CALIBRATION IS NOT ENOUGH: SUPPORT RESTORATION AND RESIDUAL CORRECTABILITY IN FEDERATED LONG-TAILED LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Federated Long-Tailed learning (Fed-LT) couples global class imbalance with fragmented client label support. Exact local-prior calibration removes locally missing classes from the client's calibrated softmax competition, even when those classes are observed elsewhere in the federation. Restoring class competition, however, does not eliminate client-update heterogeneity, and the resulting residuals alone do not determine how the server should correct the aggregate update. Therefore, we propose Federated Prior Recalibration and Residual Stabilization (FedPRS). On clients, Federation-Anchored Calibration (FAC) mixes local and federation priors to restore competition over federation-supported classes while retaining local-prior information. On the server, Label-Geometric Regression (LGR) fits composition-associated update residuals and applies a bounded affine correction around FedAvg. Our analysis separates support restoration from predictive utility and shows that server correction depends jointly on affine correctability, directional alignment, and correction scale, while the proposed LGR explicitly controls deviation from the FedAvg update. FedPRS achieves the highest mean Macro-F1 among the reported baselines across the evaluated CIFAR-10/100-LT settings.

1 INTRODUCTION

Federated Learning (FL) (McMahan et al., 2017) enables organizations and clients to train a shared model without centrally pooling their raw data. In realistic federations, however, local statistical heterogeneity, such as label skew, often coincides with severe global long-tailedness (Zhang et al., 2023; Shuai et al., 2022; Shang et al., 2022). As shown in Figure 1 (left), a few head classes dominate the federation-wide sample count, whereas tail classes are scarce and their limited examples may be concentrated on only a few clients. A representative example is federated intrusion detection in Industrial Internet of Things (IIoT) environments, where benign traffic and frequently observed attacks dominate local datasets, while rare industrial threats may be observed only at a few factories or edge sites. Such imbalance makes it difficult for conventional FedAvg (McMahan et al., 2017) and heterogeneous FL algorithms (e.g., FedProx (Li et al., 2020) and FedNTD (Lee et al., 2022)) to learn reliable decision boundaries for global minority classes, as shown in Figure 1 (right).

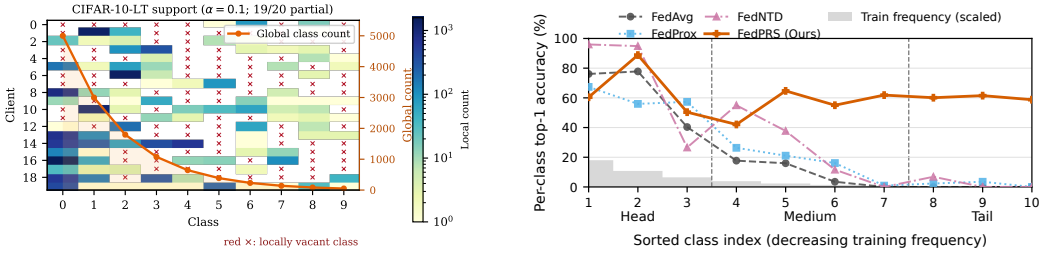


Figure 1: CIFAR-10-LT at IF= 100, $\alpha = 0.1$. Left: local class counts and vacant classes. Right: Last5 per-class recall, where gray shading denotes training frequency.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Federated Long-Tailed learning (Fed-LT) aims to address this coupled regime of global long-tailedness and client-level label heterogeneity. Existing Fed-LT methods address long-tailed heterogeneity through client-side rebalancing (Shuai et al., 2022), classifier retraining (Shang et al., 2022), representation learning with classifier alignment (Xiao et al., 2024), and logit adjustment (Zhang et al., 2022) with feature optimization (Lu et al., 2025) or self-distillation (Yan et al., 2025). Among these approaches, prior-based logit adjustment or calibration (Menon et al., 2021) is a particularly natural response to long-tailed label distributions and federated label skew. During FL local training, given raw logits $\mathbf{z}_k(x)$, client k calibrates its class scores as $\widetilde{\mathbf{z}}_k(x) = \mathbf{z}_k(x) + \log \mathbf{q}_k - \log \boldsymbol{\pi}_\star$, where $\mathbf{q}_k$ is the calibration prior and $\boldsymbol{\pi}_\star$ is the deployment prior. [1] Under an idealized pure label-shift model with complete local support, choosing the exact local training prior $\boldsymbol{\pi}_k$ ($\mathbf{q}_k = \boldsymbol{\pi}_k$) cancels the client-specific prior from the population-optimal raw logits. This identity is useful, but it does not resolve two gaps central to Fed-LT:

