Transient Analysis of Clustered Multitask Diffusion RLS Algorithm

Wei Gao † Jie Chen ⋆ Cédric Richard ‡ Wentao Shi ⋆ Qunfei Zhang ⋆

†Jiangsu University, Zhenjiang, China ⋆Northwestern Polytechnical University, Xi′an, China ‡Université Côte d'Azur, Nice, France

Abstract

We propose a novel clustered multitask diffusion RLS (MT-DRLS) algorithm over network.

- Further improve the performance of the multitask diffusion LMS (MT-DLMS) algorithm.
- Its transient behavior is investigated, in the mean and mean-square sense.
- Simulation results illustrate the significant improvement of the MT-DRLS over the MT-DLMS, as well as the accuracy of the theoretical findings.

Motivation

- The diffusion recursive least-squares (DRLS) algorithm and its steady-state performance were extensively studied in the literature, due to the superior performance of the RLS compared to the LMS.
- More recently, a transient analysis of DRLS algorithm was presented in [1].
- To the best of our knowledge, the multitask DRLS algorithm has not been considered so far except in [2], where the transient analysis of DRLS algorithm was not studied.
- This motivates us to derive the clustered MT-DRLS algorithm with adapt-then-combine (ATC) diffusion strategy.
- Furthermore, analytical models are derived to characterize its transient behavior in the mean and mean-square sense.

Clustered MT-DRLS Algorithm

Consider a connected network consisting of K nodes, indexed with $k = 1, \ldots, K$. Every node k has access to a random data pair $\{d_{k,n}, \mathbf{x}_{k,n}\}$, which is assumed to be generated by a linear regression model:

$$d_{k,n} = \mathbf{x}_{k,n}^\top \mathbf{w}_k^\star + z_{k,n}.$$

The unknown optimal weight vector $\mathbf{w}_k^\star$ are constrained to be identical within each cluster, namely, $\mathbf{w}_k^\star = \mathbf{w}_{\mathcal{C}_q}^\star$ for $\forall k \in \mathcal{C}_q$. In the context of clustered multitask networks, the objective is to estimate the unknown vectors $\{\mathbf{w}_{\mathcal{C}_q}^\star\}_{q=1}^{Q}$. By solving two least-squares problems, the proposed MT-DRLS algorithm with ATC diffusion strategy for clustered multitask networks is given by

$$\mathbf{P}_{k,n} = \lambda^{-1}\left(\mathbf{P}_{k,n-1} - \frac{\mathbf{P}_{k,n-1}\mathbf{x}_{k,n}\mathbf{x}_{k,n}^\top\mathbf{P}_{k,n-1}}{\lambda + \mathbf{x}_{k,n}^\top\mathbf{P}_{k,n-1}\mathbf{x}_{k,n}}\right)$$

$$\boldsymbol{\psi}_{k,n} = \mathbf{w}_{k,n-1} + \mathbf{P}_{k,n}\mathbf{x}_{k,n}e_{k,n} - \gamma \sum_{\ell \in \mathcal{N}_k \backslash \mathcal{C}(k)^-} \frac{\rho_{k\ell} + \rho_{\ell k}}{2}(\mathbf{w}_{k,n-1} - \mathbf{w}_{\ell,n-1})$$

$$\mathbf{w}_{k,n} = \sum_{\ell \in \mathcal{N}_k \cap \mathcal{C}k} a_{\ell k}\boldsymbol{\psi}_{\ell,n}$$

with

$$\sum_{\ell \in \mathcal{N}_k \backslash \mathcal{C}(k)^-} \rho_{k\ell} = 1, \text{ and } \begin{cases} \rho_{k\ell} > 0, & \text{if } \ell \in \mathcal{N}_k \setminus \mathcal{C}(k), \\ \rho_{kk} \geq 0, & \\ \rho_{k\ell} = 0, & \text{otherwise} \end{cases}$$

$$\sum_{\ell \in \mathcal{N}_k \cap \mathcal{C}(k)} a_{\ell k} = 1, \text{ and } \begin{cases} a_{\ell k} > 0, & \text{if } \ell \in \mathcal{N}_k \cap \mathcal{C}(k), \\ a_{\ell k} = 0, & \text{otherwise.} \end{cases}$$

