Sparsity-Aware Adaptive Link Combination Approach over Distributed Networks

Songtao Lu, Vítor H. Nascimento, Jinping Sun and Zhuangji Wang

Spatial diversity helps parameter estimation in distributed networks. In this paper, a sparsity-aware link combination strategy is proposed, which considers both the spatial sparsity in a network and the inherent sparsity of the system, where two types of zero-attracting adaptive combiners are proposed based on the least-mean-square (LMS) algorithm. The proposed algorithms exploit l_1-norm regularization through the adaptive combination of neighboring nodes' weights such that the proposed algorithms can adaptively track the variations of network topology. Simulation results illustrate the advantages of the proposed link combination algorithm in terms of convergence rate and steady-state performance for distributed sparse system learning.

Introduction: Collaborative signal processing [1, 2] is considered as a promising way to estimate parameters of interest in a distributed network. Each node collects local data and exchanges its own estimate with its neighbors. There are primarily two steps at each node, an adaptive step and a combination step. In the adaptive one, the node updates the local estimate. In the other step, each node combines its own estimates with those from its neighborhood.

In this paper we consider the problem of distributed estimation under two sparsity constraints: first, the case in which the weight vector to be estimated is itself sparse; and second, the use of sparsity-aware algorithms for link selection in order to reduce the communication cost in distributed networks.

In many situations, the parameters of interest may only contain few non-zero coefficients, so the zero-attracting least-mean-square (ZA-LMS) algorithm was proposed for sparse system identification [3], attaining better performance in terms of faster convergence rate and low steady-state errors compared to the standard LMS. Moreover, sparsity-promoting algorithms in distributed learning were also proposed [4], performing better in terms of stability and tracking behavior.

On the other hand, for a fixed node the received signal to noise ratio (SNR) among its neighbors may vary considerably. Therefore, only the nodes with higher SNRs would effectively contribute to enhance the network performance. This creates a form of sparsity in the optimal solution in the spatial domain. An adaptive convex combination scheme was proposed in [5], assigning optimized combination coefficients to neighbors' weights. As the number of nodes increases, the problem of computational complexity becomes more pressing. The exploitation of a subset of connected links is an effective way to avoid low SNR nodes and simultaneously reduce the computation burden. A sparsity-inspired link selection (SILS) algorithm was proposed in [6] which can track the network topology based on residual errors resulting from individual filters. However, the combination vector updating algorithm only depends on the maximum and minimum errors collected from each node's neighbors and the number of active nodes is strongly dependent on the number of neighbor nodes. Therefore, if there are only a few good nodes in a large network, the algorithm may consume a long time to converge.

In this paper, a sparsity-aware link selection approach is proposed for the adaptive combination of neighbor node weights, which takes both spatial and system sparsity into account. An l_1-norm penalty is incorporated to acquire the relationship of residual errors among the connected nodes. Then, a sparse online learning technique is again considered in the combination step, where the combination vector tracks the sparse pattern of the system adaptively. Due to the simplicity of the LMS algorithm, two algorithms, namely zero-attracting distributed LMS (ZA-DLMS) and joint zero-attracting distributed LMS (JZA-DLMS), are proposed to implement sparse combination. From the simulation results, the advantages of the performance of the sparsity-aware algorithms are shown in terms of convergence rate and steady-state mean square deviation (MSD).

Problem Formulation: We consider a distributed network composed of N nodes in a coordinated network, which is similar as in [5]. The set $\mathcal{N}_k$ denotes the neighborhood of node k, including itself, and the cardinality of $\mathcal{N}_k$ is N_k. Each node k collects a scalar measurement $y_k(n) \in \mathbb{R}$ and a column regressor $\boldsymbol{x}_k(n) \in \mathbb{R}^{M\times 1}$ with length M over successive time instants $n \geq 0$, that are related via a linear regression model

$$y_k(n) = \boldsymbol{x}_k^T(n)\boldsymbol{h}_\text{o} + v_k(n), \tag{1}$$

where $\boldsymbol{h}_\text{o} \in \mathbb{R}^{M\times 1}$ is the vector that we want to estimate, assumed sparse [4], $v_k(n) \in \mathbb{R}$ represents noise and modeling errors at each node, assumed zero mean with variance σ_k^2.

