Resilient multitask distributed adaptation over networks with noisy exchanges

Wang, Chengcheng; Tay, Wee Peng; Wei, Ye; Wang, Yuan

2020

Wang, C., Tay, W. P., Wei, Y. & Wang, Y. (2020). Resilient multitask distributed adaptation over networks with noisy exchanges. 2020 IEEE 11th Sensor Array and Multichannel Signal Processing Workshop (SAM). <https://dx.doi.org/10.1109/SAM48682.2020.9104281>

<https://hdl.handle.net/10356/152713>

<https://doi.org/10.1109/SAM48682.2020.9104281>

© 2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at: <https://doi.org/10.1109/SAM48682.2020.9104281>.

Downloaded on 26 Apr 2023 19:34:56 SGT

Resilient Multitask Distributed Adaptation Over Networks With Noisy Exchanges

Chengcheng Wang* Wee Peng Tay* Ye Wei† Yuan Wang*

* School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798

† School of Automation Engineering, Northeast Electric Power University, China 132012

Abstract—We develop a resilient distributed strategy over multitask networks, where individual tasks are linearly related within each neighborhood, and information exchanges between neighboring agents are noisy. In the proposed strategy, each agent follows an adapt-then-project procedure to iteratively update its local estimate. In particular, weighted projection operators are utilized in the projection step in order to attenuate the negative effect of noisy exchanges on the cooperative inference performance. We motivate a strategy for computing the weights in a distributed and adaptive manner. Simulation results demonstrate that the proposed scheme shows good resilience against noise in the information exchange between agents.

Index Terms—distributed strategies, adaptive networks, multitask networks, noisy exchanges, projection approach

I. INTRODUCTION

In multitask networks, a set of interconnected agents work collaboratively to estimate different but related parameters of interest [1]–[4]. In order to make use of the relationship between different tasks for better inference performance, local estimates are exchanged amongst agents within the same neighborhood. However, the network may suffer from imperfect information exchange. This may result from noisy communication links [5]–[7] or privacy-preserving mechanism where an additive noise is added purposefully by each agent before sharing with its neighbors in order to protect its local task [8].

The works [5]–[7], [9]–[12] show that noisy exchanges result in performance degradation of distributed strategies. Then, in order to counter the effect of such degradation, resilient distributed schemes are developed in [5]–[7], [9], [12]–[17]. These references excluding [7] assume that all agents in the network are interested in inferring the same parameters, while [7] studies how to attenuate the negative effect of running a *single-task* distributed strategy in a multitask environment in the presence of noisy links.

The objective of this work is to develop a resilient *multitask* distributed strategy over networks in the presence of noisy information exchanges between neighboring agents. Specifically, we consider multitask estimation problems where the unknown parameters of interest within each neighborhood are linearly

related with each other [3], [8]. Such problems widely exist in applications such as electrical networks, telecommunication networks, and pipeline networks [3]. In the proposed scheme, each agent follows an adapt-then-project procedure to iteratively update its local estimate. Different from the work [3], which does not take noisy exchanges into consideration, we propose to use *weighted* projection operators in the projection step in order to attenuate the negative effect of exchange noises on the cooperative inference performance. We study how to compute the involved weights in a distributed and adaptive manner. In addition, the work [8] proposes to exchange noisy local estimates amongst neighboring agents in order to protect agents’ local tasks against privacy leakage, and studies the choice guideline for powers of the exchange noises.

The rest of this paper is organized as follows. In Section II, we formulate the multitask estimation problem considered in this paper. A multitask distributed scheme is then proposed by following an adapt-then-project procedure to solve the problem in Section III-A. In particular, weighted projection operators are utilized in the projection step to enhance resilience against noisy exchanges, and these operators are introduced in Section III-B. We study the choice guideline for the involved weights, and motivate an adaptive procedure to compute these weights locally at each agent in Section III-C. We present the simulation results in Section IV. Section V concludes the paper.

