x)\boldsymbol{\sigma}_{x,\infty}\} \mathbf{v}_\infty^{(2)}\} &= [\text{vec}\{\mathbf{G}_x^\top\}]^\top \boldsymbol{\sigma}_{x,\infty} \\ &\quad + \mathbf{f}_x(\bar{\mathbf{r}}^{(1)}, \bar{\mathbf{r}}^{(2)}, \boldsymbol{\Sigma}_{x,\infty}, \mathbb{E}\{\mathbf{v}_\infty^{(1)}\}, \mathbb{E}\{\mathbf{v}_\infty^{(2)}\}, \overline{\mathbf{B}}^{(1)}, \overline{\mathbf{B}}^{(2)}), \end{aligned} \quad (130)$$

and we select $\boldsymbol{\sigma}_{x,\infty}$ to satisfy:

$$(\mathbf{I}_{(NL)^2} - \mathbf{K}_x)\boldsymbol{\sigma}_{x,\infty} = \frac{2}{N} \text{vec}\{\mathbb{E}\{\boldsymbol{\Gamma}_\infty (\mathbf{I}_{NL} - \boldsymbol{\Gamma}_\infty)\}\}. \quad (131)$$

These expressions lead to the MSD at steady state as:

$$\begin{aligned} \text{MSD}^{\text{steady}} &= [\text{vec}\{\mathbf{G}^{(1)\top}\}]^\top \boldsymbol{\sigma}_\infty^{(1)} + \mathbf{f}(\bar{\mathbf{r}}^{(1)}, \boldsymbol{\Sigma}_\infty^{(1)}, \mathbb{E}\{\mathbf{v}_\infty^{(1)}\}) + \\ & [\text{vec}\{\mathbf{G}^{(2)\top}\}]^\top \boldsymbol{\sigma}_\infty^{(2)} + \mathbf{f}(\bar{\mathbf{r}}^{(2)}, \boldsymbol{\Sigma}_\infty^{(2)}, \mathbb{E}\{\mathbf{v}_\infty^{(2)}\}) + [\text{vec}\{\mathbf{G}_x^\top\}]^\top \boldsymbol{\sigma}_{x,\infty} \\ & + \mathbf{f}_x(\bar{\mathbf{r}}^{(1)}, \bar{\mathbf{r}}^{(2)}, \boldsymbol{\Sigma}_{x,\infty}, \mathbb{E}\{\mathbf{v}_\infty^{(1)}\}, \mathbb{E}\{\mathbf{v}_\infty^{(2)}\}, \bar{\mathbf{B}}^{(1)}, \bar{\mathbf{B}}^{(2)}) \end{aligned} \quad (132)$$

with

$$\boldsymbol{\Sigma}_\infty^{(i)} = \text{vec}^{-1}\{\boldsymbol{\sigma}_\infty^{(i)}\}, \quad \forall i = 1, 2 \quad (133)$$

$$\boldsymbol{\Sigma}_{x,\infty} = \text{vec}^{-1}\{\boldsymbol{\sigma}_{x,\infty}\}. \quad (134)$$

APPENDIX H

PROOF OF STABILITY ANALYSIS RESULTS 3 & 4

By substituting (27), (28) into (20), and taking expectation, we obtain:

$$\begin{aligned} \mathbb{E}\{\gamma_{k,n+1}\} &= \mathbb{E}\{\gamma_{k,n}\} + \mathbb{E}\left\{\frac{\nu_{\gamma_k}}{\varepsilon + p_{k,n}} \gamma_{k,n} [\tilde{e}_{k,n}^{(1)} \tilde{e}_{k,n}^{(2)} - (\tilde{e}_{k,n}^{(1)})^2] \right. \\ & \left. + \frac{\nu_{\gamma_k}}{\varepsilon + p_{k,n}} (1 - \gamma_{k,n}) [(\tilde{e}_{k,n}^{(2)})^2 - \tilde{e}_{k,n}^{(1)} \tilde{e}_{k,n}^{(2)}] \right\}. \end{aligned} \quad (135)$$