Transient Performance Analysis of MT-DRLS

1. Preliminaries

The weight error vectors for node k at instant n are defined respectively as follows:

$$\widetilde{\boldsymbol{\psi}}_{k,n} \triangleq \boldsymbol{\psi}_{k,n} - \mathbf{w}_k^\star, \quad \widetilde{\mathbf{w}}_{k,n} \triangleq \mathbf{w}_{k,n} - \mathbf{w}_k^\star.$$

Let $\widetilde{\mathbf{w}}_n$ and $\mathbf{w}^\star$ denote the block weight error vector and the block optimal weight vector:

$$\widetilde{\mathbf{w}}_n \triangleq \text{col}\{\widetilde{\mathbf{w}}_{1,n}, \ldots, \widetilde{\mathbf{w}}_{K,n}\} \in \mathbb{R}^{KL}, \quad \mathbf{w}^\star \triangleq \text{col}\{\mathbf{w}_1^\star, \ldots, \mathbf{w}_K^\star\} \in \mathbb{R}^{KL}.$$

We also introduce the following required $K \times K$ block diagonal matrices defined as:

$$\mathbf{R}_{x,n} \triangleq \text{bdiag}\{\mathbf{x}_{1,n}\mathbf{x}_{1,n}^\top, \ldots, \mathbf{x}_{K,n}\mathbf{x}_{K,n}^\top\} \in \mathbb{R}^{KL \times KL},$$
$$\boldsymbol{\Phi}_n \triangleq \text{bdiag}\{\boldsymbol{\Phi}_{1,n}, \ldots, \boldsymbol{\Phi}_{K,n}\} \in \mathbb{R}^{KL \times KL},$$
$$\mathbf{P}_n \triangleq \text{bdiag}\{\mathbf{P}_{1,n}, \ldots, \mathbf{P}_{K,n}\} \in \mathbb{R}^{KL \times KL},$$
$$\boldsymbol{\mathcal{A}} \triangleq \mathbf{A}^\top \otimes \mathbf{I}_L \in \mathbb{R}^{KL \times KL},$$

and the block column vector with individual entries of size $L \times 1$ defined as:

$$\mathbf{s}_{xz,n} \triangleq \text{col}\{z_{1,n}\mathbf{x}_{1,n}, \ldots, z_{K,n}\mathbf{x}_{K,n}\} \in \mathbb{R}^{KL}.$$

2. Mean Error Behavior Analysis

By starting with the *a priori* estimation error $e_{k,n} = z_{k,n} - \mathbf{x}_{k,n}^\top\widetilde{\mathbf{w}}_{k,n-1}$, we can obtain

$$\boldsymbol{\Phi}_n\boldsymbol{\mathcal{A}}^{-1}\widetilde{\mathbf{w}}_n = \left(\lambda\boldsymbol{\Phi}_{n-1} - \gamma\boldsymbol{\Phi}_n\mathbf{Q}\right)\widetilde{\mathbf{w}}_{n-1} - \gamma\boldsymbol{\Phi}_n\mathbf{Q}\,\mathbf{w}^\star + \mathbf{s}_{xz,n}.$$

Then, the recursive relation of mean weight error vector is given by

$$\mathbb{E}\{\widetilde{\mathbf{w}}_n\} = \boldsymbol{\mathcal{A}}\left(\lambda\mathbb{E}\{\boldsymbol{\Phi}_n\}^{-1}\mathbb{E}\{\boldsymbol{\Phi}_{n-1\}} - \gamma\mathbf{Q}\right)\mathbb{E}\{\widetilde{\mathbf{w}}_{n-1}\} - \gamma\boldsymbol{\mathcal{A}}\,\mathbf{Q}\,\mathbf{w}^\star$$