Notations: We use lowercase letters to denote vectors and scalars, uppercase letters for matrices, plain letters for deterministic variables, and boldface letters for random variables. We also use $(\cdot)^\top$ to denote transposition, $(\cdot)^{-1}$ for matrix inversion, $\text{Tr}(\cdot)$ for the trace of a matrix, $\text{diag}\{\cdot\}$ for a diagonal matrix, $\text{col}\{\cdot\}$ for a column vector, $\|\cdot\|$ for the two-induced norm of a matrix or the Euclidean norm of a vector, $\|x\|_\Sigma^2$ for the weighted square value $x^\top \Sigma x$, and $\exp(\cdot)$ for the natural exponential function.

II. MULTITASK NETWORKS WITH NOISY EXCHANGES

In this section, we present our system model, and give a brief introduction to multitask networks, where neighboring agents’ tasks are linearly related. Consider a strongly-connected network of N agents, where information can flow in either direction between any two connected agents [18]. At each time instant i , each agent k has access to a scalar observation $d_k(i)$, and an $M_k \times 1$ regression vector $u_k(i)$.

This work was supported in part by the Delta-NTU Corporate Lab for Cyber-Physical Systems with funding support from Delta Electronics Inc. and the National Research Foundation (NRF) Singapore under the Corp Lab@University Scheme, by the Singapore Ministry of Education Academic Research Fund Tier 2 grant MOE2018-T2-2-019 and by A*STAR under its RIE2020 Advanced Manufacturing and Engineering (AME) Industry Alignment Fund – Pre Positioning (IAF-PP) (Grant No. A19D6a0053).

The random data $\{\mathbf{d}_k(i), \mathbf{u}_k(i)\}$ are related via the linear regression model

$$\mathbf{d}_k(i) = \mathbf{u}_k^\top(i) \mathbf{w}_k^o + \mathbf{v}_k(i) \quad (1)$$

where the scalar $\mathbf{v}_k(i)$ is measurement noise, $\mathbf{w}_k^o$ is an $M_k \times 1$ unknown parameter vector. We make the following assumptions regarding model (1): (a) firstly, the measurement noise $\mathbf{v}_k(i)$ is white over time and space with zero mean and a variance of $\sigma_{v,k}^2$; (b) secondly, the regression data $\{\mathbf{u}_k(i)\}$ are zero-mean, white over time and space with $\mathbb{E} \mathbf{u}_k(i) \mathbf{u}_k^\top(i) = \mathbf{R}_{u,k}$, where $\mathbf{R}_{u,k}$ is symmetric positive definite; and (c) thirdly, the random data $\{\mathbf{u}_k(i), \mathbf{v}_\ell(j)\}$ are independent of each other for any agents k, ℓ and time instants i, j .

The objective of each agent k is to find the minimizer of the following mean-square-error cost function:

$$J_k(w_k) = \mathbb{E}[(\mathbf{d}_k(i) - \mathbf{u}_k^\top(i) w_k)^2]. \quad (2)$$

Let $\mathcal{N}_k$ be the set of all neighboring agents of agent k , including agent k itself. We assume that neighboring tasks $\{w_k^o, w_\ell^o\}$ for any $\ell \in \mathcal{N}_k \setminus \{k\}$ are involved in at least one linear equality of the form [8]:

$$\sum_{k \in \mathcal{I}_q} D_{qk} w_k^o + b_q = 0, \quad (3)$$

where the subscript “ q ” is the index of the linear equality, the set $\mathcal{I}_q$ includes all agents involved in the q -th equality, and the coefficient matrix D_{qk} and constant vector b_q are of size $L_q \times M_k$ and $L_q \times 1$, respectively. Then, the objective for the entire network is to find the optimal solution to the following constrained optimization problem [3], [8]:

$$\min_{w_1, \dots, w_N} J(w_1, \dots, w_N) = \sum_{k=1}^N J_k(w_k) \quad (4a)$$

$$\text{s. t. } \sum_{k \in \mathcal{I}_q} D_{qk} w_k + b_q = 0, \text{ for } q = 1, \dots, Q. \quad (4b)$$