The purpose of cooperative estimation is to estimate $\boldsymbol{h}_\text{o}$ at each node k and time n in a distributed manner. Using the ZA-LMS algorithm as the core adaptive filter, we have the adaptation step and combination step [4], i.e., $\boldsymbol{\psi}_k(n) = \boldsymbol{h}_k(n-1) + \mu_k(y_k(n) - \boldsymbol{x}_k^T(n)\boldsymbol{h}_k(n-1)) - \nu_k \frac{\partial \mathcal{J}_k(n)}{\partial \boldsymbol{h}_k}$ and $\boldsymbol{h}_k(n) = \sum_{i\in\mathcal{N}_k} c_{i,k}(n)\boldsymbol{\psi}_i(n)$, where $\boldsymbol{\psi}_k(n)$ is the local estimate, $\mathcal{J}_k(n)$ is the sparse penalty function to regulate the sparsity of the system (here the l_1-norm is considered, i.e., $\mathcal{J}_k(n) = \|\boldsymbol{h}_k(n)\|_1$) and ν_k is the penalty parameter of the kth node. The kth node exploits the measured data $\{y_k(n), \boldsymbol{x}_k(n)\}$ to update its intermediate weight vector $\boldsymbol{\psi}_k(n)$, and $\boldsymbol{h}_k(n)$ combines the estimates $\{\boldsymbol{\psi}_i(n)\}$ received from each node's neighbors, $i \in \mathcal{N}_k$, and $\{c_{i,k}(n)\}$ are the time-varying combination weights. To study the behavior of the whole network, we introduce the notation $\boldsymbol{\Psi}_k(n) \triangleq [\boldsymbol{\psi}_1(n), \ldots, \boldsymbol{\psi}_{N_k}(n)]_{M\times N_k}$ and $\boldsymbol{c}_k(n) \triangleq [c_{1,k}(n), \ldots, c_{N_k,k}(n)]^T \in \mathbb{R}^{N_k\times 1}$. Accordingly, we have $\boldsymbol{h}_k(n) = \boldsymbol{\Psi}_k(n)\boldsymbol{c}_k(n)$.

In this paper, we focus on an approach to effectively update the $c_{i,k}(n)$ such that the performance of the network can be improved in terms of both convergence rate and steady-state errors. Following [5], the problem can be decoupled into N subproblems, resulting in a distributed solution. For each node k, we minimize the cost function $\mathcal{L}_k = \mathbb{E}\{(y_k(n) - \boldsymbol{x}_k^T(n)\boldsymbol{\Psi}_k(n)\boldsymbol{c}_k(n))^2 + \nu_k\mathcal{J}_k(n)\}$ with respect to $\boldsymbol{c}_k(n)$.

Sparsity-aware adaptive combiners: Recalling that there are variations in the regressor data across the distributed nodes because of spatial diversity, the number of links between nodes can be reduced. There are three main reasons for selecting a subset of $\mathcal{N}_k$, defined by $\Omega_k \subseteq \mathcal{N}_k$: i) less local noise is added through the combination step, ii) the computation load can be reduced, iii) nodes with high SNRs can help the network achieve faster convergence rate and lower steady-state MSD. Mathematically, the optimization problem is

$$\begin{aligned} \min_{\Omega_k} \quad & \mathbb{E}\{(y_k(n) - \boldsymbol{x}_k^T(n)\boldsymbol{\Psi}_k(n)\boldsymbol{c}_k(n))^2 + \nu_k\mathcal{J}_k(n)\} \\ \text{s.t.} \quad & c_{i,k}(n) > 0, \sum_{i\in\Omega_k} c_{i,k}(n) = 1, c_{i,k}(n) = 0 \text{ for } i \notin \Omega_k. \end{aligned} \tag{2}$$

where the convex constraints (2) are necessary conditions to guarantee convergence and unbiased estimates [1]. In order to find the optimized subset Ω_k, we need to determine the non-zero entries of $\boldsymbol{c}_k(n)$. Consequently, we define the instantaneous error after the local adaptive filter at the kth node as