with

$$\mathbb{E}\{\boldsymbol{\Phi}_n\} = \lambda\mathbb{E}\{\boldsymbol{\Phi}_{n-1}\} + \mathbf{R}_x$$
$$\mathbf{Q} = \frac{1}{2}\left[\text{diag}\left\{(\boldsymbol{\Theta} + \boldsymbol{\Theta}^\top)\mathbf{1}_K\right\} - (\boldsymbol{\Theta} + \boldsymbol{\Theta}^\top)\right] \otimes \mathbf{I}_L$$

where $\boldsymbol{\Theta}$ is right-stochastic matrix with the (k, ℓ)-th entry $\rho_{k\ell}$, and $\mathbf{R}_x$ is the expectation of matrix $\mathbf{R}_{x,n}$, i.e., $\mathbf{R}_x = \mathbb{E}\{\mathbf{R}_{x,n}\} = \text{bdiag}\{\mathbf{R}_{x,1}, \ldots, \mathbf{R}_{x,K}\} \in \mathbb{R}^{KL \times KL}$.

3. Mean-Square Error Behavior Analysis

The network transient mean-square deviation (MSD) at time instant n is defined by

$$\text{MSD}_n = \text{tr}\{\widetilde{\mathbf{W}}_n\}/K.$$

In order to investigate the mean-square error behavior of MT-DRLS algorithm, our next aim is to determine the update equation of $\widetilde{\mathbf{W}}_n$. For mathematical tractability of analysis, we thus introduce the following necessary approximations:

$$\mathbf{T}_0 = \mathbb{E}\{\boldsymbol{\Phi}_n\boldsymbol{\mathcal{A}}^{-1}\widetilde{\mathbf{w}}_n\widetilde{\mathbf{w}}_n^\top(\boldsymbol{\mathcal{A}}^{-1})^\top\boldsymbol{\Phi}_n\} \approx \mathbb{E}\{\boldsymbol{\Phi}_n\}\boldsymbol{\mathcal{A}}^{-1}\widetilde{\mathbf{W}}_n(\boldsymbol{\mathcal{A}}^{-1})^\top\mathbb{E}\{\boldsymbol{\Phi}_n\},$$
$$\mathbf{T}_1 = \mathbb{E}\{\boldsymbol{\Phi}_{n-1}\widetilde{\mathbf{w}}_{n-1}\widetilde{\mathbf{w}}_{n-1}^\top\boldsymbol{\Phi}_{n-1}\} \approx \mathbb{E}\{\boldsymbol{\Phi}_{n-1}\}\widetilde{\mathbf{W}}_{n-1}\mathbb{E}\{\boldsymbol{\Phi}_{n-1}\},$$
$$\mathbf{T}_2 = \mathbb{E}\{\boldsymbol{\Phi}_n\mathbf{Q}\,\widetilde{\mathbf{w}}_{n-1}\widetilde{\mathbf{w}}_{n-1}^\top\mathbf{Q}^\top\boldsymbol{\Phi}_n\} \approx \mathbb{E}\{\boldsymbol{\Phi}_n\}\mathbf{Q}\,\widetilde{\mathbf{W}}_{n-1}\mathbf{Q}^\top\mathbb{E}\{\boldsymbol{\Phi}_n\},$$
$$\mathbf{T}_3 = \mathbb{E}\{\boldsymbol{\Phi}_n\mathbf{Q}\,\mathbf{w}^\star(\mathbf{w}^\star)^\top\mathbf{Q}^\top\boldsymbol{\Phi}_n\} \approx \mathbb{E}\{\boldsymbol{\Phi}_n\}\mathbf{Q}\,\mathbf{w}^\star(\mathbf{w}^\star)^\top\mathbf{Q}^\top\mathbb{E}\{\boldsymbol{\Phi}_n\},$$
$$\mathbf{T}_4 = \mathbb{E}\{\boldsymbol{\Phi}_{n-1}\widetilde{\mathbf{w}}_{n-1}\widetilde{\mathbf{w}}_{n-1}^\top\mathbf{Q}^\top\boldsymbol{\Phi}_n\} \approx \mathbb{E}\{\boldsymbol{\Phi}_{n-1}\}\widetilde{\mathbf{W}}_{n-1}\mathbf{Q}^\top\mathbb{E}\{\boldsymbol{\Phi}_n\},$$
$$\mathbf{T}_5 = \mathbb{E}\{\boldsymbol{\Phi}_{n-1}\widetilde{\mathbf{w}}_{n-1}(\mathbf{w}^\star)^\top\mathbf{Q}^\top\boldsymbol{\Phi}_n\} \approx \mathbb{E}\{\boldsymbol{\Phi}_{n-1}\}\mathbb{E}\{\widetilde{\mathbf{w}}_{n-1}\}(\mathbf{w}^\star)^\top\mathbf{Q}^\top\mathbb{E}\{\boldsymbol{\Phi}_n\},$$
$$\mathbf{T}_6 = \mathbb{E}\{\boldsymbol{\Phi}_n\mathbf{Q}\,\widetilde{\mathbf{w}}_{n-1}(\mathbf{w}^\star)^\top\mathbf{Q}^\top\boldsymbol{\Phi}_n\} \approx \mathbb{E}\{\boldsymbol{\Phi}_n\}\mathbf{Q}\,\mathbb{E}\{\widetilde{\mathbf{w}}_{n-1}\}(\mathbf{w}^\star)^\top\mathbf{Q}^\top\mathbb{E}\{\boldsymbol{\Phi}_n\}.$$

