PIN-LoRA: Positive-Incentive Noise Aggregation for Federated Long-Tailed Learning

Chenghao Jin[1], Shihao Hou[1], Jiacheng Yang[1], Mengke Li[2], Yiqun Zhang[3], Yang Lu[1*], and Yiu-ming Cheung[4]

[1]*Key Laboratory of Multimedia Trusted Perception and Efficient Computing, Ministry of Education of China, Xiamen University, Xiamen, China*

[2]*College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China*

[3]*School of Computer Science and Technology, Guangdong University of Technology, Guangzhou, China*

[4]*Department of Computer Science, Hong Kong Baptist University, Hong Kong, China*

Email: jinchenghao@stu.xmu.edu.cn, houshihao@stu.xmu.edu.cn, jiachengyang@stu.xmu.edu.cn, mengkeli@szu.edu.cn, yqzhang@gdut.edu.cn, luyang@xmu.edu.cn*, ymc@comp.hkbu.edu.hk

***Abstract*—Federated learning on long-tailed, highly non-IID data often suffers from heterogeneity, particularly in tail classes. However, there remains an open and unexplained phenomenon in federated learning with LoRA adapters: Tail-class performance can improve as heterogeneity increases. Existing FL–LoRA variants overlook this phenomenon, as they either adopt heterogeneous low-rank structures or rely on modified aggregation rules that depart from the standard shared-rank setting. We show that standard LoRA aggregation naturally introduces a similarity-dependent cross-client interaction term, which manifests as structured heterogeneous noise. Motivated by this observation, we decompose aggregated LoRA updates into an ideal component and a similarity-induced cross-client interaction term, which manifests as structured heterogeneous noise. We propose *PIN-LoRA*, which explicitly treats aggregation-induced cross-client interactions as controllable positive-incentive noise rather than eliminating them. Within a positive-incentive noise (pi-noise) framework, we further show that this structured noise can reduce task entropy while preserving task-relevant information, thereby explaining why *PIN-LoRA* becomes beneficial in certain heterogeneous regimes. Extensive experiments on long-tailed federated benchmarks validate our analysis and show substantial and robust gains, particularly under severe non-IID conditions.**

***Index Terms*—Federated learning, LoRA, long-tailed learning, positive-incentive noise**

I. Introduction

Federated learning (FL) enables multiple clients to collaboratively train a shared model without exchanging raw data, offering an effective solution to privacy, ownership, and regulatory constraints in distributed learning systems [1]–[3]. However, adapting large Vision-Language Models (VLMs) such as CLIP [4] and ALIGN [5] in federated learning remains challenging due to their massive parameter scales and the prohibitive communication cost of full fine-tuning. To mitigate this limitation, Parameter-Efficient Fine-Tuning (PEFT) has emerged as a practical paradigm for federated adaptation. Among PEFT methods, LoRA [6] is particularly well-suited to FL, as it enables efficient communication while largely preserving model performance.

*Corresponding author: Yang Lu (luyang@xmu.edu.cn).

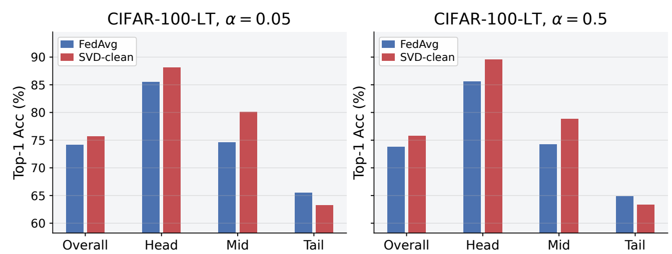

Fig. 1. **The performance comparison of LoRA aggregation on CIFAR-100-LT under different heterogeneity levels.** Suppressing cross-client interactions improves head performance but consistently degrades tail accuracy under heterogeneous long-tailed settings.

Despite LoRA's efficiency, data heterogeneity remains a fundamental challenge, as existing work does not place data heterogeneity induced by non-IID settings at the center of its investigation [7]–[12]. However, real-world data are often *long-tailed*, where head classes dominate and tail classes are sparse, presenting a compounded difficulty beyond standard non-IID settings. Crucially, in this specific long-tailed LoRA regime, we observe a counter-intuitive phenomenon: unlike in full fine-tuning, data heterogeneity can actually *enhance* tail-class representations.

In contrast to this observation, prevailing FL strategies predominantly focus on minimizing aggregation errors, operating under the premise that client discrepancies are purely detrimental noise. These methods typically employ drift-correction or variance-reduction mechanisms to align local updates, aiming to recover an "ideal" global model by eliminating non-IID interference [7]–[9], [11], [13]. However, we demonstrate that indiscriminately suppressing such discrepancies consistently degrades tail-class performance (Fig.1), suggesting that certain aggregation artifacts paradoxically encode structured information beneficial for the tail. This finding shifts the central research challenge from total error suppression to the nuanced task of *distinguishing between beneficial heterogeneity and harmful noise*.

Motivated by this, we revisit the aggregation behavior of shared-rank LoRA under heterogeneous and long-tailed federated settings. We first analyze the intrinsic structure of standard FedAvg-style LoRA aggregation and observe that client discrepancies induce structured cross-client interaction terms rather than purely random noise. Building on this observation, we reinterpret these interactions through an information-theoretic lens and formalize them as positive-incentive noise (π-noise) [14]. We further show that such perturbations act as a constructive inductive bias, reducing task entropy while preserving task-relevant information, thereby establishing heterogeneity as a necessary condition for these gains. Building on this insight, we propose *PIN-LoRA*, a principled aggregation framework that explicitly distinguishes beneficial heterogeneity from harmful noise in federated LoRA. *PIN-LoRA* addresses the fundamental challenge of preserving tail-beneficial cross-client interactions while avoiding destructive non-IID interference by treating aggregation-induced heterogeneity as positive-incentive noise and selectively injecting similarity-aware perturbations into a rank-preserving subspace, enabling controlled exploitation without sacrificing stability. Our contributions are summarized as follows:

- We propose *PIN-LoRA*, a lightweight and rank-preserving aggregation strategy consistently improving tail-class performance under severe non-IID distributions.
- We reveal a previously overlooked aggregation behavior of shared-rank LoRA under long-tailed federated settings, where heterogeneity can benefit tail representations.
- We establish a theoretical connection between LoRA aggregation and entropy-reducing π-noise, proving the existence and necessity of heterogeneity-induced gains.
- We formally characterize how heterogeneity-induced gains scale with client discrepancy, proving a monotonic relationship between heterogeneity and tail performance improvements.

