Distributed Variance Consensus with Application to Personalized Learning

Diego Deplano, Nicola Bastianello, Mauro Franceschelli, and Karl H. Johansson

ABSTRACT

This paper addresses the problem of computing the sample variance of datasets scattered across a network of interconnected agents. A general procedure is outlined to allow the agents to reach consensus on the variance of their local data, which involves two cascaded (dynamic) average consensus protocols. Our implementation of the procedure exploits the distributed ADMM, yielding a distributed protocol that does not involve the sharing of any local, private data nor any coordination of a central authority; the algorithm is proved to be convergent with linear rate and null steady-state error. The proposed distributed variance estimation scheme is then leveraged to tune personalization in "personalized learning" where agents aim at training a local model tailored to their own data, while still benefiting from the cooperation with other agents to enhance the models' generalization power. The degree to which an agent tailors its local model depends on the diversity of the local datasets, and we propose to use the variance to tune personalization. Numerical simulations test the proposed approach in a classification task of handwritten digits, drawn from the EMNIST dataset, showing the better performance of variance-tuned personalization over non-personalized training.

PROBLEM DEFINITION

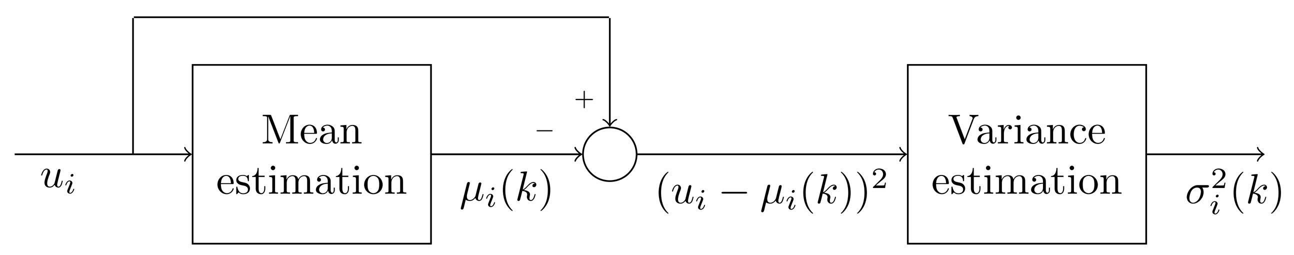
Consider a network of $n \in \mathbb{N}$ agents interacting according to an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and having access to some local data $u_i \in \mathbb{R}$. The consensus problem consists in the design of proper local interaction rules to enable the estimation of a function of the local data $u_i \in \mathbb{R}$ stacked into $\boldsymbol{u} \in \mathbb{R}^n$. We are interested in the sample variance σ^2 and average μ of all data:

$$\sigma^2 = \text{var}(\boldsymbol{u}) := \frac{1}{n}\sum_{i=1}^{n}\left(u_i - \mu\right)^2, \quad \mu = \text{avg}(\boldsymbol{u}) := \frac{1}{n}\sum_{i=1}^{n} u_i.$$

Let $s_i(k) \in \mathbb{R}$ be the signal of interest, the above consensus problems can be cast into a (time-varying) distributed optimization problem of the kind:

$$\boldsymbol{x}^\star(k) = \underset{x_1,\dots,x_n}{\text{argmin}} \sum_{i=1}^{n}\frac{1}{2}|x_i - s_i(k)|^2, \qquad \text{s.t.} \quad x_i = x_j \quad \forall (i,j) \in \mathcal{E},$$

where $s_i(k) = u_i$ for the average of the reference signals and $s_i(k) = (u_i - \mu)$ for their variance. A solution to this problem can be found by two cascaded (dynamic) average consensus protocols as shown next, where $\mu(k)_i$, $\sigma_i(k)$ denote the average and variance estimations of node $i \in \mathcal{V}$:



This strategy can be implemented by using any dynamic average consensus algorithms. We chose to use DOT-ADMM [R1], [R2], whose performance against other state-of-the-art dynamic average consensus protocols is compared in [R3].

SOLUTION VIA DOT-ADMM [R1] [R2]

Algorithm 1

Initialization: Each agent $i \in \mathcal{V}$ arbitrarily initializes the auxiliary variables $\{y_{ij}(0), z_{ij}(0)\}_{j\in\mathcal{N}_i}$, and parameters $\alpha \in (0,1)$, $\rho > 0$.

