Distributed Event-Based Learning via ADMM

Guener Dilsad ER [1] Sebastian Trimpe [2] Michael Muehlebach [1]

[1] Max Planck Institute for Intelligent Systems [2] RWTH Aachen University

ICML International Conference On Machine Learning

Summary

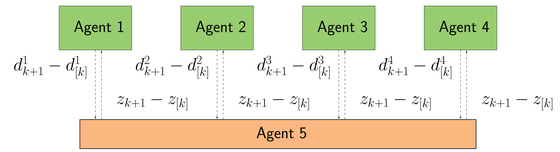
- **Objective:** Develop a **communication-efficient** distributed optimization algorithm that handles **non-i.i.d.** data distribution across agents.
- **Challenge:** High communication cost and convergence with non-i.i.d. data.
- **Solution:** An event-based communication strategy that reduces unnecessary data exchanges while maintaining model accuracy.
- **Theoretical Analysis:** Explores the trade-off between communication efficiency, model accuracy, and convergence speed.
- **Empirical Results:** Over 30% reduction in comm. cost while preserving high classification accuracy.
- **Impact:** Scalable, robust solution for large-scale distributed optimization.

We consider a **distributed optimization problem** where multiple agents with local datasets collaborate to minimize a global objective function, represented as the sum of their individual objectives. The **non-identically distributed (non-i.i.d.) data** across agents poses challenges for global optimality, requiring communication to ensure convergence. The problem is formulated as a consensus problem:

$$\min_{x^1,\dots,x^N\in\mathbb{R}^n} \sum_{i=1}^{N} f^i(x^i) + g(z), \quad \text{subject to } x^i = z, \quad \forall i = 1,\dots,N,$$

where $g(z)$ represents nonsmooth part of the objective function, and z is the global consensus variable.

Algorithm



Algorithm 1 Event-Based Distributed Learning with Over-Relaxed ADMM

Require: Local objective functions f^i, parameters ρ, Δ^d, Δ^z, reset period T

Require: Initialize $\hat{x}^i_0 = x_0$, $\hat{z}_0 = \zeta_0 = x_0$, $\hat{u}^i_{-1} = u^i_0$

for $k = 0$ to $t_{\max}$ **do**

for $i = 1$ to N **do**

$\hat{z}^i_k \leftarrow$ receive $z_k - z_{[k-1]}$ {Agent i}

$u^i_k = u^i_{k-1} + \alpha x^i_k - \hat{z}^i_k + (1-\alpha)z^i_{k-1}$

$x^i_{k+1} = \arg\min_{x^i} f^i(x^i) + \frac{\rho}{2}|x^i - z^i_k + u^i_k|^2$ {-> Local Training}

event-based send of $d^i_{k+1} - d^i_{[k]}$ $\{|d^i_{k+1} - d^i_{[k]}| > \Delta^d\}$

end for

$\hat{\zeta}_k \leftarrow$ receive $\frac{1}{N}\sum_{i\in\mathcal{C}^d_{k+1}}(d^i_{k+1} - d^i_{[k]})$ {Agent $N+1$}

$z_{k+1} = \arg\min_z g(z) + \frac{N\rho}{2}|z - \hat{\zeta}_k - (1-\alpha)z_k|^2$ {-> Global Aggregation}

event-based send of $z_{k+1} - z_{[k]}$ $\{|z_{k+1} - z_{[k]}| > \Delta^z\}$

if $\mathrm{mod}(k+1, T) = 0$ **then**

perform reset $\rightarrow \hat{\zeta}_k = \zeta_k$, $\hat{z}_k = z_k$

end if

end for

Theorem (strongly convex and smooth objective -> linear convergence)

Let $f = \sum_{i=1}^N f^i$ be m-**strongly convex and** L-**smooth** with $\kappa = L/m$, and g be convex. Let the step-size be $\rho = (mL)^{\frac{1}{2}}\kappa^{\epsilon}$ with $\epsilon \in [0, \infty)$, and $\alpha = 1$. For large enough κ, we have

$$|z_k - z_*|^2 \le 4\left(1 - \frac{1}{4\,\kappa^{\epsilon+\frac{1}{2}}}\right)^{2k} D_0 + \frac{5}{N}\kappa^{2+2\epsilon}\Delta^2,$$

where z_* is the optimal value for the consensus variable z, and D_0 represents the initial error, $D_0 = |z_0 - z_*|^2 + \frac{1}{N}\sum_{i=1}^N |u^i_0 - u^i_*|^2$, with u^i_* denoting the optimal values of the dual variables associated with each agent. Here, $\Delta = N\Delta^d + \Delta^z + T(N\bar{\chi}^d + \bar{\chi}^z)$ captures the error arising from the event-based communication.