Let $w = \text{col}\{w_1, \dots, w_N\}$. We proceed to write the constraints (4b) more compactly as $\mathcal{D}w + b = 0$, where the matrix $\mathcal{D}$ is a $Q \times N$ block matrix with blocks $\{D_{qk}\}$ for $q = 1, \dots, Q$ and $k = 1, \dots, N$, and the vector b is a $Q \times 1$ block vector with blocks $\{b_q\}$ for $q = 1, \dots, Q$. We assume that matrix $\mathcal{D}$ is full row-rank to ensure that there is at least one solution to $\mathcal{D}w + b = 0$ [3]. Let j_k be the total number of linear equality constraints that agent k is involved in. We assume that each agent k is involved in at least one linear equality constraint, i.e., $j_k \geq 1$, and that each agent k has access to all j_k linear equalities that it is involved in. In addition, it is assumed that all agents involved in these j_k linear equalities are inside $\mathcal{N}_k$, i.e., $\mathcal{I}_q \subset \mathcal{N}_k$ if $k \in \mathcal{I}_q$ [3].

As demonstrated in [3], under perfect information exchange conditions, each agent k can benefit through cooperation with neighboring agents by sharing their local parameter estimates with their neighbors, which enables it to leverage the linear relationships (3) and its neighbors’ parameter estimates to improve its own inference accuracy. In this paper, we consider the scenario where the network suffers from imperfect information

exchanges. This may result from noisy communication links [5]–[7] or privacy-preserving mechanism where an additive noise is added purposefully by each agent before sharing with its neighbors in order to protect its local task [8]. Our objective is to motivate and develop a resilient distributed adaptation scheme over multitask networks in the presence of imperfect information exchanges, which is able to counter the negative effect of noisy exchanges on the cooperative inference performance.

Before proceeding, we make assumptions about the local constraints and exchange noises. Let $n_k = |\mathcal{N}_k|$ be the cardinality of $\mathcal{N}_k$. We introduce an $n_k \times 1$ block vector $w_{k,e} = \text{col}\{w_\ell\}_{\ell \in \mathcal{N}_k}$, where the subscript ‘ k, e ’ denotes that the quantity is extended by the neighboring agents $\{\ell \in \mathcal{N}_k\}$. Then, we proceed to write the j_k constraints that agent k is involved in more compactly as:

$$\mathcal{D}_k w_{k,e} + b_k = 0, \quad (5)$$

where matrix $\mathcal{D}_k$ is a $j_k \times n_k$ block matrix, and vector b_k is a $j_k \times 1$ block vector. We assume that the local constraints satisfy the following conditions.

Assumption 1. (Conditions on local constraint) Matrix $\mathcal{D}_k$ is full row-rank and column-rank deficient for any agent k .

Let $M_\ell \times 1$ random vector $\mathbf{n}_{\ell k}(i)$ be the exchange noise that perturbs the exchange of information from agent ℓ to agent k at time instant i . We assume that the exchange noise processes satisfy the following conditions.

Assumption 2. (Conditions on exchange noise) The exchange noises $\{\mathbf{n}_{\ell k}(i)\}$ are zero-mean, with a diagonal covariance matrix given by $\mathbb{E}[\mathbf{n}_{\ell k}(i) \mathbf{n}_{\ell k}^\top(i)] = \mathbf{R}_{n,\ell k}$. The random noises $\{\mathbf{n}_{\ell k}(i)\}$ are white over time and space. The random process $\{\mathbf{n}_{\ell k}(i)\}$ is independent of any other random processes.

Note that Assumption 1 enables us to use weighted projection operators to counter the negative effect of exchange noises as shown in future Section III-B, and Assumption 2 simplifies our choice guideline for the involved weights as shown by future (19).

III. RESILIENT MULTITASK DISTRIBUTED ADAPTATION SCHEME

In this section, we propose a distributed adaptation and learning strategy over multitask networks with noisy exchanges by following an *adapt-then-project* (ATP) procedure. We call the proposed scheme ATP algorithm. In order to enhance resilience against exchange noises, we introduce *weighted* projection operators at each agent k by solving a weighted-least-squares problem subject to local linear equality constraints (5) as shown by future (10). We then develop a distributed adaptation scheme to learn the involved weights in real time.