$$e_k(n) = y_k(n) - (\boldsymbol{\Psi}_k(n)\boldsymbol{c}_k(n))^T\boldsymbol{x}_k(n) = y_k(n) - \boldsymbol{c}_k(n)^T\widetilde{\boldsymbol{x}}_k(n) \tag{3}$$

where

$$\widetilde{\boldsymbol{x}}_k(n) \triangleq \boldsymbol{\Psi}_k^T(n)\boldsymbol{x}_k(n) = [\widehat{y}_{1,k}(n), \ldots, \widehat{y}_{i,k}(n), \ldots, \widehat{y}_{N_k,k}(n)]^T, \tag{4}$$

and $\widehat{y}_{i,k}(n) = \boldsymbol{\psi}_i^T(n)\boldsymbol{x}_k(n)$, $i \in \mathcal{N}_k$ represent the estimates of $y_k(n)$ relative to the kth node. Since $\boldsymbol{c}_k^T \mathbb{1}_{N_k} = 1$ where $\mathbb{1}_N$ represents the $N \times 1$ vector whose entries are all one, the instantaneous error can be expressed further by

$$e_k(n) = \boldsymbol{c}_k^T(n)(\mathbb{1}_{N_k}y_k(n) - \widetilde{\boldsymbol{x}}_k(n)) = \sum_{i\in\mathcal{N}_k} c_{i,k}(n)v_{i,k}(n), \tag{5}$$

where $v_{i,k}(n)$ denotes the individual error resulting from applying the estimate $\boldsymbol{\psi}_i(n)$ from the ith node to the data at the kth node. Since only the relative values of $c_{i,k}(n)$ are the parameters of interest, we can simplify the optimization if we i) ignore the constraints, ii) obtain the combination coefficient vector $\boldsymbol{c}_k(n)$ by the proposed combination approach as (7) or (9) below, iii) take $\max\{c_{i,k}(n), 0\}$, $i \in \mathcal{N}_k$, and normalize them each time such that they satisfy (2), iv) eliminate the weights whose entries $c_{i,k}(n)$ are very small in the combination, where the threshold is defined as β_k, v) normalize $c_{i,k}(n)$ again.

ZA-DLMS combiner: In the combination step a new cost function is defined by minimizing the received instantaneous errors with an l_1-norm

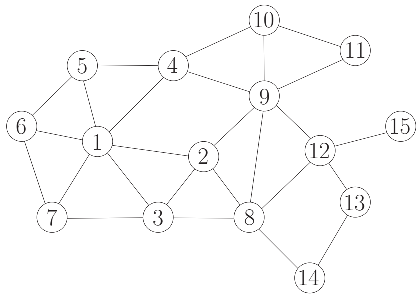

Fig. 1. Topology of the distributed network.

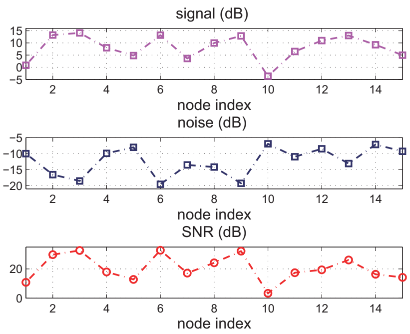

Fig. 2 *(Top) received signal variances $\sigma^2_{\boldsymbol{x},k}$. (Middle) local noise variances σ^2_k. (Bottom) SNR at each node.*

penalty on the coefficient vector, which is expressed by

$$\mathcal{L}_k^{\mathrm{ZA}}(n) = \frac{1}{2}e_k^2(n) + \gamma_k^{\mathrm{ZA}}\|\boldsymbol{c}_k(n)\|_1 \tag{6}$$

where γ_k^{ZA} is the regulation parameter that controls a tradeoff between the residual error and sparsity of the combination coefficients. According to (3) and (4), $\widetilde{\boldsymbol{x}}_k(n)$ can be considered as a new input vector. Using gradient descent updating, the ZA-DLMS filter update is obtained by