- **Fragmented support breaks exact local-prior calibration.** For zero entries in the empirical prior, exact Local-Only calibration is understood as the corresponding masked-softmax limit; $\log 0$ is not evaluated numerically. When $\mathbf{q}_k = \boldsymbol{\pi}_k$, local-prior calibration is population-consistent under pure label shift provided that every class has positive local support. If $\pi_k(c) = 0$, however, class c is excluded from the calibrated softmax competition, yielding zero data-loss gradient for its logit and reducing the logit-Hessian rank from $C - 1$ to $|S_k| - 1$, where S_k denotes the set of classes observed by client k. Replacing the local prior with the global prior, $\mathbf{q}_k = \boldsymbol{\pi}_g$, preserves federation-wide support but no longer exactly cancels the client-specific training prior. This reveals a fundamental trade-off between exact local-prior correction and federation-wide support preservation.
- **Support restoration does not eliminate optimization heterogeneity.** Even when all federation-supported classes remain active in the local objective, client updates may still differ due to class-conditional differences, client-dependent calibration, multi-step local optimization, and stochasticity. Under partial participation, the selected-client aggregate may further deviate from its full-participation counterpart. These residual update discrepancies motivate an additional server-side correction after support restoration.

Therefore, we present Federated Prior Recalibration and Residual Stabilization (FedPRS). (*i*) On the client side, we explicitly disentangle the local training prior $\boldsymbol{\pi}_k$, the federation-level training prior $\boldsymbol{\pi}_g$, and the deployment target $\boldsymbol{\pi}_\star$. This three-prior view leads to a Federation-Anchored Calibration (FAC) strategy that preserves competition over globally supported classes while controlling the departure from exact local-prior correction. (*ii*) On the server side, we propose Label-Geometric Regression (LGR), which fits a label-associated response from finite-round client residuals and uses it to construct a correction toward the federation label center. A radial projection further bounds the correction relative to the FedAvg update. We formulate LGR as a bounded affine aggregation rule, rather than as an estimator guaranteed to recover the full-participation update. The main contributions are summarized as follows:

- We establish a three-prior view of Fed-LT calibration and reveal the trade-off between exact local-prior correction and federation-wide support preservation.
- We propose FedPRS, combining FAC to restore global class competition with LGR for bounded affine server correction.
- We characterize the correction mechanism through affine reachability, correctability, directional alignment, and scale, and derive bounds on its deviation from FedAvg.

2 Related Work

Federated Long-Tailed Learning. Fed-LT methods intervene in local objectives, representations, and classifiers. BalanceFL (Shuai et al., 2022) combines knowledge inheritance with local balancing; CReFF (Shang et al., 2022) learns federated features for server-side classifier retraining. FedGraB (Xiao et al., 2023) balances gradients using estimated global priors, and FedLoGe (Xiao et al., 2024) uses structured classifiers and feature realignment. FedLF (Lu et al., 2025) combines adaptive

[1] Under balanced evaluation, the latter is class-invariant, so the effective adjustment is simply $\log \mathbf{q}_k$.

logit adjustment with class-centred and feature-decorrelation objectives. FedYoYo (Yan et al., 2025) combines self-bootstrap distillation with distribution-aware calibration. More recently, OFLT (Ma et al., 2026) combines discriminative representation learning with adaptive logit calibration under one-shot Fed-LT. We instead study how client-side support restoration interacts with the geometry of subsequent affine server correction.

Calibration and Label-Aware Aggregation. Logit adjustment was first developed for centralized long-tailed learning to correct bias from imbalanced class priors (Menon et al., 2021), and was later extended to federated label skew (Zhang et al., 2022; Wang et al., 2026), e.g., through the class-count-dependent margins in FedLC (Zhang et al., 2022). FedLF (Lu et al., 2025) and FedYoYo (Yan et al., 2025) further bring logit adjustment into Fed-LT. More recent studies have further explored post-hoc calibration in Fed-LT. FedMDD (Wang et al., 2025) calibrates the global decision boundary through model-aware deliberation, and FedNPC (Gao et al., 2026) performs data-free post-hoc classifier calibration using stochastic noise. Our FedPRS differs in both analysis and correction. It introduces a three-prior view for support-aware calibration and further models the residual update discrepancies after support restoration, using LGR for bounded affine server correction. Moreover, FedPALS (Listo Zec et al., 2026) is closely related to LGR, as it optimizes regularized label-aware aggregation weights over the probability simplex. In contrast, our FedPRS couples support-aware client calibration with residual-response modeling and bounded affine server correction, emphasizing how label-distribution geometry can be used to correct the remaining update discrepancies after support restoration.

3 Federated Prior Recalibration and Residual Stabilization

We consider a Fed-LT setting with K clients, where client k holds a local dataset D_k of $N_k > 0$ samples drawn from a heterogeneous label distribution over $C \geq 2$ globally observed classes. The pooled federation exhibits a long-tailed class distribution, while local class priors may differ substantially and may even have incomplete support. FedPRS combines FAC for local class competition with LGR for bounded server correction. Let $N_{k,c}$ denote the number of samples of class c at client k. The pooled empirical prior is $\pi_g(c) = \sum_k N_{k,c} / \sum_k N_k$, which has full support over the federation label space. The deployment prior $\pi_\star$ is also assumed to have full support. Algorithm 1 summarizes the training procedure.