II. Background and Related Work

A. PEFT in Federated Learning

Recent work has explored parameter-efficient fine-tuning (PEFT) as a practical solution for federated learning, with LoRA-based methods becoming a dominant choice due to their communication efficiency and compatibility with large pre-trained models [6]. Early FL+LoRA studies focus on transmitting low-rank updates or decoupling shared and client-specific adapters to balance personalization and efficiency [10], [15]–[17]. More recent works extend this line by introducing heterogeneous or structured aggregation mechanisms, allowing clients to adopt different ranks or selectively participate via rank alignment, weighting, or parameter selection [9], [11], [12], [18], [19]. Despite these advances, existing approaches consistently treat heterogeneity as instability to be corrected, suppressing aggregation artifacts without examining their potential constructive role in long-tailed representation learning.

B. Federated Long-tailed Learning

Existing solutions for data heterogeneity in FL mainly operate from two perspectives: optimization constraints to limit local drift [7], [20], [21] and server-side aggregation corrections to mitigate non-IID bias [8], [13], [22]. Although these approaches stabilize non-IID training, they typically treat aggregation neutrally and neglect global long-tailed distributions, resulting in suboptimal performance on tail classes. Conversely, strategies addressing class imbalance often rely on client selection based on uploaded distributions [23], [24] or auxiliary server data [25], [26]. However, these approaches risk violating privacy principles or assume unrealistic data availability. While centralized long-tailed techniques exist [27]–[30], adapting them to FL without global statistics remains challenging, with CReFF [31] being a rare exception addressing this intersection.

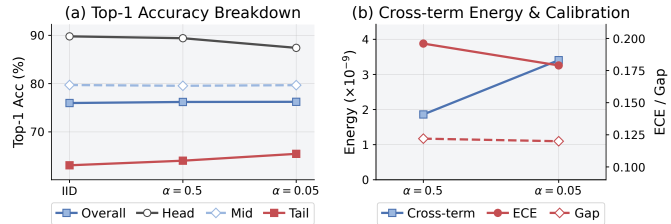

Fig. 2. **Heterogeneity-induced gain in federated LoRA.** As client heterogeneity increases, cross-term energy grows while calibration improves (lower ECE / Gap), accompanied by higher tail-class accuracy.

C. Positive-Incentive Noise

Most prior noise-based learning methods introduce explicit stochastic perturbations for regularization or robustness. In contrast, positive-incentive noise (π-noise) posits that certain perturbations can reduce task complexity while preserving relevant information, serving as a beneficial inductive bias rather than pure interference [14]. Supported by entropy-based analyses of representation learning [32], this framework provides a criterion to distinguish constructive noise from detrimental errors. Unlike prior works relying on explicit noise injection, we focus on structured perturbations arising *implicitly* from heterogeneous aggregation, analyzing conditions where these interactions yield beneficial entropy reduction.

III. Preliminaries

A. Motivation

In federated long-tailed settings, we observe a counter-intuitive phenomenon: as client data become more non-IID, tail-class performance improves rather than degrades. As shown in Fig. 2(a), increasing heterogeneity (smaller α) leads to a monotonic increase in tail-class accuracy. Meanwhile, Fig. 2(b) shows that this improvement is accompanied by a significant growth in cross-client interaction energy and better calibration (lower ECE / Gap), suggesting that the tail gains are closely tied to aggregation artifacts induced by heterogeneity.

Motivated by these observations, we turn to the aggregation mechanism itself. Under standard FedAvg, LoRA factors are averaged separately across clients. For clarity, consider the