$$\begin{aligned}\boldsymbol{c}_k(n+1) &= \boldsymbol{c}_k(n) - \mu_k^{\mathrm{ZA}}\frac{\partial\mathcal{L}_k^{\mathrm{ZA}}(n)}{\partial\boldsymbol{c}_k(n)}\\ &= \boldsymbol{c}_k(n) + \mu_k^{\mathrm{ZA}}e_k(n)\widetilde{\boldsymbol{x}}_k(n) - \rho_k^{\mathrm{ZA}}\mathrm{sgn}(\boldsymbol{c}_k(n))\end{aligned} \tag{7}$$

where the $\rho_k^{\mathrm{ZA}} = \mu_k^{\mathrm{ZA}}\gamma_k^{\mathrm{ZA}}$ and $\mathrm{sgn}(\cdot)$ is a component-wise sign function defined as $\mathrm{sgn}(x) = x/|x|, x \neq 0$ and $\mathrm{sgn}(x) = 0, x = 0$. μ_k^{ZA} is the step-size of the filter at node k. The complexity of the proposed ZA-DLMS is still as low as that of LMS, $\mathcal{O}(M)$.

JZA-DLMS combiner: After combining the individual weights, the updated weight at the kth node may add residual errors in the combined weight. Therefore, taking the sparsity of the system into consideration, the l_1-norm penalty still must be imposed on the combined weight $\boldsymbol{h}_k(n)$, resulting in the cost function $\mathcal{L}_k^{\mathrm{JZA}}(n)$, i.e.,

$$\mathcal{L}_k^{\mathrm{JZA}}(n) = \frac{1}{2}e_k^2(n) + \underbrace{\gamma_k^{\mathrm{JZA}}\|\boldsymbol{c}_k(n)\|_1}_{\text{spatial sparsity}} + \underbrace{\bar{\gamma}_k^{\mathrm{JZA}}\|\boldsymbol{\Psi}_k(n)\boldsymbol{c}_k(n)\|_1}_{\text{system sparsity}} \tag{8}$$

where the two regulation parameters γ_k^{JZA} and $\bar{\gamma}_k^{\mathrm{JZA}}$ control the balance between error and sparsity. Similarly, the updating equation of the adaptive combines for the JZA-DLMS algorithm is obtained by

$$\begin{aligned}\boldsymbol{c}_k(n+1) &= \boldsymbol{c}_k(n) - \mu_k^{\mathrm{JZA}}\frac{\partial\mathcal{L}_k^{\mathrm{JZA}}(n)}{\partial\boldsymbol{c}_k(n)}\\ &= \boldsymbol{c}_k(n) + \mu_k^{\mathrm{JZA}}e_k(n)\widetilde{\boldsymbol{x}}_k(n) - \rho_k^{\mathrm{JZA}}\mathrm{sgn}(\boldsymbol{c}_k(n))\\ &\quad - \bar{\rho}_k^{\mathrm{JZA}}\boldsymbol{\Psi}_k^T(n)\mathrm{sgn}\big(\boldsymbol{\Psi}_k(n)\boldsymbol{c}_k(n)\big)\end{aligned} \tag{9}$$

where $\rho_k^{\mathrm{JZA}} = \mu_k^{\mathrm{JZA}}\gamma_k^{\mathrm{JZA}}$ and $\bar{\rho}_k^{\mathrm{JZA}} = \mu_k^{\mathrm{JZA}}\bar{\gamma}_k^{\mathrm{JZA}}$.

The proposed JZA-DLMS algorithm considers jointly the sparsity both of the network links and of the optimal weight vector such that the quality of estimation is improved further with the prior information about sparsity.