for $k = 1, 2, \dots$ **each active agent** $i \in \mathcal{V}$

applies the local updates

$$\mu_i(k) = \frac{u_i + \sum_{j\in\mathcal{N}_i} y_{ij}(k-1)}{1 + \rho\eta_i}$$

$$\sigma_i^2(k) = \frac{(u_i - \mu_i(k))^2 + \sum_{j\in\mathcal{N}_i} z_{ij}(k-1)}{1 + \rho\eta_i}$$

for each agent $j \in \mathcal{N}_i$

transmits the packets

$$p_{i\to j}(k) = 2\rho\mu_i(k) - y_{ij}(k-1)$$
$$q_{i\to j}(k) = 2\rho\sigma_i^2(k) - z_{ij}(k-1)$$

for each neighbor $j \in \mathcal{N}_i$

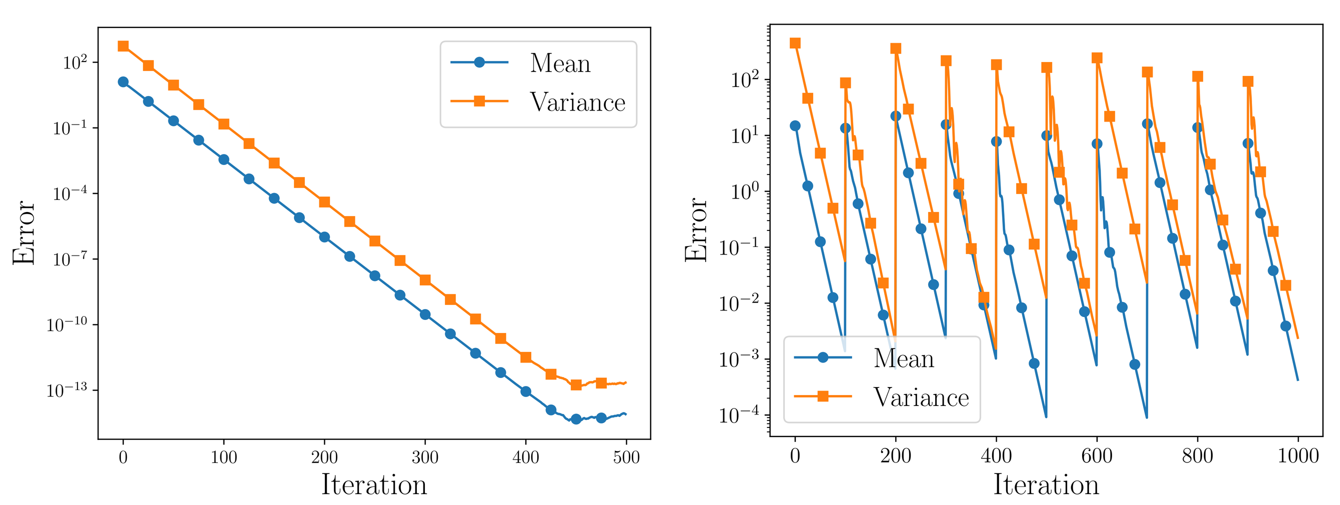
updates the auxiliary variables

$$y_{ij}(k) = (1-\alpha)y_{ij}(k-1) + \alpha p_{j\to i}(k)$$
$$z_{ij}(k) = (1-\alpha)z_{ij}(k-1) + \alpha q_{j\to i}(k)$$

Theorem 3: Consider a network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ executing Algorithm 1 with local data $u_i \in \mathbb{R}$. If the graph $\mathcal{G}$ is connected, the root mean squared errors on the estimation of the mean and variance, respectively, converge to zero with a linear rate.

DOWNLOAD ME!

NUMERICAL SIMULATIONS



Error trajectories of Algorithm 1 on a random geometric graph with $n = 25$ nodes: (left) constant data u_i; (right) time-varying data $u_i(k)$.

PERSONALIZED LEARNING

In distributed personalized learning, each agent $i \in \mathcal{V}$ holds a local data set $\{\boldsymbol{a}_{i,h}, \boldsymbol{b}_{i,h}\}_{h=1}^{m_i}$ of dimension $m_i \in \mathbb{N}$ and aims at estimating the parameters of the common regression model g such that the sum of local partial costs f_i is minimized:

$$f_i(\boldsymbol{x}) = \sum_{h=1}^{m_i} g(\boldsymbol{x}, \boldsymbol{a}_{i,h}, \boldsymbol{b}_{i,h}).$$

In many applications there may be an interest in providing personalization: the model trained by each agent is tailored according to the specific data that it stores. The method we chose to implement personalization is by regularizing the costs as follows:

$$(\boldsymbol{x}^\star_{\text{PER}}, \boldsymbol{y}^\star_{1,\text{PER}}, \cdots, \boldsymbol{y}^\star_{n,\text{PER}}) \min_{\boldsymbol{x}_1,\dots,\boldsymbol{x}_n,\boldsymbol{y}_1,\dots,\boldsymbol{y}_n} \sum_{i\in\mathcal{V}} f_i(\boldsymbol{x}) + \frac{\lambda_i}{2}\|\boldsymbol{y}_i - \boldsymbol{x}_i\|^2, \qquad \text{s.t.} \quad \boldsymbol{x}_i = \boldsymbol{x}_j \text{ if } (i,j) \in \mathcal{E}.$$

