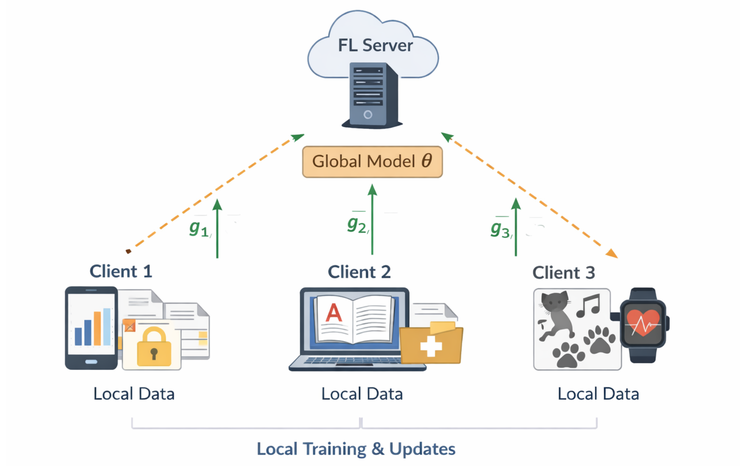
Distributed Quasi-Newton Method for Fair and Fast Federated Learning

Shayan Mohajer Hamidi, Linfeng Ye

ICLR

Federated Learning Setup



Motivation

Problem: First-order (FedAvg): slow & Second-order: fast but unfair
Issue: Global update may increase loss for some clients

Key Idea

Goal: achieve **fast** and **fair** federated learning

Key observation:

- Averaging updates → may hurt some clients
- Need a **common descent direction**

Our idea: construct a global update d_t such that

$$g_k \cdot d_t > 0 \quad \forall k$$

⇒ decreases **all local losses** (fairness)

Method Overview

At each round t, the server coordinates K clients.

Client k:

- Receive global model θ_t
- Compute local gradient $g_k = \nabla f_k(\theta_t)$
- Estimate curvature (quasi-Newton): $\tilde{d}_k = g_k^\top B_k g_k$
- Send $(g_k, \tilde{d}_k)$ to server

Server:

- Build a **common descent direction**
- Ensure update improves **all clients**
- Preserve **quasi-Newton speed**

Core Update

Step 1: Gradient Orthogonalization
Transform $\{g_k\}$ into orthogonal set $\{\tilde{g}_k\}$

Step 2: Optimal Weights (closed-form) $\lambda_k^\star = \frac{1/\|\tilde{g}_k\|^2}{\sum_{i=1}^K 1/\|\tilde{g}_i\|^2}$

Step 3: Global Direction $d_t = \sum_{k=1}^K \lambda_k^\star \tilde{g}_k$

Step 4: Model Update $\theta_{t+1} = \theta_t - \eta_t d_t$

Why It Works: $g_k \cdot d_t > 0 \quad \forall k$

- Descent for all clients, Fair updates, Curvature-aware
- Converges to Pareto solution
- Linear–quadratic rate

$$T_\epsilon = \begin{cases} \mathcal{O}(\log\log\frac{1}{\epsilon}) \\ \mathcal{O}(\log\frac{1}{\epsilon}) \end{cases}$$

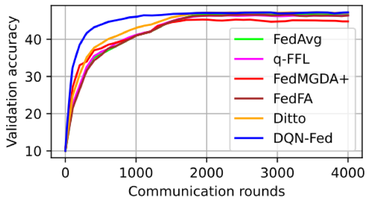
Experiments

Datasets: CIFAR-10, CIFAR-100, FEMNIST, Shakespeare

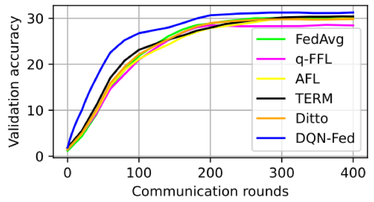
Results:

- Higher accuracy
- Lower variance (fairness)
- Faster convergence

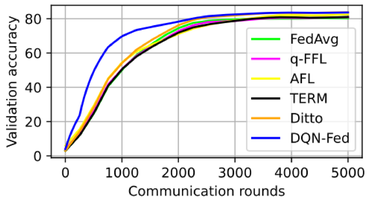
Participation	Algorithm	10% Participation			50% Participation		
		$\bar{a}$	σ_a	B(10%)	$\bar{a}$	σ_a	B(10%)
Naive First-Order	FedAvg	61.41	6.36	68.12	62.74	6.01	70.15
Fair First-Order	q-FFL	56.55	6.24	65.89	58.97	5.95	68.22
	FedMGDA	61.05	5.80	69.07	62.57	5.54	70.52
	FedHEAL	61.26	6.12	68.52	62.92	5.71	70.72
	TERM	63.22	5.93	70.56	63.81	5.66	71.83
	Ditto	62.87	6.07	69.92	63.51	5.76	71.24
Second-Order	FedNL	63.25	6.16	70.81	64.15	5.90	71.92
	FedNew	63.80	6.07	71.21	64.56	5.80	72.08
	DQN-Fed	**64.14**	**5.67**	**72.21**	**64.71**	**5.48**	**72.55**

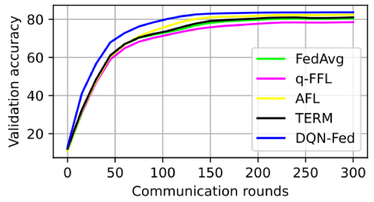


(a) CIFAR-10



(b) CIFAR-100





smohajer@stanford.edu; linfeng.ye@mail.utoronto.ca