By resorting to EMSEs $J_{\text{ex},k,n}^{(i)}$ and cross-EMSE $J_{\text{ex},k,n}^{(1,2)}$ defined in Section IV-A, as well as **Ap₂**, **Ap₅**, **Ap₆**, expression (135) becomes

$$\begin{aligned} \mathbb{E}\{\gamma_{k,n+1}\} &= \mathbb{E}\{\gamma_{k,n}\} \left[1 - \frac{\nu_{\gamma_k}}{\varepsilon + \bar{p}_{k,n}} (\Delta J_{k,n}^{(1)} + \Delta J_{k,n}^{(2)}) \right] \\ & + \frac{\nu_{\gamma_k}}{\varepsilon + \bar{p}_{k,n}} (J_{\text{ex},k,n}^{(2)} - J_{\text{ex},k,n}^{(1,2)}), \end{aligned} \quad (136)$$

where we adopt the following approximation to simplify the derivation:

$$\begin{aligned} \mathbb{E}\left\{\frac{\nu_{\gamma_k}}{\varepsilon + p_{k,n}}\right\} &\approx \frac{\nu_{\gamma_k}}{\varepsilon + \bar{p}_{k,n}} \quad \text{with} \\ \bar{p}_{k,n} &= \eta \bar{p}_{k,n-1} + (1 - \eta)(\Delta J_{k,n}^{(1)} + \Delta J_{k,n}^{(2)}). \end{aligned} \quad (137)$$

From (136) a sufficient condition can be derived for the step-size ν_{γ_k} to ensure the mean stability of the power-normalized scheme [24], [26]:

$$\left| 1 - \frac{\nu_{\gamma_k}}{\varepsilon + \bar{p}_{k,n}} (\Delta J_{k,n}^{(1)} + \Delta J_{k,n}^{(2)}) \right| < 1 - \phi, \quad \forall n, \quad (138)$$

with a positive constant ϕ , i.e., the left term is uniformly bounded away from one. A sufficient condition to ensure (138) is given by:

$$0 < \nu_{\gamma_k} < \min_n \left\{ \frac{\varepsilon + \bar{p}_{k,n}}{\Delta J_{k,n}^{(1)} + \Delta J_{k,n}^{(2)}} \right\}. \quad (139)$$

Substituting $\bar{p}_{k,n}$ of (137) into (139), and after some mathematical manipulations, we obtain condition (56). By further taking the limit of (136) as $n \rightarrow \infty$, and solving for $\mathbb{E}\{\gamma_{k,\infty}\}$, we arrive at equation (95).

$$\left| 1 + 3 \left(\frac{\nu_{\gamma_k}}{\varepsilon + \bar{p}_{k,n}} \right)^2 \left(\Delta J_{k,n}^{(1)} + \Delta J_{k,n}^{(2)} \right)^2 - \frac{2\nu_{\gamma_k}}{\varepsilon + \bar{p}_{k,n}} \left(\Delta J_{k,n}^{(1)} + \Delta J_{k,n}^{(2)} \right) \right| < 1 - \phi, \quad \forall n \quad (146)$$

Next, we evaluate the mean-square behavior of $\gamma_{k,n}$. To simplify the notation, define $\nu_{\gamma_{k,n}} \triangleq \frac{\nu_{\gamma_k}}{\varepsilon + \bar{p}_{k,n}}$. By substituting (27) and (28) into (20), we obtain:

$$\begin{aligned} \gamma_{k,n+1} = & \underbrace{\gamma_{k,n} \left[1 - \nu_{\gamma_{k,n}} \left(\tilde{e}_{k,n}^{(2)} - \tilde{e}_{k,n}^{(1)} \right)^2 \right]}_{\alpha_1} \\ & + \underbrace{\nu_{\gamma_{k,n}} \left[\left(\tilde{e}_{k,n}^{(2)} \right)^2 - \tilde{e}_{k,n}^{(1)} \tilde{e}_{k,n}^{(2)} \right]}_{\alpha_2} \\ & + \underbrace{\nu_{\gamma_{k,n}} \left(\tilde{e}_{k,n}^{(2)} - \tilde{e}_{k,n}^{(1)} \right) z_{k,n}}_{\alpha_3}. \end{aligned} \quad (140)$$