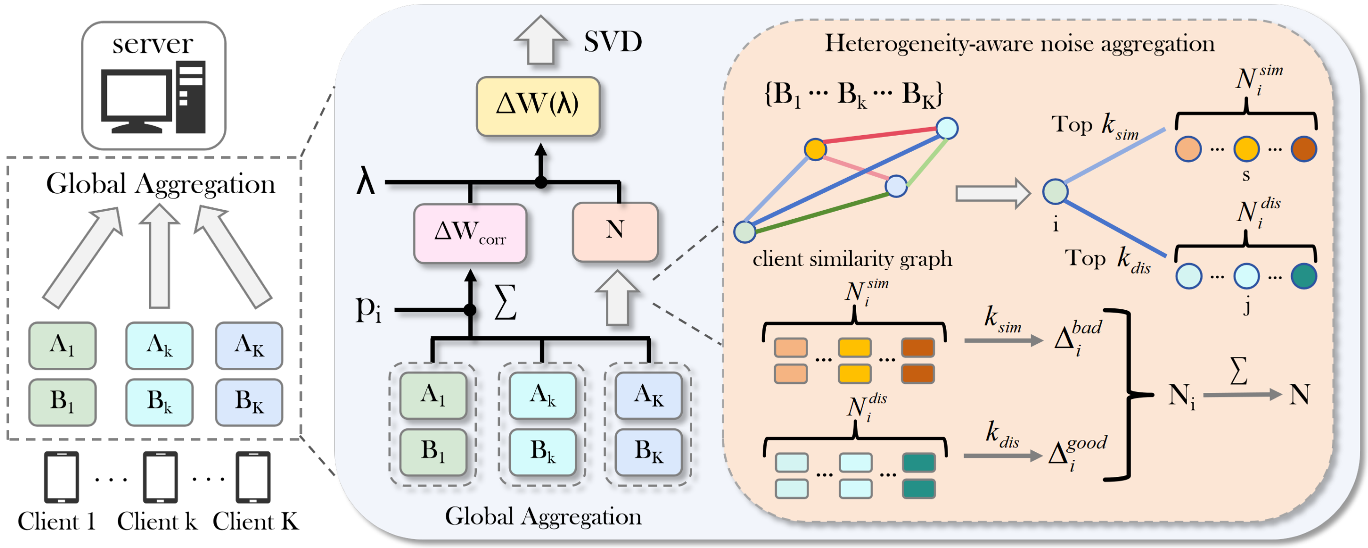

Fig. 3. **Overview of *PIN-LoRA*.** Clients locally update shared-rank LoRA adapters and upload factors $(\mathbf{A}_i, \mathbf{B}_i)$ to the server. The server first performs the correct low-rank aggregation $\Delta\mathbf{W}_{\text{corr}} = \sum_i p_i \mathbf{B}_i \mathbf{A}_i$, then constructs a heterogeneity-aware structured noise term $\mathbf{N}$ from cross-client interactions using a similarity graph over $\{\mathbf{B}_i\}$. The final update is obtained by injecting positive-incentive noise to form $\Delta\mathbf{W}_{\text{corr}} + \lambda\mathbf{N}$, followed by a rank-$r$ SVD projection.

case of two clients ($K = 2$) with aggregation weights p_0 and p_1. The resulting LoRA update can be written as

$$\begin{aligned} \Delta\mathbf{W}_{\text{FedAvg}} &= (p_0\mathbf{B}_0 + p_1\mathbf{B}_1)(p_0\mathbf{A}_0 + p_1\mathbf{A}_1) \\ &= p_0^2\mathbf{B}_0\mathbf{A}_0 + p_1^2\mathbf{B}_1\mathbf{A}_1 \\ &\qquad + p_0 p_1(\mathbf{B}_0\mathbf{A}_1 + \mathbf{B}_1\mathbf{A}_0), \end{aligned} \tag{1}$$

where $\mathbf{A}_i$ and $\mathbf{B}_i$ denote the LoRA low-rank factors learned on client i, and $\mathbf{B}_i\mathbf{A}_i$ corresponds to the local LoRA weight update. The last term consists of *cross-client interaction terms* that do not appear in the ideal "correct" aggregation $\sum_i p_i\mathbf{B}_i\mathbf{A}_i$. These interactions are an inherent consequence of FedAvg under heterogeneous LoRA updates and are closely linked to the empirical trends in Fig. 2. This observation motivates the need to explicitly model and regulate cross-client interactions during aggregation, rather than treating them as undifferentiated noise.

Guided by this perspective, we propose *PIN-LoRA*, which introduces controlled aggregation of FedAvg-induced cross-client interactions. To assess the theoretical feasibility of this design, we further analyze heterogeneity-induced LoRA noise through an information-theoretic framework based on positive-incentive noise (π-noise). The formal analysis is presented in Section IV-B.

B. Settings and Notations

We consider a federated setting with K clients, where client i holds a dataset $\mathcal{D}_i$ of size n_i with aggregation weight $p_i = n_i/n$. The collective data distribution is highly non-IID and globally long-tailed. With a frozen backbone $f(x;\Theta_0)$, adaptation is performed via rank-r LoRA updates:

$$\Delta\mathbf{W}_i = \mathbf{B}_i\mathbf{A}_i, \quad \text{with } \mathbf{B}_i \in \mathbb{R}^{d_{\text{out}}\times r}, \mathbf{A}_i \in \mathbb{R}^{r\times d_{\text{in}}}. \tag{2}$$

where $(\mathbf{B}_i, \mathbf{A}_i)$ denote the trainable rank-r LoRA parameters of client i. The server aggregates the resulting increments and projects the update back to rank-r form via truncated SVD to maintain communication efficiency. This formulation allows us to explicitly characterize cross-client interactions as structured aggregation noise in Section IV-B.

IV. Methodology

A. PIN-LoRA: Structured-Noise Aggregation

PIN-LoRA operates in a standard federated training loop where, in each communication round, clients locally update their shared-rank LoRA adapters and upload the resulting factors to the server. The server first forms the correct low-rank aggregation of client updates, then identifies and reweights cross-client interactions according to client similarity. Based on this decomposition, a structured heterogeneity-aware noise term is constructed and injected into the aggregation, followed by a rank-r projection to preserve the shared LoRA parameterization. Algorithm 1 summarizes the complete client–server workflow of the proposed aggregation procedure and clarifies the execution order of these steps.

Structured-noise-driven aggregation. Guided by our analysis in Section IV-B, our method constructs a practical heterogeneity-aware noise term that preserves the informative components of cross-client interactions while suppressing redundant or harmful ones. The construction operates directly on LoRA increments of the form $\Delta\mathbf{W}_i = \mathbf{B}_i\mathbf{A}_i$, which are the only trainable parameters in our federated setting. The key idea is to use client similarity, measured in the space of LoRA matrices, to decompose cross interactions into beneficial (dissimilar) and interfering (similar) components.

Client-similarity graph. To quantify cross-client relationships, we compute pairwise similarity using the LoRA down-projection matrices $\{\mathbf{B}_i\}_{i=1}^K$, adopting cosine similarity as the metric. For each client i, we select the k_{sim} most similar clients as its near neighbors N_i^{sim} and the k_{dis} most dissimilar clients as its far neighbors N_i^{dis}. The corresponding normalized similarity scores serve as weighting coefficients when aggregating cross noise. This graph encodes the heterogeneity structure upon which our noise construction is built.