A. Adapt-then-Project Distributed Strategy

In the proposed ATP algorithm, we initialize $w_k(-1) = 0$ for every agent k in the network. Given data $\{\mathbf{d}_k(i), \mathbf{u}_k(i)\}$

for each time instant $i \geq 0$, and for each agent $k = 1, \dots, N$, we perform the following steps iteratively:

- 1) Adaptation. Each agent k updates its local current estimate $\mathbf{w}_k(i-1)$ of $\mathbf{w}_k^o$ to an intermediate estimate $\psi_k(i)$ by following the stochastic gradient descent (SGD) algorithm

$$\psi_k(i) = \mathbf{w}_k(i-1) + \mu_k \mathbf{u}_k(i) (\mathbf{d}_k(i) - \mathbf{u}_k^T(i) \mathbf{w}_k(i-1)) \quad (6)$$

where $\mu_k > 0$ is the step-size parameter at agent k .

- 2) Exchange. Each agent k collects estimates $\{\psi_{\ell k}(i)\}$ from neighboring agents $\{\ell \in \mathcal{N}_k\}$

$$\psi_{\ell k}(i) = \begin{cases} \psi_{\ell}(i) + \mathbf{n}_{\ell k}(i), & \text{if } \ell \in \mathcal{N}_k \setminus \{k\}, \\ \psi_k(i), & \text{if } \ell = k \end{cases} \quad (7)$$

where $\psi_{\ell k}(i)$ is the noisy intermediate estimate received by agent k from its neighbor $\ell \in \mathcal{N}_k$, and the random additive noise vector $\mathbf{n}_{\ell k}(i)$ is the exchange noise.

- 3) Projection. Each agent k projects the estimates received from its neighborhood $\{\psi_{\ell k}(i)\}$ onto the linear manifold $\{\mathbf{w}_{k,e} \mid \mathcal{D}_k \mathbf{w}_{k,e} + \mathbf{b}_k = 0\}$:

$$\mathbf{w}_k(i) = \mathcal{P}_k \text{col} \{\psi_{\ell k}(i)\}_{\ell \in \mathcal{N}_k} - \mathbf{f}_k \quad (8)$$

where the matrix $\mathcal{P}_k$ and vector $\mathbf{f}_k$ will be introduced in future expressions (16) and (17) respectively.

Note that the additive noises $\{\mathbf{n}_{\ell k}(i)\}$ in (7) can be used to model either communication link noises [5]–[7] or privacy mechanism noises [8].

Remark 1: The projection approach (8) differs from that in [3], which considers perfect information exchanges. In [3], neighboring estimates of agent k are projected onto the linear manifold corresponding to each of the j_k constraints separately, which generates j_k intermediate estimates, and the new estimate is the average of these intermediate estimates. In addition, the work [8] focuses on quantifying powers of the exchange noises in order to protect each individual task against privacy leakage, which is different from our objective.

B. Weighted Projection Operators

For each agent k , let us collect the noisy intermediate estimates $\{\psi_{\ell k}(i)\}$ from its neighboring agents $\{\ell \in \mathcal{N}_k\}$ into an $n_k \times 1$ block vector $\psi_{k,e}(i) = \text{col} \{\psi_{\ell k}(i)\}_{\ell \in \mathcal{N}_k}$. Then, for each random realization $\psi_{k,e}(i) = \psi_{k,e}(i)$, we are interested in seeking the optimal solution

$$\mathbf{w}_{k,e}(i) = \text{col} \{\mathbf{w}_{\ell k}(i)\}_{\ell \in \mathcal{N}_k} \quad (9)$$

to the following optimization problem:

$$\min_{\tilde{\mathbf{w}}_{k,e}(i)} \|\psi_{k,e}(i) - \tilde{\mathbf{w}}_{k,e}(i)\|_{\Omega_{k,e}^{-1}}^2 \quad (10a)$$

$$\text{s. t. } \mathcal{D}_k \tilde{\mathbf{w}}_{k,e}(i) + \mathbf{b}_k = 0, \quad (10b)$$

where the weight matrix

$$\Omega_{k,e}^{-1} = \text{diag} \{\omega_{\ell k} I_{M_\ell}\}_{\ell \in \mathcal{N}_k} \quad (11)$$

with I_M being an $M \times M$ identity matrix, and where

$$\omega_{\ell k} > 0, \text{ for any } \ell \in \mathcal{N}_k, \text{ and } \sum_{\ell \in \mathcal{N}_k} \omega_{\ell k} = 1. \quad (12)$$