As a consequence, we have:

$$\mathbb{E}\{\gamma_{k,n+1}^2\} = \mathbb{E}\{\alpha_1^2\} + \mathbb{E}\{\alpha_2^2\} + \mathbb{E}\{\alpha_3^2\} + 2\mathbb{E}\{\alpha_1\alpha_2\}. \quad (141)$$

By resorting to approximations **Ap**₂, **Ap**₅–**Ap**₇, and using expression (140), we obtain:

$$\begin{aligned} \mathbb{E}\{\alpha_1^2\} = & \mathbb{E}\{\gamma_{k,n}^2\} \left[1 + 3\mathbb{E}\{\nu_{\gamma_{k,n}}^2\} \left(\Delta J_{k,n}^{(1)} + \Delta J_{k,n}^{(2)} \right)^2 \right. \\ & \left. - 2\mathbb{E}\{\nu_{\gamma_{k,n}}\} \left(\Delta J_{k,n}^{(1)} + \Delta J_{k,n}^{(2)} \right) \right], \end{aligned} \quad (142)$$

$$\begin{aligned} \mathbb{E}\{\alpha_2^2\} = & \mathbb{E}\{\nu_{\gamma_{k,n}}^2\} J_{\text{ex},k,n}^{(2)} \left(\Delta J_{k,n}^{(1)} + \Delta J_{k,n}^{(2)} \right) \\ & + 2\mathbb{E}\{\nu_{\gamma_{k,n}}^2\} \left(\Delta J_{k,n}^{(2)} \right)^2, \end{aligned} \quad (143)$$

$$\mathbb{E}\{\alpha_3^2\} = \sigma_{z,k}^2 \mathbb{E}\{\nu_{\gamma_{k,n}}^2\} \left(\Delta J_{k,n}^{(1)} + \Delta J_{k,n}^{(2)} \right), \quad (144)$$

$$\begin{aligned} \mathbb{E}\{\alpha_1\alpha_2\} = & \mathbb{E}\{\gamma_{k,n}\} \left[\mathbb{E}\{\nu_{\gamma_{k,n}}\} \Delta J_{k,n}^{(2)} \right. \\ & \left. - 3\mathbb{E}\{\nu_{\gamma_{k,n}}^2\} \left(\Delta J_{k,n}^{(1)} + \Delta J_{k,n}^{(2)} \right) \Delta J_{k,n}^{(2)} \right]. \end{aligned} \quad (145)$$

By adopting approximation $\mathbb{E}\{\nu_{\gamma_{k,n}}^2\} \approx [\mathbb{E}\{\nu_{\gamma_{k,n}}\}]^2$ in (142)–(145), we obtain the expression of $\mathbb{E}\{\gamma_{k,n+1}^2\}$. Taking the limit of (141) with $n \rightarrow \infty$, and solving for $\mathbb{E}\{\gamma_{k,\infty}^2\} \triangleq \lim_{n \rightarrow \infty} \mathbb{E}\{\gamma_{k,n}^2\}$, we obtain the steady-state value (78), with $\bar{\nu}_\infty \triangleq \mathbb{E}\{\nu_{\gamma_{k,\infty}}\} = \lim_{n \rightarrow \infty} \mathbb{E}\{\nu_{\gamma_{k,n}}\}$.

From (141) and the term $\mathbb{E}\{\alpha_1^2\}$, the range of step-sizes to ensure the mean-square stability is given by (146). With the use of approximation (137), then for a constant step-size, a sufficient condition to ensure (138) is given by:

$$0 < \nu_{\gamma_k} < \min_n \left\{ \frac{\varepsilon + \bar{p}_{k,n}}{3 \left(\Delta J_{k,n}^{(1)} + \Delta J_{k,n}^{(2)} \right)} \right\}. \quad (147)$$

Substituting $\bar{p}_{k,n}$ of (137) into (147), and after some mathematical manipulations, we arrive at condition (73).