Local cross-noise decomposition. We define two types of local cross-noise components via the similarity graph. The *bad* noise captures interference from strongly similar clients:

$$\Delta_i^{\mathrm{bad}} = \frac{1}{k_{\mathrm{sim}}} \sum_{s \in N_i^{\mathrm{sim}}} s_{is} (\mathbf{B}_i - \mathbf{B}_s)(\mathbf{A}_s - \mathbf{A}_i), \tag{3}$$

where N_i^{sim} denotes the set of the k_{sim} most similar clients to client i under cosine similarity of LoRA matrices, and s_{is} is the corresponding normalized similarity weight. Complementarily, the *good* noise is defined over dissimilar clients as

$$\Delta_i^{\mathrm{good}} = \frac{1}{k_{\mathrm{dis}}} \sum_{j \in N_i^{\mathrm{dis}}} d_{ij} (\mathbf{B}_i - \mathbf{B}_j)(\mathbf{A}_j - \mathbf{A}_i). \tag{4}$$

These definitions are motivated by the cross-client interaction decomposition formalized in Section IV-B, where differences $(\mathbf{B}_i - \mathbf{B}_j)(\mathbf{A}_j - \mathbf{A}_i)$ isolate cross-client interaction terms while eliminating redundant diagonal components.

Structured heterogeneity noise. For each client, we combine the two components by subtracting near-client interference from far-client contributions:

$$\widehat{\mathbf{N}}_i = \Delta_i^{\mathrm{good}} - \Delta_i^{\mathrm{bad}}. \tag{5}$$

Aggregating over all clients yields the global structured noise $\mathbf{N} = \sum_{i=1}^{K} \widehat{\mathbf{N}}_i$, which is designed to align with a similarity-weighted projection of the theoretical cross-client noise analyzed in Section IV-B.

Correct aggregation with noise injection. Let the correct low-rank LoRA aggregation be

$$\Delta \mathbf{W}_{\mathrm{corr}} = \sum_{i=1}^{K} p_i \mathbf{B}_i \mathbf{A}_i. \tag{6}$$

We inject the structured noise via

$$\Delta \mathbf{W}^{(\lambda)} = \Delta \mathbf{W}_{\mathrm{corr}} + \lambda \mathbf{N}, \tag{7}$$

where $\lambda \geq 0$ controls the noise strength. This mirrors the injected family $\Delta \mathbf{W}(\lambda)$ analyzed in Section IV-B, ensuring consistency between the aggregation mechanism and the subsequent analysis.

Rank-constrained projection via SVD. To maintain rank-r LoRA parameterization, we apply truncated SVD to $\Delta \mathbf{W}^{(\lambda)}$:

$$\Delta \mathbf{W}^{(\lambda)} \approx \mathbf{U}_r \Sigma_r \mathbf{V}_r^{\top}, \tag{8}$$

where $\mathbf{U}_r$, Σ_r, and $\mathbf{V}_r$ denote the rank-r truncated singular vectors and singular values of $\Delta \mathbf{W}^{(\lambda)}$. The global LoRA factors are then updated as

$$\mathbf{B}^{t+1} = \mathbf{U}_r \Sigma_r^{1/2}, \qquad \mathbf{A}^{t+1} = \Sigma_r^{1/2} \mathbf{V}_r^{\top}. \tag{9}$$

These updated factors are broadcast to clients for the next round of federated training. This procedure ensures compatibility with existing LoRA pipelines while injecting entropy-reducing structured noise into the aggregation step.

We next present an information-theoretic analysis framework (Section IV-B) to characterize when and why the induced cross-client noise becomes beneficial.

Algorithm 1 Noise Aggregation of *PIN-LoRA*

Input: round index t, clients $\mathcal{A}^t$, LoRA rank r, noise scale λ
Output: aggregated LoRA factors $(\mathbf{A}^{t+1}, \mathbf{B}^{t+1})$
Client-side.
for each client $i \in \mathcal{A}^t$ **in parallel do**
    $(\mathbf{A}_i, \mathbf{B}_i) \leftarrow \mathrm{LocalUpdate}(\mathbf{A}^t, \mathbf{B}^t)$
    Upload $(\mathbf{A}_i, \mathbf{B}_i)$ to server
end for
Server-side aggregation.
Correct aggregation $\Delta \mathbf{W}_{\mathrm{corr}} \leftarrow \sum_{i \in \mathcal{A}^t} p_i \mathbf{B}_i \mathbf{A}_i$
Construct similarity graph from $\{\mathbf{B}_i\}_{i \in \mathcal{A}^t}$
Aggregate structured heterogeneity noise:
    $\mathbf{N} \leftarrow \sum_{i \in \mathcal{A}^t} \big(\Delta_i^{\mathrm{good}} - \Delta_i^{\mathrm{bad}}\big)$
Inject noise: $\Delta \mathbf{W}(\lambda) \leftarrow \Delta \mathbf{W}_{\mathrm{corr}} + \lambda \mathbf{N}$
Project $\Delta \mathbf{W}(\lambda)$ back to rank r via truncated SVD:
    $(\mathbf{A}^{t+1}, \mathbf{B}^{t+1}) \leftarrow \mathrm{Rank}\text{-}r\mathrm{Proj}\big(\Delta \mathbf{W}(\lambda)\big)$
Return $(\mathbf{A}^{t+1}, \mathbf{B}^{t+1})$

B. Analysis of Heterogeneity-Induced LoRA Noise

Cross-client interactions in FedAvg. We analyze the explicit form of LoRA aggregation under standard FedAvg. For the two-client case, the resulting update expands as

$$\begin{aligned} \Delta \mathbf{W}_{\mathrm{FedAvg}} = p_0^2 \mathbf{B}_0 \mathbf{A}_0 + p_1^2 \mathbf{B}_1 \mathbf{A}_1 \\ + p_0 p_1 (\mathbf{B}_0 \mathbf{A}_1 + \mathbf{B}_1 \mathbf{A}_0). \end{aligned} \tag{10}$$

where the last term corresponds to cross-client interaction terms that do not appear in the ideal aggregation $\Delta \mathbf{W}_{\mathrm{corr}} = \sum_i p_i \mathbf{B}_i \mathbf{A}_i$, and is referred to as the *cross-client noise*.