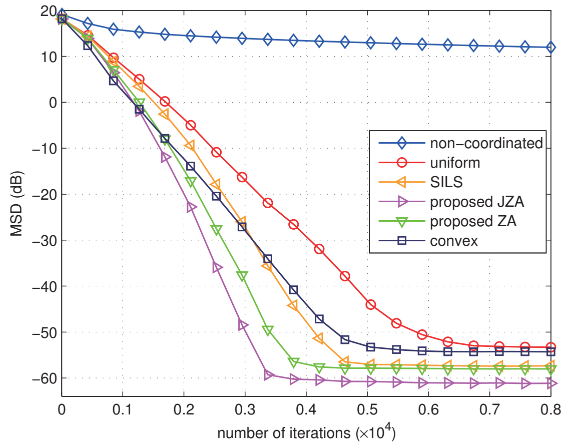

Fig. 3 *Comparison of the different combination strategies in terms of the network MSD. The step size and sparsity parameters are as follows for $k = 1, \ldots, 15$. Local LMS, $\mu_k = 1\times10^{-4}$, $\nu_k = 1\times10^{-6}$. Combiner, $\mu_k^{Convex} = \mu_k^{ZA} = \mu_k^{JZA} = 1\times10^{-4}$, $\rho_k^{JZA} = 1\times10^{-6}$, $\bar{\rho}_k^{JZA} = 1\times10^{-6}$, $\beta_k = 0.05$, and $\rho_k^{SILS} = 5\times10^{-3}$, $\varepsilon_k^{SILS} = 10$.*

Simulation results: The topology of the network is considered as shown in Fig. 1, where the link combination algorithms (uniform [4], convex [5], SILS [6] and the proposed ones) are applied at each node. The input signal and the observed noise are both white Gaussian random sequences with zero mean, where the covariance matrices are $\boldsymbol{R}_{\boldsymbol{x},k} = \sigma^2_{\boldsymbol{x},k}\boldsymbol{I}_M$ and noise variances are σ_k^2 respectively, where $k = 1, \ldots, 15$. It can be observed in Fig. 2 that the 2nd, 3rd, 6th and 9th node have relative higher SNRs, which illustrates the sparsity in spatial domain. We simulate a 200-tap regressor with 10 non-zero coefficients at the first 10 taps in Fig. 3, with a sparsity ratio of $10/200$. Based on 100 Monte Carlo trials, the proposed ZA-DLMS and JZA-DLMS show faster convergence rate and lower steady-state MSD than the other approaches, where the JZA-DLMS performs better than ZA-DLMS since it still considers the sparsity of the system.

Conclusion: In this paper, we proposed a sparsity-aware adaptive combination strategy for sparse system learning. The proposed ZA-DLMS and JZA-DLMS adaptively combine the individual nodes' weights by the l_1-norm penalization such that the network achieves better performance with faster convergence rate and lower steady-state MSD. Simulation results showed sparsity-aware algorithms are superior to the traditional ones in distributed sparse system learning.

Songtao Lu (*Department of Electrical and Computer Engineering, Iowa State University, USA*)

Vítor H. Nascimento (*Department of Electronic Systems Engineering, University of São Paulo, Brazil*)

E-mail: vitor@lps.usp.br

Jinping Sun (*School of Electronic and Information Engineering, Beihang University, China*)

Zhuangji Wang (*Department of Agronomy, Iowa State University, USA*)

References

1 Sayed, A., Tu S.-Y., Chen J., Zhao X., and Towfic Z.: 'Diffusion strategies for adaptation and learning over networks: an examination of distributed strategies and network behavior', *IEEE Signal Process. Mag.*, 2013, **30**, p. 155-171

2 Lopes, C. and Sayed A.: 'Diffusion least-mean squares over adaptive networks: Formulation and performance analysis', *IEEE Trans. Signal Process.*, 2008, **56**, pp. 3122-3136

3 Chen Y., Gu Y., and Hero A.: 'Sparse LMS for system identification', *Proc. of IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2009, pp. 3125-3128

4 Di Lorenzo, P. and Sayed A.: 'Sparse distributed learning based on diffusion adaptation', *IEEE Trans. Signal Process.*, 2013, **61**, pp. 1419-1433

5 Takahashi, N., Yamada I., and Sayed A.: 'Diffusion least-mean squares with adaptive combiners: formulation and performance analysis', *IEEE Trans. Signal Process.*, 2010, **58**, pp. 4795-4810

6 Xu S., de Lamare R. C., and Poor H. V.: 'Adaptive link selection strategies for distributed estimation in diffusion wireless networks', *Proc. of IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2013, pp. 5402-5405