Remark 2: In (10a), the weight matrix $\Omega_{k,e}^{-1}$ is introduced in order to mitigate the negative effect of exchange noises on the projection step (8). To be specific, (10a) can be rewritten as $\|\psi_{k,e}(i) - \tilde{\mathbf{w}}_{k,e}(i)\|_{\Omega_{k,e}^{-1}}^2 = \sum_{\ell \in \mathcal{N}_k} \omega_{\ell k} \|\psi_{\ell k}(i) - \tilde{\mathbf{w}}_{\ell k}(i)\|^2$. Then, our objective is to assign a small weight to the square Euclidean distance $\|\psi_{\ell k}(i) - \tilde{\mathbf{w}}_{\ell k}(i)\|^2$ if the received intermediate estimate $\psi_{\ell k}(i)$ involves an exchange noise with large power. In the special case when $\Omega_{k,e}^{-1} = \text{diag} \{1/n_k I_{M_\ell}\}_{\ell \in \mathcal{N}_k}$, the optimal solution of (10) is reduced to the Euclidean projection of $\psi_{k,e}(i)$ on the affine set $\{\mathbf{w}_{k,e} \mid \mathcal{D}_k \mathbf{w}_{k,e} + \mathbf{b}_k = 0\}$ [19, p.398].

In view of (11) and (12), it holds that $\Omega_{k,e}^{-1}$ is an $n_k \times n_k$ block-diagonal, symmetric and positive definite matrix. Let $M_{k,e} = \sum_{\ell \in \mathcal{N}_k} M_\ell$. We introduce

$$\mathcal{P}_{k,e} = I_{M_{k,e}} - \Omega_{k,e} \mathcal{D}_k^T (\mathcal{D}_k \Omega_{k,e} \mathcal{D}_k^T)^{-1} \mathcal{D}_k, \quad (13)$$

$$\mathbf{f}_{k,e} = \Omega_{k,e} \mathcal{D}_k^T (\mathcal{D}_k \Omega_{k,e} \mathcal{D}_k^T)^{-1} \mathbf{b}_k. \quad (14)$$

It follows that matrix $\mathcal{P}_{k,e}$ is a projection matrix, i.e., $\mathcal{P}_{k,e}^2 = \mathcal{P}_{k,e}$. Then, it holds that the minimizer of (10) is

$$\mathbf{w}_{k,e}(i) = \mathcal{P}_{k,e} \psi_{k,e}(i) - \mathbf{f}_{k,e}. \quad (15)$$

Let $[\mathcal{A}]_k$ be the block row of matrix $\mathcal{A}$ corresponding to agent k . We introduce

$$\mathcal{P}_k = [\mathcal{P}_{k,e}]_k, \quad (16)$$

$$\mathbf{f}_k = [\mathbf{f}_{k,e}]_k, \quad (17)$$

$$\mathbf{w}_k(i) = [\mathbf{w}_{k,e}(i)]_k. \quad (18)$$

Then, we arrive at the update equation (8). Now, we proceed to assign values to the weights $\{\omega_{\ell k}\}$ defined in (11). Note that the projection step (8) at each agent k involves noisy intermediate estimates $\{\psi_{\ell k}(i)\}$ from its neighbors $\{\ell \in \mathcal{N}_k \setminus \{k\}\}$ and the uncontaminated intermediate estimate $\psi_k(i)$. Then, in view of the fact that the noisy intermediate estimate $\psi_{\ell k}(i)$ involves an exchange noise $\mathbf{n}_{\ell k}(i)$ with a *diagonal* covariance matrix of $R_{n,\ell k}$ (from Assumption 2), we set