APPENDIX I

PROOF OF STABILITY ANALYSIS RESULTS 7 & 8

We first evaluate the mean behavior of $\gamma_{k,n}$ for the sign-regressor diffusion scheme in (79). To make the analysis tractable, we introduce approximation **Ap**₅ and joint Gaussian assumption **Ap**₇ in the same way. Furthermore, by resorting to Price's theorem [24], the following approximations hold:

$$\mathbb{E}\left\{\tilde{e}_{k,n}^{(2)} \text{sgn}\{\tilde{e}_{k,n}^{(2)} - \tilde{e}_{k,n}^{(1)}\}\right\} \approx \sqrt{\frac{2}{\pi}} \frac{\Delta J_{k,n}^{(2)}}{\sqrt{\Delta J_{k,n}^{(1)} + \Delta J_{k,n}^{(2)}}} \quad (148)$$

$$\mathbb{E}\left\{|\tilde{e}_{k,n}^{(2)} - \tilde{e}_{k,n}^{(1)}|\right\} \approx \sqrt{2(\Delta J_{k,n}^{(1)} + \Delta J_{k,n}^{(2)})/\pi}. \quad (149)$$

It is noted that approximation (148) is used in evaluating the transient and steady-state behaviors of $\mathbb{E}\{\gamma_{k,n}\}$ and $\mathbb{E}\{\gamma_{k,n}^2\}$.

Then, under **Ap**₅ and **Ap**₇, and with the use of approximation (149), iteration (79) becomes:

$$\begin{aligned} \mathbb{E}\{\gamma_{k,n+1}\} &\approx \mathbb{E}\{\gamma_{k,n}\} + \nu_{\gamma_k} \mathbb{E}\left\{\tilde{e}_{k,n}^{(2)} \text{sgn}\{\tilde{e}_{k,n}^{(2)} - \tilde{e}_{k,n}^{(1)}\}\right\} \\ &\quad - \nu_{\gamma_k} \mathbb{E}\{\gamma_{k,n}\} \sqrt{2(\Delta J_{k,n}^{(1)} + \Delta J_{k,n}^{(2)})/\pi}. \end{aligned} \quad (150)$$

To ensure the mean stability of $\gamma_{k,n}$, the step-size parameter ν_{γ_k} must satisfy:

$$\left|1 - \nu_{\gamma_k} \sqrt{2(\Delta J_{k,n}^{(1)} + \Delta J_{k,n}^{(2)})/\pi}\right| < 1 - \phi, \quad \forall n \quad (151)$$

with a positive ϕ . A sufficient condition is then given by:

$$0 < \nu_{\gamma_k} < \frac{1}{\max_n \left\{ \sqrt{2(\Delta J_{k,n}^{(1)} + \Delta J_{k,n}^{(2)})/\pi} \right\}}, \quad (152)$$

which leads to condition (82) directly. In addition, by taking the limit of (150) as $n \rightarrow \infty$, we obtain the same steady-state value for $\mathbb{E}\{\gamma_{k,n}\}$ as in (95).

To evaluate the mean-square behavior of $\gamma_{k,n}$, by substituting (27), (28) into (22) and rearranging terms, we obtain:

$$\begin{aligned} \gamma_{k,n+1} &= \underbrace{(1 - \nu_{\gamma_k} |\tilde{e}_{k,n}^{(2)} - \tilde{e}_{k,n}^{(1)}|)}_{\alpha_4} \gamma_{k,n} \\ &\quad + \underbrace{\nu_{\gamma_k} \tilde{e}_{k,n}^{(2)} \text{sgn}\{\tilde{e}_{k,n}^{(2)} - \tilde{e}_{k,n}^{(1)}\}}_{\alpha_5} \\ &\quad + \underbrace{\nu_{\gamma_k} z_{k,n} \text{sgn}\{\tilde{e}_{k,n}^{(2)} - \tilde{e}_{k,n}^{(1)}\}}_{\alpha_6}. \end{aligned} \quad (153)$$