From raw cross noise to structured noise. Motivated by the cross-client interactions identified above, we collect all off-diagonal terms into a theoretical cross-noise

$$\mathbf{N}_{\mathrm{cross}} := \sum_{i \neq j} p_i p_j \, \mathbf{B}_i \mathbf{A}_j, \tag{11}$$

which is dense and entangled with diagonal-aligned components, making it unsuitable for direct aggregation. We therefore reparameterize cross interactions via a bilinear expansion over pairwise differences between LoRA factors, yielding the structured noise

$$\mathbf{N}_{\mathrm{practical}} := \sum_{i<j} \Omega_{ij} (\mathbf{B}_i - \mathbf{B}_j)(\mathbf{A}_j - \mathbf{A}_i), \tag{12}$$

where Ω_{ij} are signed similarity-based weights. This construction preserves cross-client interaction structure while canceling diagonal redundancy, resulting in a compact and heterogeneity-aware reparameterization of $\mathbf{N}_{\mathrm{cross}}$.

Importantly, Eqs. (11) and (12) show that $\mathbf{N}_{\mathrm{practical}}$ is a similarity-weighted reparameterization of $\mathbf{N}_{\mathrm{cross}}$ that preserves cross-client interaction structure while removing diagonal redundancy. Consequently, the noise term $\mathbf{N}$ used in *PIN-LoRA* can be viewed as a sparse and implementable approximation of $\mathbf{N}_{\mathrm{cross}}$ that retains the same class of heterogeneity-induced perturbations, allowing theoretical analysis to directly transfer to the practical aggregation mechanism.

Definition 1 (Positive-Incentive Noise)**.** *Let z denote a representation and let $\tilde{z} = (\mathbf{I} + \mathbf{Q})z$ be a linearized perturbation* *induced by aggregation. The perturbation is called* positive-incentive noise *if the induced entropy change* $\Delta S := H(z) - H(\tilde{z})$ *satisfies* $\Delta S > 0$.

Injected aggregation family. To isolate the effect of cross-client noise, we consider the one-parameter family of aggregated updates:

$$\Delta \mathbf{W}(\lambda) = \Delta \mathbf{W}_{\mathrm{corr}} + \lambda \mathbf{N}_{\mathrm{cross}}, \quad \lambda \geq 0, \tag{13}$$

where $\lambda = 1$ recovers standard FedAvg and $\lambda = 0$ removes all cross-client interactions. This formulation allows us to analyze whether the cross noise introduced by FedAvg can act as positive-incentive noise.

Linearized representation perturbation. Under a small-perturbation assumption, cross-client noise induces a linearized representation shift of the form $\tilde{z} = (\mathbf{I} + \lambda \mathbf{Q})z$, where $\mathbf{Q}$ is a linear operator derived from cross-client interactions. We assume $\mathbf{Q}$ satisfies mild regularity conditions: $\mathrm{tr}(\mathbf{Q}) = 0$, $\mathbf{Q} \neq \mathbf{0}$ under heterogeneity, and $\|\mathbf{Q}\|$ is bounded, which are sufficient for the analysis.

Proposition IV.1 (Entropy Change under Linear Perturbation)**.** *If* $\mathbf{I} + \lambda \mathbf{Q}$ *is invertible, then*

$$\Delta S(\lambda) = H(z) - H((\mathbf{I} + \lambda \mathbf{Q})z) = -\log |\mathbf{I} + \lambda \mathbf{Q}|.$$

Theorem IV.1 (Existence of Beneficial Cross-Client Noise)**.** *If* $\mathbf{Q} \neq \mathbf{0}$ *and* $\mathrm{tr}(\mathbf{Q}) = 0$*, then there exists* $\lambda^{\star} > 0$ *such that for all* $0 < \lambda \leq \lambda^{\star}$*,*

$$\Delta S(\lambda) > 0.$$

This follows from a second-order expansion of $\log |\mathbf{I} + \lambda \mathbf{Q}|$*, where the first-order term vanishes due to* $\mathrm{tr}(\mathbf{Q}) = 0$ *and the leading quadratic term is strictly negative.*

Theorem IV.2 (Necessity of Client Heterogeneity)**.** *If all clients share identical or collinear LoRA factors* $(\mathbf{A}_i, \mathbf{B}_i)$*, then* $\mathbf{N}_{cross} \propto \Delta \mathbf{W}_{\mathrm{corr}}$ *and* $\mathbf{Q} = \mathbf{0}$ *after baseline orthogonalization. Consequently,* $\Delta S(\lambda) = 0$*. Conversely, client heterogeneity is necessary for any positive entropy reduction.*

Theorem IV.3 (Heterogeneity–Gain Scaling Law)**.** *Define the heterogeneity index*

$$\begin{aligned} H_{cross} := \sum_{i<j} p_i p_j \Big(&\|\mathbf{B}_i - \mathbf{B}_j\|_F \, \sigma_{\min}(\mathbf{A}_j - \mathbf{A}_i) \\ &+ \|\mathbf{A}_i - \mathbf{A}_j\|_F \, \sigma_{\min}(\mathbf{B}_j - \mathbf{B}_i) \Big). \end{aligned} \tag{14}$$

There exist constants $c > 0$ *and* $\lambda^{\star} > 0$ *such that for all* $0 < \lambda \leq \lambda^{\star}$*,*

$$\Delta S(\lambda) \geq c \, \lambda^2 H_{cross}^2.$$

This result implies that, for sufficiently small λ*, the entropy reduction induced by cross-client noise increases monotonically with the degree of client heterogeneity.*

V. Experiments

A. Experimental Setup

We evaluate the proposed structured cross-noise injection strategy under a federated long-tailed learning setting. All experiments are conducted on CIFAR-style long-tailed benchmarks, with a particular focus on CIFAR-100-LT as the primary evaluation dataset. Unless otherwise specified, the imbalance factor is fixed at IF=100. Client data heterogeneity is simulated using a Dirichlet distribution with concentration parameter α. We consider $\alpha \in \{0.05, 0.1, 0.5\}$ to represent different degrees of heterogeneity, where smaller α corresponds to more severe non-IID data distributions. In Section V-C, we mainly report results under $\alpha = 0.05$, which represents a highly heterogeneous scenario.