$$\omega_{\ell k} = \begin{cases} \frac{\exp(-\text{Tr}(R_{n,\ell k}))}{\sum_{\ell \in \mathcal{N}_k \setminus \{k\}} \exp(-\text{Tr}(R_{n,\ell k})) + 1}, & \text{if } \ell \in \mathcal{N}_k \setminus \{k\}, \\ \frac{1}{\sum_{\ell \in \mathcal{N}_k \setminus \{k\}} \exp(-\text{Tr}(R_{n,\ell k})) + 1}, & \text{if } \ell = k, \end{cases} \quad (19)$$

which satisfies the requirements in (12). In addition, the weight $\omega_{\ell k}$ is proportionate to the value of $\exp(-\text{Tr}(R_{n,\ell k}))$ for all $\{\ell \in \mathcal{N}_k \setminus \{k\}\}$, and is set to a larger value for $\ell = k$ (note that $\exp(0) = 1$), since $\psi_k(i)$ is not contaminated. In view of (19), it holds that if the quantity $\text{Tr}(R_{n,\ell k})$ is large, a small weight $\omega_{\ell k}$ is assigned to the square Euclidean distance $\|\psi_{\ell k}(i) - \tilde{\mathbf{w}}_{\ell k}(i)\|^2$. In the extreme case when $\{\text{Tr}(R_{n,\ell k}) \rightarrow \infty\}$ for all $\{\ell \in \mathcal{N}_k \setminus \{k\}\}$, we have $\{\omega_{\ell k} \rightarrow 0\}$ for all $\{\ell \in \mathcal{N}_k \setminus \{k\}\}$, and $\omega_{kk} \rightarrow 1$. Then, the optimization problem in (10) is reduced to

$$\min_{\tilde{\mathbf{w}}_{k,e}(i)} \|\psi_k(i) - [\tilde{\mathbf{w}}_{k,e}(i)]_k\|^2 \quad (20a)$$

$$\text{s. t. } \mathcal{D}_k \tilde{\mathbf{w}}_{k,e}(i) + \mathbf{b}_k = 0, \quad (20b)$$

where we used the fact that $[\psi_{k,e}(i)]_k = \psi_{kk}(i) = \psi_k(i)$. Under Assumption 1, it holds that equation (20b) has infinitely

many solutions. Then, the optimal solution to (20) is $w_k(i) = \psi_k(i)$, where $w_k(i)$ is defined by (18), while the $\{w_{\ell k}(i)\}$ for all $\{\ell \in \mathcal{N}_k \setminus \{k\}\}$ are adjusted such that $\mathcal{D}_k w_{k,e}(i) + b_k = 0$ is satisfied. Then, the proposed ATP algorithm will be reduced to the noncooperative least-mean-squares (LMS) algorithm [20, p.165].

C. Adaptive Weight Matrix

Note that in order to evaluate the weights $\{\omega_{\ell k}\}$ via (19), agent k needs to know the quantities $\{\text{Tr}(R_{n,\ell k})\}$ for all $\{\ell \in \mathcal{N}_k \setminus \{k\}\}$, which are usually not available beforehand. We now describe a procedure that allows every agent k to learn $\{\text{Tr}(R_{n,\ell k})\}$ of its neighbors in an adaptive manner. By following similar techniques as shown in [6, p.414], we have

$$\lim_{i \rightarrow \infty} \mathbb{E} \|\psi_{\ell k}(i) - w_{\ell}(i-1)\|^2 \approx \mu_{\ell}^2 \sigma_{v,\ell}^2 \text{Tr}(R_{u,\ell}) + \text{Tr}(R_{n,\ell k}) \quad (21)$$

for all $\{\ell \in \mathcal{N}_k \setminus \{k\}\}$, and where the second term on the right hand side is dominant for sufficiently small enough μ_{ℓ} . Let $\gamma_{\ell k}(i)$ denote an estimate for $\text{Tr}(R_{n,\ell k})$ that is computed by agent k at time instant i for any neighboring agent $\ell \in \mathcal{N}_k \setminus \{k\}$. Then, each agent k computes $\gamma_{\ell k}(i)$ by using the following instantaneous approximation:

