THEORETICAL CONVERGENCE ANALYSIS

1. SUPPLEMENTARY THEORETICAL CONVERGENCE ANALYSIS

To complement the empirical evaluation, we provide a theoretical analysis showing that HFedAdv converges under standard federated optimization assumptions, with adversarial training introducing only a bounded perturbation.

1.1. Assumptions

Following prior works [1, 2], we state the assumptions required for convergence.

Assumption 1 (Smoothness). Each local loss $F_k(\omega)$ is L-smooth:

$$\|\nabla F_k(\omega) - \nabla F_k(\omega')\| \le L\|\omega - \omega'\|, \quad \forall \omega, \omega'. \tag{1}$$

This ensures gradients do not change abruptly.

Assumption 2 (Bounded variance). Stochastic gradients have bounded variance:

$$\mathbb{E}\left[\|\nabla F_k(\omega) - \nabla F(\omega)\|^2\right] \le \sigma^2. \tag{2}$$

This captures randomness from sampling and client heterogeneity.

Assumption 3 (Bounded gradients). Gradients are bounded:

$$\|\nabla F_k(\omega)\| \le G. \tag{3}$$

Assumption 4 (Stable adversarial training). For each client, the adversarial subproblem

$$L_k(\omega_k, \theta_k, \lambda_k) = \mathcal{L}_k^{ce} + \lambda_k \cdot \mathcal{L}_k^{dom} \tag{4}$$

admits an optimal λ_k^* per iteration, and the adversarial perturbation on updates is bounded by Δ.

1.2. Convergence Theorem

Theorem 1 (Average Convergence). If the learning rate $\eta \le 1/L$, then after T rounds HFedAdv satisfies

$$\frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}[\|\nabla F(\omega^t)\|^2] \le O\Big(\tfrac{1}{\sqrt{T}}\Big) + O(\sigma^2) + O(\Delta). \tag{5}$$

Hence, HFedAdv converges to a neighborhood of a stationary point, with the gap controlled by stochastic noise σ^2 and bounded adversarial effect Δ.

1.3. Proof

By Assumption 1, each $F(\cdot)$ is L-smooth. Thus, by a first-order Taylor expansion with a quadratic remainder, for any t we have

$$F(\omega^{t+1}) \le F(\omega^t) + \langle \nabla F(\omega^t),\, \omega^{t+1} - \omega^t\rangle + \tfrac{L}{2}\|\omega^{t+1} - \omega^t\|^2. \tag{6}$$

The update rule in HFedAdv can be written as

$$\omega^{t+1} = \omega^t - \eta \nabla F(\omega^t) + \xi^t, \tag{7}$$

where ξ^t denotes the error term caused by local updates, client heterogeneity, and bounded adversarial perturbations. Substituting (7) into (6) and expanding, we obtain

$$\langle \nabla F(\omega^t), \omega^{t+1} - \omega^t\rangle = -\eta\|\nabla F(\omega^t)\|^2 + \langle \nabla F(\omega^t), \xi^t\rangle, \tag{8}$$

$$\|\omega^{t+1} - \omega^t\|^2 \le \eta^2\|\nabla F(\omega^t)\|^2 + 2\eta\langle \nabla F(\omega^t), \xi^t\rangle + \|\xi^t\|^2. \tag{9}$$

Substituting (7) into (6), and using the bounds for the inner product and squared norm, we obtain

$$\mathbb{E}[F(\omega^{t+1})] \le \mathbb{E}[F(\omega^t)] - \eta\mathbb{E}[\|\nabla F(\omega^t)\|^2] + L\eta^2 G^2 + O(\sigma^2 + \Delta). \tag{10}$$

Summing from $t = 0$ to $T - 1$ and dividing by T, we obtain

$$\frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}[\|\nabla F(\omega^t)\|^2] \le \frac{F(\omega^0) - F(\omega^T)}{\eta T} + L\eta G^2 + O(\sigma^2 + \Delta). \tag{11}$$

Finally, by setting $\eta = O(1/\sqrt{T})$, we conclude that

$$\frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}[\|\nabla F(\omega^t)\|^2] \le O\Big(\tfrac{1}{\sqrt{T}}\Big) + O(\sigma^2 + \Delta), \tag{12}$$

which establishes the convergence of HFedAdv to a neighborhood of a stationary point under bounded stochastic noise and adversarial perturbations. □

Choosing $\eta = O(1/\sqrt{T})$ gives the stated result.

1.4. Discussion

This result implies that HFedAdv enjoys the same $O(1/\sqrt{T})$ convergence rate as standard FL methods, while the adversarial term Δ introduces only a bounded bias. Crucially, this bias is not harmful: by explicitly separating generalized and personalized features, adversarial training improves personalization without sacrificing convergence guarantees.

2. REFERENCES

[1] Y. Tan, G. Long, L. Liu, T. Zhou, Q. Lu, J. Jiang, and C. Zhang, "FedProto: Federated prototype learning across heterogeneous clients," vol. 36, no. 8, 2022, pp. 8432–8440.

[2] L. Yi, G. Wang, X. Liu, Z. Shi, and H. Yu, "FedGH: Heterogeneous federated learning with generalized global header." ACM, 2023, pp. 8686–8696.