We adopt a standard federated learning pipeline with a fixed backbone network and employ LoRA for parameter-efficient adaptation. All methods share the same backbone architecture and initialization to ensure fair comparison. Unless otherwise specified, the LoRA rank is set to $r = 16$. Model performance is evaluated using Top-1 accuracy on the test set. Following common practice in long-tailed recognition, we report accuracy for Overall, Head, Mid, and Tail class groups.

B. Comparison with Baseline Methods

We compare *PIN-LoRA* with representative federated learning baselines under long-tailed and heterogeneous data distributions, including FedAvg and FedProx with LoRA adapters, FedSA-LoRA, and an SVD-based clean aggregation variant (SVD-clean), under a unified federated setup and backbone configuration. Results on CIFAR-10/100-LT across different heterogeneity levels are summarized in Table I, while Table II reports the corresponding results on ImageNet-LT with a fine-grained breakdown over Head, Mid, and Tail classes.

Results on CIFAR-10/100-LT. Across both CIFAR benchmarks and all heterogeneity levels, *PIN-LoRA* consistently achieves strong overall performance, with the advantage being most pronounced under severe heterogeneity. On CIFAR-100-LT with $\alpha = 0.05$, *PIN-LoRA* improves overall accuracy by 3.43% over FedAvg and 1.12% over SVD-clean. Notably, the gains on tail classes are substantially larger. For instance, under $\alpha = 0.05$ on CIFAR-100-LT, *PIN-LoRA* improves tail accuracy by over 4% compared to both baselines, with similar trends on CIFAR-10-LT, where *PIN-LoRA* achieves the highest tail accuracy across all heterogeneity settings.

Results on ImageNet-LT. On ImageNet-LT, under stronger heterogeneity (e.g., $\alpha = 0.05$ and $\alpha = 0.1$), some baselines achieve slightly higher overall or head accuracy than *PIN-LoRA*. This is expected in large-scale long-tailed datasets, where head classes dominate the sample budget and implicit alignment in aggregation tends to favor dominant representations under highly non-IID partitions. In contrast, *PIN-LoRA* explicitly regulates cross-client interaction terms in LoRA aggregation, preventing the global update from being dominated by head-heavy clients. As a result, *PIN-LoRA* consistently achieves stronger tail performance across all α, while its

Method	CIFAR-100-LT						CIFAR-10-LT					
	α=0.05		α=0.1		α=0.5		α=0.05		α=0.1		α=0.5	
	Overall	Tail	Overall	Tail	Overall	Tail	Overall	Tail	Overall	Tail	Overall	Tail
FedAvg (LoRA)	74.13	65.50	73.77	64.85	73.81	63.66	90.99	86.16	89.70	83.52	90.23	82.43
FedProx (LoRA)	76.23	65.45	76.20	64.03	77.54	64.77	90.42	84.88	90.23	83.62	90.39	83.22
FedSA-LoRA	76.34	64.23	76.74	65.25	76.59	66.42	90.54	85.54	90.35	85.44	90.85	84.44
SVD-clean	76.44	63.15	76.97	65.92	77.38	67.23	93.71	90.62	93.70	90.58	94.41	91.40
PIN-LoRA (ours)	**77.56**	**67.30**	**78.16**	**67.95**	**78.19**	**68.60**	**94.25**	**92.28**	**94.38**	**92.46**	**95.24**	**93.18**

TABLE I
TOP-1 TEST ACCURACY (%) (OVERALL / TAIL) ON CIFAR-100-LT AND CIFAR-10-LT UNDER DIFFERENT α (IF=100, R=16).

Method	ImageNet-LT											
	α=0.05				α=0.1				α=0.5			
	Overall	Head	Mid	Tail	Overall	Head	Mid	Tail	Overall	Head	Mid	Tail
FedAvg (LoRA)	**69.88**	**78.21**	69.24	59.37	69.40	**81.75**	**69.15**	57.34	70.84	82.35	69.06	56.86
FedProx (LoRA)	68.90	74.66	68.37	62.26	68.42	74.34	67.91	61.47	68.46	75.06	67.48	61.30
FedSA-LoRA	68.62	75.68	69.11	62.30	68.86	75.03	68.15	61.49	69.52	75.92	68.13	62.12
SVD-clean	68.90	74.11	69.49	63.53	69.07	76.32	67.48	62.93	69.32	77.73	68.49	61.73
PIN-LoRA (ours)	69.60	74.98	**69.97**	**64.38**	**70.43**	76.42	68.61	**64.00**	**71.24**	**78.03**	**69.84**	**64.28**

TABLE II
TOP-1 TEST ACCURACY (%) (OVERALL / HEAD / MID / TAIL) ON IMAGENET-LT UNDER DIFFERENT α (IF=100, R=16).

overall accuracy becomes competitive again as heterogeneity decreases (e.g., attaining the best overall result at $\alpha = 0.5$).

Overall, the results suggest that explicitly decomposing cross-client interactions in LoRA aggregation and reintroducing them in a structured manner enables the global model to better balance shared and client-specific information. This design makes the aggregated updates less dominated by head-class clients, thereby improving robustness to data heterogeneity and benefiting tail-class generalization.

C. Model Validation

We further validate the key modeling assumptions underlying *PIN-LoRA* through controlled ablation studies, focusing on the effect of noise strength and noise structure.