Theorem (nonconvex objective -> sublinear convergence)

Let each $f^i : \mathbb{R}^n \to \mathbb{R}$ be **smooth (potentially nonconvex)** and let $g : \mathbb{R}^n \to \bar{\mathbb{R}}$ be a proper, closed convex function. Let the relaxation parameter be $\alpha = 1$, and the communication threshold Δ_k decay as $\Delta_k = \Delta_0/(k+1)^2$. Then, the gradients and residuals converge with a rate of $\mathcal{O}(1/k)$, and the following bound holds:

$$\frac{1}{K+1}\sum_{k=0}^{K}\left(\frac{2}{3N}\sum_{i=1}^{N}|x^i_{k+1} - z_{k+1}|^2 + \frac{1}{6N}\left|G_{k+1}\right|^2\right) = \mathcal{O}\left(\frac{1}{K}\right),$$

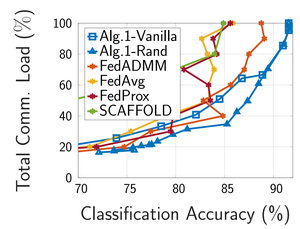
where $|x^i_{k+1} - z_{k+1}|$ are the residuals, and the gradient terms are given by $G_{k+1} \in \frac{1}{\rho N}\left(\sum_{i=1}^N \nabla f^i(x^i_{k+1}) + \partial g(z_{k+1})\right)$.

Empirical Evaluation

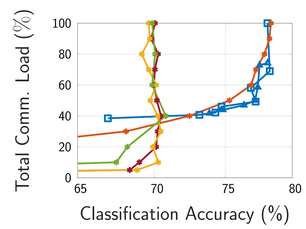
Event-based communication reduces communication cost and results in a better trade-off compared to well-known baselines FedAvg, FedProx, SCAFFOLD and FedADMM.

Algorithm	MNIST Target Acc.			CIFAR-10 Target Acc.			
	80%	85%	90%	70%	75%	77%	78%
Alg. 1 - Randomized	629	693	1723	12531	13422	15008	18376
Alg. 1 - Vanilla	816	1285	1710	12214	14780	14780	20690
FedADMM	800	1200	>2000	12000	15000	21000	27000
FedAvg	800	2000	n/a	3000	n/a	n/a	n/a
FedProx	1000	2000	n/a	6000	n/a	n/a	n/a
SCAFFOLD	1600	2000	3200	12000	n/a	n/a	n/a

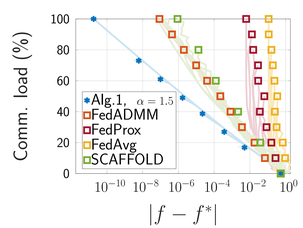
MNIST Classifier



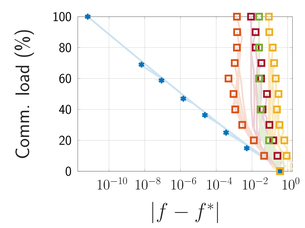
CIFAR-10 Classifier



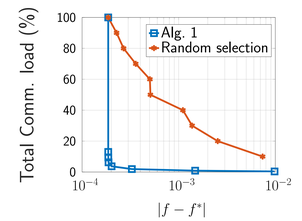
Linear Regression

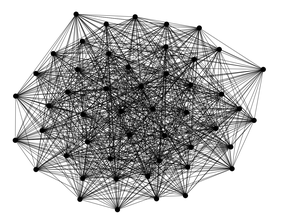


LASSO



Linear Regression over a 50-agent network with 1762 edges



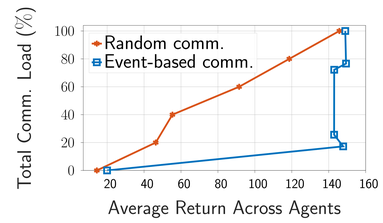


Extensions

Event-Based Federated Q-Learning

G. D. Er, M. Muehlebach,
ICML 2024 Workshop

This work integrates **event-triggered communication** principles into **federated reinforcement learning**. By only sharing significant updates, it reduces communication overhead while preserving learning performance across distributed agents.



Controlling Participation in Federated Learning with Feedback

M. Cummins, G. D. Er, M. Muehlebach,
L4DC 2025

This method **dynamically regulates the communication threshold** (Δ^i_k) based on real-time feedback on participation levels (L^i_k), thereby improving the communication and computational efficiency.