After some derivation steps, we finally arrive at the recursion of $\widetilde{\mathbf{W}}_n$ as follows:

$$\begin{aligned}\widetilde{\mathbf{W}}_n = \boldsymbol{\mathcal{A}}\Big[&\lambda^2\mathbb{E}\{\boldsymbol{\Phi}_n\}^{-1}\mathbb{E}\{\boldsymbol{\Phi}_{n-1}\}\widetilde{\mathbf{W}}_{n-1}\mathbb{E}\{\boldsymbol{\Phi}_{n-1}\}\mathbb{E}\{\boldsymbol{\Phi}_n\}^{-1} + \gamma^2\big(\mathbf{Q}\,\widetilde{\mathbf{W}}_{n-1}\mathbf{Q}^\top + \mathbf{Q}\,\mathbf{w}^\star(\mathbf{w}^\star)^\top\mathbf{Q}^\top\big)\\ &- \gamma\lambda\big(\mathbb{E}\{\boldsymbol{\Phi}_n\}^{-1}\mathbb{E}\{\boldsymbol{\Phi}_{n-1}\}\widetilde{\mathbf{W}}_{n-1}\mathbf{Q}^\top + \mathbf{Q}\,\widetilde{\mathbf{W}}_{n-1}^\top\mathbb{E}\{\boldsymbol{\Phi}_{n-1}\}\mathbb{E}\{\boldsymbol{\Phi}_n\}^{-1}\big)\\ &- \gamma\lambda\big(\mathbb{E}\{\boldsymbol{\Phi}_n\}^{-1}\mathbb{E}\{\boldsymbol{\Phi}_{n-1}\}\mathbb{E}\{\widetilde{\mathbf{w}}_{n-1}\}(\mathbf{w}^\star)^\top\mathbf{Q}^\top + \mathbf{Q}\,\mathbf{w}^\star\mathbb{E}\{\widetilde{\mathbf{w}}_{n-1}\}^\top\mathbb{E}\{\boldsymbol{\Phi}_{n-1}\}\mathbb{E}\{\boldsymbol{\Phi}_n\}^{-1}\big)\\ &+ \gamma^2\big(\mathbf{Q}\,\mathbb{E}\{\widetilde{\mathbf{w}}_{n-1}\}(\mathbf{w}^\star)^\top\mathbf{Q}^\top + \mathbf{Q}\,\mathbf{w}^\star\mathbb{E}\{\widetilde{\mathbf{w}}_{n-1}\}^\top\mathbf{Q}^\top\big) + \mathbb{E}\{\boldsymbol{\Phi}_n\}^{-1}\boldsymbol{\Sigma}_z\mathbf{R}_x\mathbb{E}\{\boldsymbol{\Phi}_n\}^{-1}\Big]\boldsymbol{\mathcal{A}}^\top\end{aligned}$$