By squaring (153) and taking expectation, we have

$$\mathbb{E}\{\gamma_{k,n+1}^2\} = \mathbb{E}\{\alpha_4^2\} + \mathbb{E}\{\alpha_5^2\} + \mathbb{E}\{\alpha_6^2\} + 2\mathbb{E}\{\alpha_4\alpha_5\}. \quad (154)$$

To obtain an explicit expression for the mean-square behavior of $\gamma_{k,n}$, we have to evaluate the terms on RHS of (154) one by one under the joint Gaussian assumption **Ap**₇. By resorting to **Ap**₅, **Ap**₇ and utilizing (153), we obtain:

$$\begin{aligned} \mathbb{E}\{\alpha_4\alpha_5\} &= \mathbb{E}\{\gamma_{k,n}\} \left[\nu_{\gamma_k} \mathbb{E}\left\{\tilde{e}_{k,n}^{(2)} \text{sgn}\{\tilde{e}_{k,n}^{(2)} - \tilde{e}_{k,n}^{(1)}\}\right\} - \nu_{\gamma_k}^2 \Delta J_{k,n}^{(2)} \right] \\ &\approx \mathbb{E}\{\gamma_{k,n}\} \left[\sqrt{\frac{2}{\pi}} \frac{\nu_{\gamma_k} \Delta J_{k,n}^{(2)}}{\sqrt{\Delta J_{k,n}^{(1)} + \Delta J_{k,n}^{(2)}}} - \nu_{\gamma_k}^2 \Delta J_{k,n}^{(2)} \right], \end{aligned} \quad (155)$$

$$\begin{aligned} \mathbb{E}\{\alpha_4^2\} &\approx \mathbb{E}\{\gamma_{k,n}^2\} \left[1 + \nu_{\gamma_k}^2 (\Delta J_{k,n}^{(1)} + \Delta J_{k,n}^{(2)}) \right. \\ &\quad \left. - 2\nu_{\gamma_k} \sqrt{2(\Delta J_{k,n}^{(1)} + \Delta J_{k,n}^{(2)})/\pi} \right], \end{aligned} \quad (156)$$

$$\mathbb{E}\{\alpha_5^2\} \approx \nu_{\gamma_k}^2 J_{\text{ex},k,n}^{(2)}, \quad (157)$$

$$\mathbb{E}\{\alpha_6^2\} \approx \nu_{\gamma_k}^2 \sigma_{z,k}^2, \quad (158)$$

and we used $[\text{sgn}\{\tilde{e}_{k,n}^{(2)} - \tilde{e}_{k,n}^{(1)}\}]^2 \approx 1$ in the derivations of (157) and (158). Equations (154) and (155)–(158) constitute the iteration of the mean-square behavior of $\gamma_{k,n}$. From (154) and the term $\mathbb{E}\{\alpha_4^2\}$, we obtain the condition to ensure the mean-square stability of $\gamma_{k,n}$ as:

$$\left|1 + \nu_{\gamma_k}^2 (\Delta J_{k,n}^{(1)} + \Delta J_{k,n}^{(2)}) - \nu_{\gamma_k} \sqrt{8(\Delta J_{k,n}^{(1)} + \Delta J_{k,n}^{(2)})/\pi}\right| < 1 - \phi \quad (159)$$

for all n . For a constant step-size, a sufficient condition to ensure (159) is given by equation (83). In addition, by taking the limit of (154) as $n \rightarrow \infty$, we obtain the steady-state value of $\mathbb{E}\{\gamma_{k,n}^2\}$ as (84).