$$\gamma_{\ell k}(i) = \|\psi_{\ell k}(i) - w_{\ell k}(i-1)\|^2,$$

where the quantity $w_{\ell}(i-1)$ in (21) is replaced by $w_{\ell k}(i-1)$, which is the block row of $w_{k,e}(i-1)$ corresponding to agent ℓ as shown by (9). The quantity $w_{k,e}(i-1)$ is computed via (15), and is available to agent k at time instant $i-1$. The variables $\{\gamma_{\ell k}(i)\}$ are then used in place of the quantities $\{\text{Tr}(R_{n,\ell k})\}$ in the rule (19) to adapt entries of the weight matrix $\Omega_{k,e}^{-1}$ over time.

IV. SIMULATION RESULTS

As shown by Fig. 1a, we consider the case when there are $N = 6$ agents in the network. The lengths of the unknown parameter vectors $\{w_k^o\}$ are $\{M_k \equiv 2\}$. The agents in the network are involved in $Q = 5$ linear equality constraints, each of the form [3]: $\sum_{k \in \mathcal{I}_q} d_{qk} w_k^o + b_q \mathbf{1}_2 = 0$ with the scalar parameters $\{d_{qk}, b_q\}$ randomly selected from $[-3, -1] \cup [1, 3]$, and where $\mathbf{1}_M$ denotes an $M \times 1$ vector with all its entries equal to one. Random data $\{u_k(i), v_k(i)\}$ are independent, normally distributed with zero mean and covariance or variance given by $\{R_{u,k} = \sigma_{u,k}^2 I_2, \sigma_{v,k}^2\}$, and white over time and space. Let $\text{SNR}_k = 10 \log_{10} \left(\mathbb{E}[(u_k^T(i) w_k^o)^2] / \sigma_{v,k}^2 \right)$ be the signal-to-noise ratio (SNR) at agent k . Then, the parameters $\{\sigma_{u,k}^2, \sigma_{v,k}^2\}$ are adjusted to make $\{\text{SNR}_k\}$ as shown by Fig. 1b. Exchange noises $\{n_{\ell k}(i)\}$ are zero-mean independent and identically distributed (i.i.d.) Gaussian variables with covariances $\{R_{n,\ell k} \equiv \sigma_n^2 I_2\}$. We test three levels of exchange noises by setting $\sigma_n^2 = 0$, $\sigma_n^2 = 0.5$, and $\sigma_n^2 = 5$, respectively. We compare performance of three algorithms: the multitask diffusion algorithm (algorithm (40) in [3]), the noncooperative LMS algorithm, and the proposed ATP algorithm. For the step-size parameters, we set $\{\mu_k \equiv 0.02\}$ in the noncooperative

LMS and ATP algorithms, and $\{\mu_k/j_k \equiv 0.02\}$ in the multitask diffusion algorithm. Let

$$\text{MSD}_{\text{net}}(i) = \frac{1}{N} \sum_{k=1}^N \mathbb{E} \|w_k^o - w_k(i)\|^2$$

be the network mean-square-deviation (MSD) at time instant i [18, p.583]. Fig. 2 shows the network MSD learning curves of the tested strategies, which are averaged over 200 independent runs. It is clear from that figure that performance of both cooperative algorithms deteriorates as the level of exchange noises increases, and that the proposed ATP algorithm achieves lower steady-state MSD levels in all the tested cases. In addition, the ATP algorithm shows better resilience against noisy exchanges than the existing multitask diffusion algorithm by always delivering superior performance compared to the noncooperative case in all the tested scenarios.

Fig. 1. Network topology with constraints (left) and SNRs across the agents (right).

Fig. 2. Network MSD learning curves of the multitask diffusion algorithm [3], the noncooperative LMS algorithm, and the proposed ATP algorithm.