with the block diagonal matrix $\boldsymbol{\Sigma}_z = \text{bdiag}\{\sigma_{z,1}^2\mathbf{I}_L, \ldots, \sigma_{z,K}^2\mathbf{I}_L\}$. We can characterize the transient mean-square errors of clustered MT-DRLS by the above recursive relation.

Numerical Tests

We consider a connected network consisting of 14 nodes grouped into 3 clusters as shown in Fig. 1 (Left). The optimal weight vectors to be estimated in each cluster are $\mathbf{w}_{\mathcal{C}_1}^\star = [0.5196, -0.3667]^\top$, $\mathbf{w}_{\mathcal{C}_2}^\star = [0.4952, -0.3783]^\top$, and $\mathbf{w}_{\mathcal{C}_3}^\star = [0.4951, -0.4079]^\top$, respectively. The variances $\sigma_{x,k}^2$ and $\sigma_{z,k}^2$ are depicted in Fig. 1 (Right), respectively.

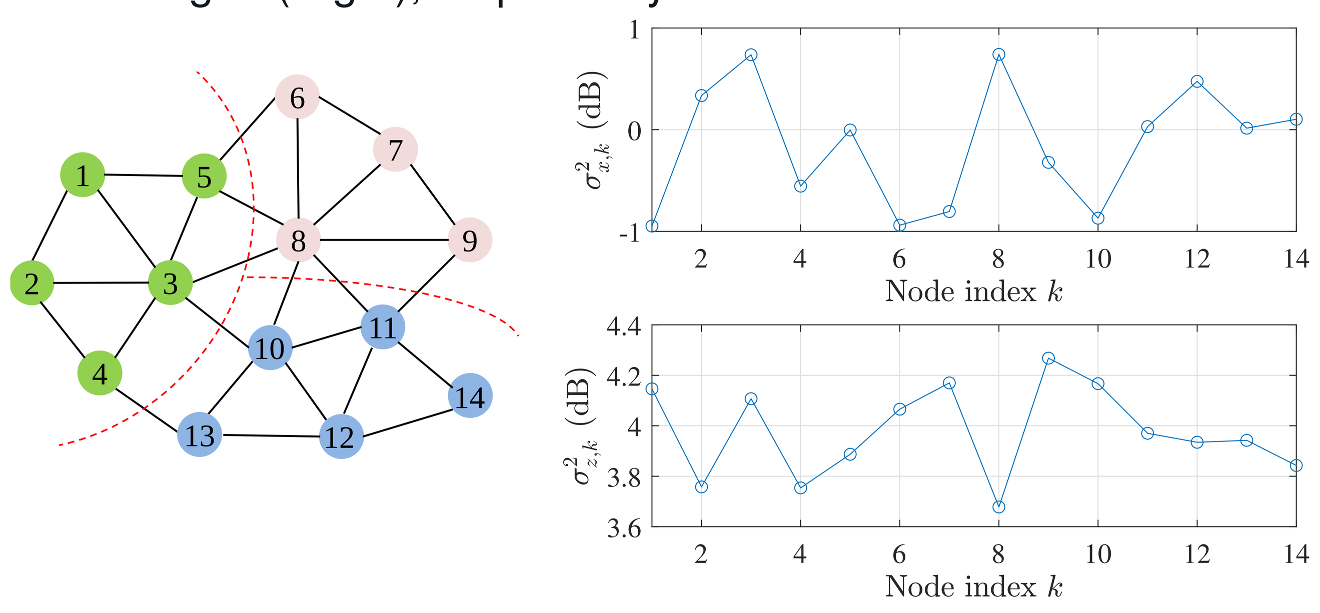

Fig. 1: Network topology (Left). Input and noise variances (Right).