REFERENCES

- [1] P. Braca, S. Marano, and V. Matta, “Enforcing consensus while monitoring the environment in wireless sensor networks,” *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3375–3380, Jul. 2008.
- [2] A. Nedic and A. Ozdaglar, “Distributed subgradient methods for multi-agent optimization,” *IEEE Trans. Autom. Control*, vol. 54, no. 1, pp. 48–61, Jan. 2009.
- [3] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, “Gossip algorithms for distributed signal processing,” *Proc. of the IEEE*, vol. 98, no. 11, pp. 1847–1864, Nov. 2010.
- [4] K. Srivastava and A. Nedic, “Distributed asynchronous constrained stochastic optimization,” *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 4, pp. 772–790, Aug. 2011.
- [5] A. H. Sayed, “Diffusion adaptation over networks,” in *Academic Press Library in Signal Processing*, R. Chellapa and S. Theodoridis, Eds., vol. 3, pp. 322–454. Elsevier, 2014.
- [6] A. H. Sayed, “Adaptive networks,” *Proc. of the IEEE*, vol. 102, no. 4, pp. 460–497, Apr. 2014.
- [7] S.-Y. Tu and A. H. Sayed, “Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks,” *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6217–6234, Dec. 2012.
- [8] J. Chen, C. Richard, and A. H. Sayed, “Multitask diffusion adaptation over networks,” *IEEE Trans. Signal Process.*, vol. 62, no. 16, pp. 4129–4144, Aug. 2014.
- [9] J. Chen, C. Richard, and A. H. Sayed, “Diffusion LMS over multitask networks,” *IEEE Trans. Signal Process.*, vol. 63, no. 11, pp. 2733–2748, Jun. 2015.
- [10] 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, 2017.
- [11] L. Li and J. Chambers, “Distributed adaptive estimation based on the APA algorithm over diffusion networks with changing topology,” in *Proc. IEEE SSP*, Cardiff, UK, 2009, pp. 757–760.
- [12] F. Cattivelli, C. G. Lopes, and A. H. Sayed, “Diffusion recursive least-squares for distributed estimation over adaptive networks,” *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1865–1877, May 2008.
- [13] X. Zhao and A. H. Sayed, “Clustering via diffusion adaptation over networks,” in *Proc. Int. Workshop Cognitive Inf. Process.*, Baiona, Spain, May. 2012, pp. 1–6.
- [14] J. Arenas-Garcia, A. R. Figueiras-Vidal, and A. H. Sayed, “Mean-square performance of a convex combination of two adaptive filters,” *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1078–1090, Mar. 2006.
- [15] J. Arenas-Garcia, L. A. Azpicueta-Ruiz, M. T. M. Silva, V. H. Nascimento, and A. H. Sayed, “Combinations of adaptive filters: Performance and convergence properties,” *IEEE Signal Process. Mag.*, vol. 33, no. 1, pp. 120–140, Jan. 2016.
- [16] S. S. Kozat, A. T. Erdogan, A. C. Singer, and A. H. Sayed, “Steady-state MSE performance analysis of mixture approaches to adaptive filtering,” *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4050–4063, Aug. 2010.
- [17] J. Arenas-Garcia, M. Martinez-Ramon, A. Navia-Vazquez, and A. R. Figueiras-Vidal, “Plant identification via adaptive combination of transversal filters,” *Signal Process.*, vol. 86, no. 9, pp. 2430 – 2438, 2006.
- [18] R. Alain, B. Francis, C. Stephane, and G. Yves, “Simple MKL,” *Journal of Mach. Learn. Research*, vol. 9, no. 3, pp. 2491–2521, 2008.
- [19] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *Proc. of CVPR*, Boston, MA, USA, June 2015, pp. 1–9.