V. CONCLUSION

We have developed a resilient distributed strategy over multitask networks with noisy exchanges based on the adapt-then-project principle. We proposed weighted projection operators to attenuate the negative effect of exchange noises on the cooperative inference performance. We motivated a procedure for computing the involved weights in a distributed and adaptive manner. Simulation results demonstrate that the proposed scheme shows good resilience against noisy exchanges.

REFERENCES

- [1] 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.
- [2] —, "Diffusion LMS over multitask networks," *IEEE Trans. Signal Process.*, vol. 63, no. 11, pp. 2733 – 2748, Jun. 2015.
- [3] R. Nassif, C. Richard, A. Ferrari, and A. H. Sayed, "Diffusion LMS for multitask problems with local linear equality constraints," *IEEE Trans. Signal Process.*, vol. 65, no. 19, pp. 4979 – 4993, Oct. 2017.
- [4] Y. Wang, W. P. Tay, and W. Hu, "A multitask diffusion strategy with optimized inter-cluster cooperation," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 3, pp. 504 – 517, Apr. 2017.
- [5] X. Zhao, S.-Y. Tu, and A. H. Sayed, "Diffusion adaptation over networks under imperfect information exchange and non-stationary data," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3460 – 3475, Jul. 2012.
- [6] A. H. Sayed, "Diffusion adaptation over networks," in *Academic Press Library in Signal Processing*, R. Chellapa and S. Theodoridis, Eds. Academic Press, Elsevier, 2014, vol. 3, pp. 323 – 453.
- [7] R. Nassif, C. Richard, J. Chen, A. Ferrari, and A. H. Sayed, "Diffusion LMS over multitask networks with noisy links," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Shanghai, China, Mar. 2016.
- [8] C. Wang, W. P. Tay, Y. Wang, and Y. Wei, "A privacy-preserving diffusion strategy over multitask networks," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Brighton, UK, May 2019.
- [9] S.-Y. Tu and A. H. Sayed, "Adaptive networks with noisy links," in *Proc. IEEE Global Telecommun. Conf.*, Houston, TX, Dec. 2011.
- [10] A. Khalili, M. A. Tinati, A. Rastegarnia, and J. A. Chambers, "Steady-state analysis of diffusion LMS adaptive networks with noisy links," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 974 – 979, Feb. 2012.
- [11] —, "Transient analysis of diffusion least-mean squares adaptive networks with noisy channels," *Int. J. Adapt. Control Signal Process.*, vol. 26, no. 2, pp. 171 – 180, 2012.
- [12] X. Zhao and A. H. Sayed, "Combination weights for diffusion strategies with imperfect information exchange," in *Proc. IEEE Int. Conf. on Commun.*, Ottawa, Canada, Jun. 2012.
- [13] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 355 – 369, Jan. 2009.
- [14] G. Mateos, I. D. Schizas, and G. B. Giannakis, "Performance analysis of the consensus-based distributed LMS algorithm," *EURASIP J. Adv. Signal Process.*, pp. 1 – 19, Jan. 2009.
- [15] R. Abdoolee and B. Champagne, "Diffusion LMS algorithms for sensor networks over non-ideal inter-sensor wireless channels," in *Proc. IEEE Int. Conf. Dist. Comput. Sensor Systems*, Barcelona, Spain, Jun. 2011.
- [16] R. Abdoolee, B. Champagne, and A. H. Sayed, "Diffusion LMS strategies for parameter estimation over fading wireless channels," in *Proc. IEEE Int. Conf. on Commun.*, Budapest, Hungary, Jun. 2013.
- [17] Y. He, F. Wang, S. Wang, P. Ren, and B. Chen, "Maximum total correntropy diffusion adaptation over networks with noisy links," *IEEE Trans. Circuits Syst. II*, vol. 66, no. 2, pp. 307 – 311, Feb. 2019.
- [18] A. H. Sayed, "Adaptation, learning, and optimization over networks," *Foundations and Trends in Machine Learning*, vol. 7, no. 4-5, pp. 311 – 801, 2014. [Online]. Available: <http://dx.doi.org/10.1561/22000000051>
- [19] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.
- [20] A. H. Sayed, *Adaptive Filters*. New York, USA: Wiley-IEEE Press, 2008.