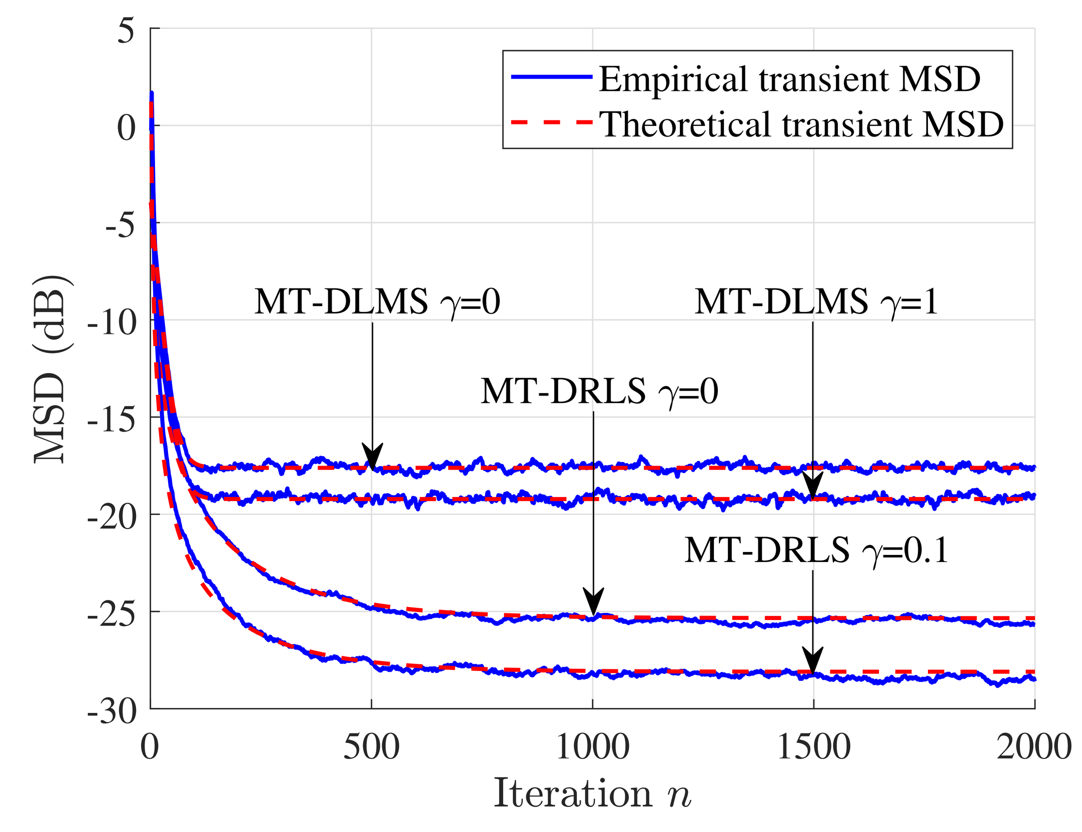

Fig. 2: Empirical vs. theoretical network MSD.

- Fig. 2 shows that the clustered MT-DRLS algorithms significantly outperforms the counterpart clustered MT-DLMS algorithms in terms of convergence rate, steady-state errors, and parameter estimation accuracy.
- We can also see that the consistent agreement between empirical and theoretical MSD curves validates the accuracy and effectiveness of transient theoretical analysis for clustered MT-DRLS algorithm.
- The good consistency also verified that all the necessary approximations introduced in the analysis are correct and reasonable.

References

[1] W. Gao, J. Chen, and C. Richard, "Transient theoretical analysis of diffusion RLS algorithm for cyclostationary colored inputs," IEEE Signal Process. Lett., vol. 28, pp. 1160–1164, 2021.
[2] X. Cao and K. J. R. Liu, "Decentralized sparse multitask RLS over networks," IEEE Trans. on Signal Process., vol. 65, no. 23, pp. 6217–6232, Dec. 2017.
[3] 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.

Contact Information: wei_gao@ujs.edu.cn