- [20] N. J. Bershad, J. C. M. Bermudez, and J. Y. Tourneret, "An affine combination of two LMS adaptive filters—transient mean-square analysis," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1853–1864, May 2008.
- [21] J. Fernandez-Bes, L. A. Azpicueta-Ruiz, J. Arenas-Garcia, and M. T. M. Silva, "Distributed estimation in diffusion networks using affine least-squares combiners," *Digital Signal Process.*, vol. 36, pp. 1 – 14, 2015.
- [22] D. Jin, J. Chen, and J. Chen, "Convex combination of diffusion strategies over distributed networks," in *Proc. Asia-Pacific Signal Inf. Process. Association (APSIPA)*, Hawaii, USA, Nov. 2018, pp. 224–228.
- [23] J. Fernandez-Bes, J. Arenas-Garcia, M. T. M. Silva, and L. A. Azpicueta-Ruiz, "Adaptive diffusion schemes for heterogeneous networks," *IEEE Trans. Signal Process.*, vol. 65, no. 21, pp. 5661–5674, Nov. 2017.
- [24] A. H. Sayed, *Adaptive Filters*, John Wiley & Sons, Inc., 2008.
- [25] L. A. Azpicueta-Ruiz, A. R. Figueiras-Vidal, and J. Arenas-Garcia, "A normalized adaptation scheme for the convex combination of two adaptive filters," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Las Vegas, NV, USA, Mar. 2008, pp. 3301–3304.
- [26] R. Candido, M. T. M. Silva, and V. H. Nascimento, "Transient and steady-state analysis of the affine combination of two adaptive filters," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4064–4078, Aug. 2010.
- [27] A. H. Sayed, *Adaptation, Learning, and Optimization over Networks*, vol. 7, Now Publishers Inc., Hanover, MA, USA, Jul. 2014.
- [28] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill Higher Education, 4 edition, 2002.
- [29] J. Chen, C. Richard, Y. Song, and D. Brie, "Transient performance analysis of zero-attracting LMS," *IEEE Signal Process. Letters*, vol. 23, no. 12, pp. 1786–1790, Dec. 2016.
- [30] S. Haykin, *Adaptive Filter Theory*, Pearson Education India, 4th edition, 2005.
- [31] V. H. Nascimento, M. T. M. Silva, R. Candido, and J. Arenas-Garcia, "A transient analysis for the convex combination of adaptive filters," in *Proc. IEEE SSP*, Cardiff, UK, Aug. 2009, pp. 53–56.
- [32] M. T. M. Silva, V. H. Nascimento, and J. Arenas-Garcia, "A transient analysis for the convex combination of two adaptive filters with transfer of coefficients," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Dallas, TX, USA, Mar. 2010, pp. 3842–3845.
- [33] D. Jin, J. Chen, C. Richard, and J. Chen, "Model-driven online parameter adjustment for zero-attracting LMS," *Signal Process.*, vol. 152, pp. 373–383, Nov. 2018.
- [34] R. Nassif, C. Richard, A. Ferrari, and A. H. Sayed, "Multitask diffusion LMS with sparsity-based regularization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Australia, Apr. 2015, pp. 3516–3520.
- [35] M. Fiedler, "Algebraic connectivity of graphs," *Czechoslovak Mathematical Journal*, vol. 23, no. 2, pp. 298–305, 1973.
- [36] A. Simões and J. Xavier, "FADE: Fast and asymptotically efficient distributed estimator for dynamic networks," *IEEE Trans. Signal Process.*, vol. 67, no. 8, pp. 2080–2092, Apr. 2019.
- [37] S. Das and J. M. F. Moura, "Distributed state estimation in multi-agent networks," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Vancouver, BC, Canada, May 2013, pp. 4246–4250.
- [38] A. Bertrand, M. Moonen, and A. H. Sayed, "Diffusion bias-compensated RLS estimation over adaptive networks," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5212–5224, Nov. 2011.
- [39] M. J. Piggott and V. Solo, "Diffusion LMS with correlated regressors I: Realization-wise stability," *IEEE Trans. Signal Process.*, vol. 64, no. 21, pp. 5473–5484, Nov. 2016.
- [40] M. J. Piggott and V. Solo, "Diffusion LMS with correlated regressors II: Performance," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 3934–3947, Aug. 2017.
- [41] H. H. Sohrab, *Basic Real Analysis*, Birkhäuser, 2014.
- [42] P. Di Lorenzo and A. H. Sayed, "Sparse distributed learning based on diffusion adaptation," *IEEE Trans. Signal Process.*, vol. 61, no. 6, pp. 1419–1433, Mar. 2013.
- [43] R. Nassif, C. Richard, A. Ferrari, and A. H. Sayed, "Proximal multitask learning over networks with sparsity-inducing coregularization," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6329–6344, Dec. 2016.
- [44] J. Chen, C. Richard, S. K. Ting, and A. H. Sayed, "Chapter # - Multitask learning over adaptive networks with grouping strategies," in *Cooperative and Graph Signal Processing*, pp. 107 – 129. Academic Press, 2018.