FedScar: Correcting Geometric Bias for Flatness-Consistent Federated Learning

Jianfeng Lu [1] Yuzhao Xiang [1] Yue Chen [2*] Gang Li [3] Shuqin Cao [4] Guanghui Wen [5*]

ICML International Conference On Machine Learning

Problem Definition and Contribution

Motivations:

- The flawed assumption that local flatness transfers globally causes severe trajectory divergence under data heterogeneity.
- Existing methods only address first-order parameter drift, failing to correct the second-order geometric misalignment.

Key Contributions:

- We propose FedScar to explicitly correct geometric bias and enforce flatness consistency across clients.
- We introduce a novel method to steer local optimization trajectories toward globally flat minima.
- We prove that FedScar achieves exponential geometric contraction and empirically demonstrate its superior generalization under severe statistical heterogeneity.

Proposed Method: FedScar

Split-Dual ADMM Formulation:

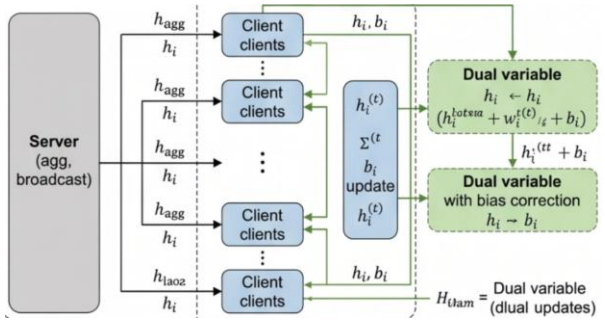
$$\mathcal{L} = \sum_{i=1}^{N} p_i \Big(f_i(w_i) + \frac{\beta}{2} \| w_i - z \|^2 + \beta \langle w_i - z, h_i \rangle + \langle w_i - z, \phi(b_i) \rangle \Big)$$

History-Accumulated Bias Tracking:

$$b_i \leftarrow (1-\gamma) b_i + \gamma \Delta_i, \quad v_i^t = \Delta w^t - \Delta w_i^t$$

Variance-Aware Perturbation Injection:

$$h_i^{t+1} = h_i^t + \frac{1}{\eta K}(w^t - w_i^t)$$



Process of FedScar

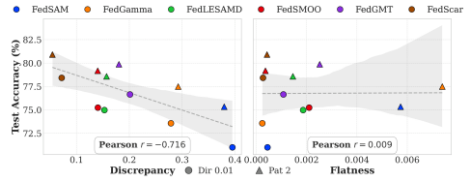
Rethink SAM-based Methods

$$|\max_{\|\delta\| \le \rho} F(w+\delta) - F(w) - [\sum_{i=1}^{N} \frac{m_i}{m} \max_{\|\delta_i\| \le \rho} F_i(w_i+\delta_i) - F_i(w_i)]| \longrightarrow \nabla f_i(w_i^{(t)})^\top (w^t - w_i^{(t)}) + \frac{1}{2}(w^t - w_i^{(t)})^\top \nabla^2 f_i(w_i^{(t)})(w^t - w_i^{(t)}).$$

$$|\underbrace{[\max_{\epsilon} f(w+\epsilon) - \sum_{i=1}^{N} p_i \max_{\epsilon_i} f_i(w+\epsilon_i)]}_{\text{Perturbed Loss Discrepancy}} - \underbrace{[f(w) - \sum_{i=1}^{N} p_i f_i(w)]}_{\text{Original Loss Discrepancy}}| \longrightarrow \tilde{\Delta}_{\mathcal{F}}^{(t)} := \underbrace{\Big| f(w) - \sum_{i=1}^{N} p_i f_i(w) \Big|}_{\text{First-Order Trajectory Deviation}} + \underbrace{\Big| \sum_{i=1}^{N} p_i f_i(w^t) - \sum_{i=1}^{N} p_i f_i(w_i^{(t)}) \Big|}_{\text{Higher-Order Geometric Bias}}$$

Research work	Flatness	Flatness Discrepancy	Bytes	FLOPS
FedSAM	✓	×	1×	2×
FedSpeed	✓	×	1×	2×
FedSMOO	✓	×	2×	2×
FedLESAM	✓	×	1×	1×
FedGMT	✓	×	1.5×	1.33×
FedScar (Ours)	✓	✓	1.5×	1×

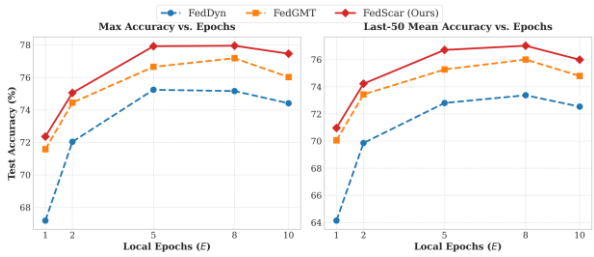
Comparison of different algorithms.

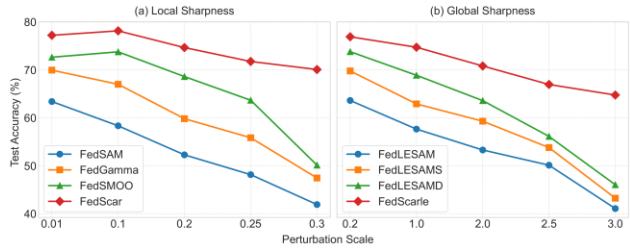


flatness_correlation

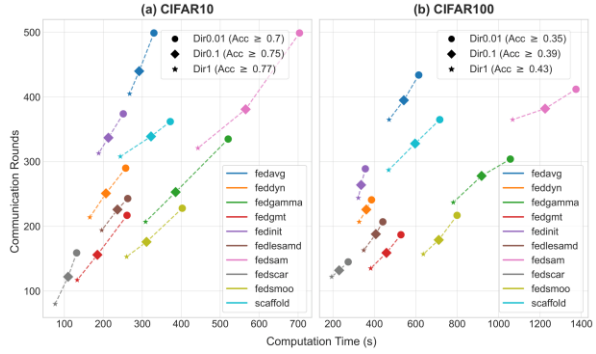
Experiments and Results

Robust of Training:

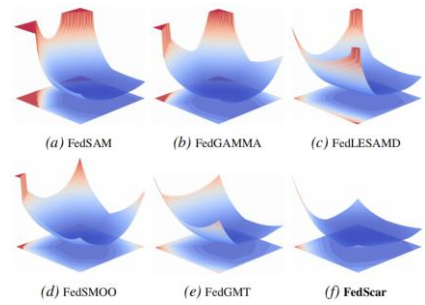




Efficiency of Training:



Flatness of Training:



Resources:

Paper

Code