Effect of Noise Injection Strength λ. We investigate the effect of the noise injection strength λ in *PIN-LoRA*, recalling that λ controls the magnitude of the injected cross-client noise term, which is theoretically shown to act as positive-incentive noise under a small perturbation regime. Fig. 4 visualizes the relative accuracy change (Δ Accuracy) with respect to the no-noise baseline ($\lambda = 0$) on CIFAR-100-LT with $\alpha = 0.05$. When $\lambda = 0$, *PIN-LoRA* reduces to the baseline aggregation without noise injection. Introducing a small amount of noise ($\lambda = 0.01$) consistently improves overall performance and yields the best tail accuracy, which indicates that moderate heterogeneous cross-noise can enhance representation generalization under severe data heterogeneity. Specifically, when $\lambda = 0.01$, tail accuracy improves by approximately +0.8% compared to the no-noise baseline, while overall accuracy increases by around +0.4%, confirming the existence of a beneficial small-noise regime. As λ increases further, performance begins to degrade, particularly for tail classes. This observation is consistent with our theoretical analysis, which predicts that the entropy-reduction benefit of

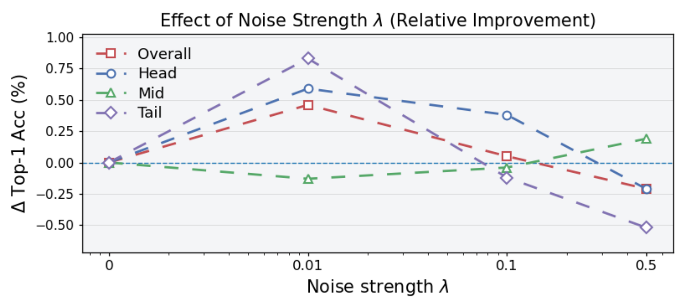

Fig. 4. Relative accuracy changes (Δ Accuracy) under different noise strengths λ on CIFAR-100-LT with $\alpha = 0.05$.

cross-noise injection holds only within a bounded small-λ regime, as well as excessive noise violates the linear perturbation assumption and may introduce harmful distortion rather than beneficial regularization. Overall, these results empirically confirm that the effect of heterogeneous noise is non-monotonic and can be effectively controlled via the noise strength parameter λ.

Effect of Noise Structure. To examine whether the performance gain of *PIN-LoRA* depends on the structured construction of the injected noise, we specifically compare different noise structures, including no noise ($\lambda = 0$), random noise, noise constructed from only beneficial interactions or detrimental interactions, and the full structured noise defined as the difference between beneficial and detrimental interactions. Table III reports the Top-1 accuracy across Overall, Head, Mid, and Tail classes under different noise structures on CIFAR-100-LT with $\alpha = 0.05$. As shown in Table III, random noise injection leads to only marginal tail improvement (+0.27% over no noise), whereas using only beneficial or only detrimental interactions fails to consistently improve tail performance. In contrast, the full structured noise (Good −

Noise structure	CIFAR-100-LT			
	Overall	Head	Mid	Tail
No noise ($\lambda = 0$)	76.44	90.28	80.65	63.15
Random noise	76.28	90.12	80.38	63.42
Good interactions only	76.95	90.08	80.72	65.67
Bad interactions only	75.82	**90.35**	80.21	62.18
Good $-$ Bad (ours)	**77.56**	90.19	**80.9**	**67.3**

TABLE III
TOP-1 TEST ACCURACY (%) OF DIFFERENT NOISE STRUCTURES ON CIFAR-100-LT UNDER HETEROGENEOUS FEDERATED LEARNING.

Bad) achieves the highest tail accuracy (67.30), outperforming random noise by +1.88% and no-noise aggregation by +4.15%. This behavior can be attributed to the way structured noise selectively amplifies heterogeneous cross-client interactions while suppressing redundant or conflicting ones. Specifically, isolating only beneficial or only detrimental interactions leads to biased perturbations, whereas their structured combination (Good $-$ Bad) preserves complementary information across clients and avoids reinforcing spurious correlations. As a result, the injected noise reshapes the aggregated update in a more balanced manner, which is particularly favorable for tail classes under severe heterogeneity.

VI. CONCLUSION

In this paper, we have revisited federated learning with shared-rank LoRA on heterogeneous, long-tailed client data and show that cross-client interaction terms introduced by standard aggregation can serve as a beneficial inductive bias, especially for tail classes. By decomposing standard LoRA aggregation into an ideal update and a heterogeneity-dependent interaction term, we have interpreted the latter as a structured perturbation and show that, under a small-perturbation regime, it can act as π-noise by reducing representation entropy without erasing task-relevant information. From an information-theoretic perspective, we have established the existence of a beneficial noise interval and further show that client heterogeneity is necessary. Guided by these insights, we have proposed *PIN-LoRA*, a lightweight server-side aggregation mechanism that selectively retains beneficial cross-client interactions while suppressing detrimental interference. Experiments on federated long-tailed benchmarks have validated our analysis, demonstrating consistent gains in tail-class accuracy and calibration under severe non-IID data partitions.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China under Grants 62376233, 62431004, and 62476063; in part by the Natural Science Foundation of Fujian Province under Grant 2024J09001; in part by the Natural Science Foundation of Guangdong Province under Grant 2025A1515011293; in part by the Shenzhen Science and Technology Program under Grant RCBS20231211090659101; in part by the National Key Laboratory of Radar Signal Processing under Grant JKW202403; in part by the General Research Fund of Research Grants Council under Grants 12201323 and 12202025.