This novel problem formulation uses the regularization terms to keep locally trained models $\boldsymbol{y}^\star_{i,\text{PER}}$ close to the global model $\boldsymbol{x}^\star_{\text{PER}}$, which are computed simultaneously. The personalization is tuned by the personalized weights $\lambda_i \in [0, \infty]$:

- Complete personalization: if $\lambda_i = 0$, then agent i only relies on its own data to train its local model.
- No personalization: if $\lambda_i = \infty$, then agent i fully collaborates in order to train a common model with other agents.
- Partial personalization: if $\lambda_i \in (0, \infty)$, then agent i partially collaborates with other agents to train both a common model and a personalized model.

CHOICE OF THE PERSONALIZED WEIGHTS

The strategy we propose is outlined in three main steps:

1) **Local training**: Each agent independently trains its own model $\boldsymbol{y}^\star_{i,\text{LOC}} = \text{argmin}_{\boldsymbol{y}} f_i(\boldsymbol{y})$ without considering the models of other nodes.

2) **Model distribution analysis**: The agents collaboratively estimate the component-wise mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\sigma}^2$ of their local models to assess their alignment or divergence.

3) **Weight selection**: Each node personalizes its model by integrating knowledge from others, adjusting the regularization weight λ_i based on how far its local model $\boldsymbol{y}^\star_{i,\text{LOC}}$ deviates from the average $\boldsymbol{\mu}$, relative to the standard deviation $\boldsymbol{\sigma}$. In particular, we let:

$$\lambda_i = \frac{w}{\text{avg}(\max\{\boldsymbol{0}, |\boldsymbol{y}^\star_{i,\text{LOC}} - \boldsymbol{\mu}| - \boldsymbol{\sigma}\})},$$

A larger deviation results in a smaller λ_i, promoting more personalization, while smaller deviations (within one standard deviation) yield $\lambda_i = \infty$, meaning no personalization is applied and the agent adopts the global model $\boldsymbol{x}^\star_{\text{PER}}$.

RESULTS ON THE EMNIST DATASET

We simulate a binary classification task between handwritten digits 3 and 8 using the EMNIST dataset, which groups characters by writer. To ensure a solid set-up, we select writers with at least 15 training and 5 test samples, resulting in a total of $n = 40$ writers. The model used is a standard linear regression. The accuracy results of our simulations are displayed in the next table for different levels of personalization.

	Param.	Model	Min	Mean	Max
Global model without personalization		$\boldsymbol{x}^\star_{\text{DIS}}$	**98.1%**	98.1%	98.1%
Global model with personalization	$\forall w \geq 0$	$\boldsymbol{x}^\star_{\text{PER}}$	**98.1%**	98.1%	98.1%
Local model with personalization	$w = 10$	$\boldsymbol{y}^\star_{i,\text{PER}}$	97.6%	98.1%	98.6%
Local model with personalization	$w = 1$	$\boldsymbol{y}^\star_{i,\text{PER}}$	96.2%	**98.3%**	**99.5%**
Local model with personalization	$w = 0.1$	$\boldsymbol{y}^\star_{i,\text{PER}}$	94.3%	98.1%	**100%**
Local model without cooperation		$\boldsymbol{y}^\star_{i,\text{LOC}}$	59.5%	85.0%	94.3%

Results show that agents benefit from the personalized cooperative learning approach because it maintains or improves expected accuracy (potentially reaching 100%) while still enabling the computation of a global model with comparable performance.

BIBLIOGRAPHY AND AFFILIATIONS

[R1] Bastianello, N. and Carli, R. (2022). ADMM for dynamic average consensus over imperfect networks. In IFAC–PapersOnLine, volume 55, 228–233.

[R2] Bastianello, N., Deplano, D., Franceschelli, M., and Johansson, K.H., "Robust online learning over networks". IEEE Transactions on Automatic Control (2025).

[R3] Deplano, D., Bastianello, N., Franceschelli, M., and Johansson, K.H., "A unified approach to solve the dynamic consensus on the average, maximum, and median values with linear convergence". In IEEE 62nd Conference on Decision and Control (2023b).

D. Deplano and M. Franceschelli are with DIEE, University of Cagliari, 09123 Cagliari, Italy. Emails: {diego.deplano,mauro.franceschelli}@unica.it

N. Bastianello and K. H. Johansson are with the School of Electrical Engineering and Computer Science, and Digital Futures, KTH Royal Institute of Technology, Stockholm, Sweden. Emails: {nicolba,kallej}@kth.se