REFERENCES

[1] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," 2023. [Online]. Available: https://arxiv.org/abs/1602.05629
[2] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," 2017. [Online]. Available: https://arxiv.org/abs/1610.05492
[3] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," 2019. [Online]. Available: https://arxiv.org/abs/1902.04885
[4] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, "Learning transferable visual models from natural language supervision," 2021. [Online]. Available: https://arxiv.org/abs/2103.00020
[5] C. Jia, Y. Yang, Y. Xia, Y.-T. Chen, Z. Parekh, H. Pham, Q. V. Le, Y. Sung, Z. Li, and T. Duerig, "Scaling up visual and vision-language representation learning with noisy text supervision," 2021. [Online]. Available: https://arxiv.org/abs/2102.05918
[6] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "Lora: Low-rank adaptation of large language models," 2021. [Online]. Available: https://arxiv.org/abs/2106.09685
[7] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," 2020. [Online]. Available: https://arxiv.org/abs/1812.06127
[8] J. Wang, Q. Liu, H. Liang, G. Joshi, and H. V. Poor, "Tackling the objective inconsistency problem in heterogeneous federated optimization," 2020. [Online]. Available: https://arxiv.org/abs/2007.07481
[9] P. Guo, S. Zeng, Y. Wang, H. Fan, F. Wang, and L. Qu, "Selective aggregation for low-rank adaptation in federated learning," 2025. [Online]. Available: https://arxiv.org/abs/2410.01463
[10] Y. Yang, G. Long, T. Shen, J. Jiang, and M. Blumenstein, "Dual-personalizing adapter for federated foundation models," 2024. [Online]. Available: https://arxiv.org/abs/2403.19211
[11] Y. J. Cho, L. Liu, Z. Xu, A. Fahrezi, and G. Joshi, "Heterogeneous lora for federated fine-tuning of on-device foundation models," 2024. [Online]. Available: https://arxiv.org/abs/2401.06432
[12] Y. Byun and J. Lee, "Towards federated low-rank adaptation of language models with rank heterogeneity," 2025. [Online]. Available: https://arxiv.org/abs/2406.17477
[13] S. P. Karimireddy, S. Kale, M. Mohri, S. J. Reddi, S. U. Stich, and A. T. Suresh, "Scaffold: Stochastic controlled averaging for federated learning," 2021. [Online]. Available: https://arxiv.org/abs/1910.06378
[14] X. Li, "Positive-incentive noise," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 6, pp. 8708–8714, 2024.
[15] L. Yi, H. Yu, G. Wang, X. Liu, and X. Li, "pfedlora: Model-heterogeneous personalized federated learning with lora tuning," 2024. [Online]. Available: https://arxiv.org/abs/2310.13283
[16] X.-Y. Liu, R. Zhu, D. Zha, J. Gao, S. Zhong, M. White, and M. Qiu, "Differentially private low-rank adaptation of large language model using federated learning," 2024. [Online]. Available: https://arxiv.org/abs/2312.17493
[17] J. QI, Z. Luan, S. Huang, C. Fung, H. Yang, and D. Qian, "Fdlora: Personalized federated learning of large language model via dual lora tuning," 2024. [Online]. Available: https://arxiv.org/abs/2406.07925
[18] S. Chen, O. Tavallaie, N. Nazemi, and A. Y. Zomaya, "Rbla: Rank-based-lora-aggregation for fine-tuning heterogeneous models in flaas," 2024. [Online]. Available: https://arxiv.org/abs/2408.08699
[19] Z. Wang, Z. Shen, Y. He, G. Sun, H. Wang, L. Lyu, and A. Li, "Flora: Federated fine-tuning large language models with heterogeneous low-rank adaptations," 2024. [Online]. Available: https://arxiv.org/abs/2409.05976
[20] D. A. E. Acar, Y. Zhao, R. M. Navarro, M. Mattina, P. N. Whatmough, and V. Saligrama, "Federated learning based on dynamic regularization," 2021. [Online]. Available: https://arxiv.org/abs/2111.04263
[21] Q. Li, B. He, and D. Song, "Model-contrastive federated learning," 2021. [Online]. Available: https://arxiv.org/abs/2103.16257
[22] T.-M. H. Hsu, H. Qi, and M. Brown, "Federated visual classification with real-world data distribution," 2020. [Online]. Available: https://arxiv.org/abs/2003.08082
[23] M. Duan, D. Liu, X. Chen, R. Liu, Y. Tan, and L. Liang, "Self-balancing federated learning with global imbalanced data in mobile systems," *IEEE*

Transactions on Parallel and Distributed Systems, vol. 32, no. 1, pp. 59–71, 2021.
[24] M. Yang, A. Wong, H. Zhu, H. Wang, and H. Qian, "Federated learning with class imbalance reduction," 2020. [Online]. Available: https://arxiv.org/abs/2011.11266
[25] D. Sarkar, A. Narang, and S. Rai, "Fed-focal loss for imbalanced data classification in federated learning," 2020. [Online]. Available: https://arxiv.org/abs/2011.06283
[26] L. Wang, S. Xu, X. Wang, and Q. Zhu, "Addressing class imbalance in federated learning," 2020. [Online]. Available: https://arxiv.org/abs/2008.06217
[27] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," 2018. [Online]. Available: https://arxiv.org/abs/1708.02002
[28] K. Cao, C. Wei, A. Gaidon, N. Arechiga, and T. Ma, "Learning imbalanced datasets with label-distribution-aware margin loss," 2019. [Online]. Available: https://arxiv.org/abs/1906.07413
[29] J. Wang, W. Zhang, Y. Zang, Y. Cao, J. Pang, T. Gong, K. Chen, Z. Liu, C. C. Loy, and D. Lin, "Seesaw loss for long-tailed instance segmentation," 2021. [Online]. Available: https://arxiv.org/abs/2008.10032
[30] B. Kang, S. Xie, M. Rohrbach, Z. Yan, A. Gordo, J. Feng, and Y. Kalantidis, "Decoupling representation and classifier for long-tailed recognition," 2020. [Online]. Available: https://arxiv.org/abs/1910.09217
[31] X. Shang, Y. Lu, G. Huang, and H. Wang, "Federated learning on heterogeneous and long-tailed data via classifier re-training with federated features," 2022. [Online]. Available: https://arxiv.org/abs/2204.13399
[32] X. Yu, Z. Huang, Y. Xue, and D. Zhu, "Noise-augmented deep neural networks for image classification: Insights from information theory," 2025. [Online]. Available: https://openreview.net/forum?id=o0DQDGaVIY