3.1 Prior Cancellation and Support Restoration

Motivation. For a client distribution P_k, calibration prior q, and deployment prior $\pi_\star$, consider

$$\mathcal{R}_k(z; q) = \mathbb{E}_{(X,Y)\sim P_k}\left[-\log \frac{e^{z_Y(X)} q(Y)/\pi_\star(Y)}{\sum_{j=1}^{C} e^{z_j(X)} q(j)/\pi_\star(j)}\right]. \tag{1}$$

Theorem 1 (Three-prior population optimum)**.** *Assume pure label shift with common class-conditional densities $p(x \mid c)$, full-support π_k, q, and $\pi_\star$, and finite, strictly positive $p(x \mid c)$ for every class at $P_{k,X}$-almost every x. Over measurable finite-valued logits, z minimizes $\mathcal{R}_k(z; q)$ if and only if*

$$z^*_{k,c}(x) = \log p(x \mid c) + \log \pi_k(c) - \log q(c) + \log \pi_\star(c) + b_k(x) \tag{2}$$

for $P_{k,X}$-almost every x, where $b_k(x)$ is an arbitrary finite measurable function independent of c.

Proof sketch. Conditional cross-entropy equals the posterior entropy plus a nonnegative KL divergence. Matching the adjusted softmax to the positive client posterior and applying Bayes' rule gives the result (Appendix A.1.1).

Thus $q = \pi_k$ cancels the client-prior term under the theorem's assumptions, whereas $q = \pi_g$ retains $\log(\pi_k/\pi_g)$ when π_g has full support. This population characterization does not select a calibration prior for finite neural-network training, and fragmented local support lies outside its assumptions.

Proposition 1 (Support truncation)**.** *Let $S_k = \{c : \pi_k(c) > 0\}$, assume $\pi_\star$ has full support, and consider exact Local-Only calibration $q = \pi_k$. For finite logits and any local example with $y \in S_k$, the adjusted softmax probability p satisfies*

$$p_c = 0, \qquad \frac{\partial \ell}{\partial z_c} = 0 \quad (c \notin S_k), \qquad \operatorname{rank}\big(\operatorname{diag}(p) - pp^\top\big) = |S_k| - 1. \tag{3}$$

If q instead has full support, then all adjusted probabilities are positive and the logit Hessian has rank $C-1$.

Proof sketch. The cross-entropy gradient is $p - e_y$, while the Hessian quadratic form is a weighted variance over the active support (Appendix A.1.2).

Consequently, exact Local-Only calibration reduces the local objective from a C-class competition problem to an $|S_k|$-class subproblem, leaving federation-valid but locally absent classes unconstrained by the local data loss. Since clients may have different support sets S_k, they effectively optimize different truncated class-competition spaces, creating a structural mismatch with the global C-class task. This motivates introducing federation-level support into the calibration prior while retaining the client-specific prior information, leading to our Federation-Anchored Calibration (FAC).

Federation-Anchored Calibration (FAC). Using the empirical prior $\pi_k(c) = N_{k,c}/N_k$, FAC constructs

$$\widehat{\pi}_k = (1-\lambda)\pi_k + \lambda\pi_g, \qquad \lambda = \frac{\rho}{1+\rho}, \tag{4}$$

and optimizes

$$\mathcal{L}_k^{\mathrm{FAC}}(\theta) = \frac{1}{N_k} \sum_{(x,y)\in D_k} \mathrm{CE}\left(z_\theta(x) + \log\widehat{\pi}_k - \log\pi_\star, y\right). \tag{5}$$

For every federation-supported class c with $\pi_g(c) > 0$, any $\lambda > 0$ ensures $\widehat{\pi}_k(c) > 0$, so locally absent classes remain in the calibrated softmax competition. We set $\rho = 0.1$, corresponding to $\lambda = 1/11$, and use raw logits at inference.

3.2 Residual Heterogeneity and Affine Correctability

Motivation. Let $\Delta_{k,t} = \theta_{k,t}^+ - \theta_t$ denote the trainable-parameter update from client k. For a nonempty selected set $\mathcal{S}_t$, define $w_{k,t} = N_k / \sum_{j\in\mathcal{S}_t} N_j$, $\bar{\Delta}_t = \sum_{k\in\mathcal{S}_t} w_{k,t}\Delta_{k,t}$, and $r_{k,t} = \Delta_{k,t} - \bar{\Delta}_t$. The selected residual subspace is $\mathcal{U}_t = \mathrm{span}\{r_{k,t} : k \in \mathcal{S}_t\}$. For any affine weights satisfying $\sum_{k\in\mathcal{S}_t}\alpha_{k,t} = 1$, the correction relative to FedAvg satisfies

$$C_{\alpha,t} = \sum_{k\in\mathcal{S}_t}(\alpha_{k,t} - w_{k,t})\Delta_{k,t} = \sum_{k\in\mathcal{S}_t}\alpha_{k,t}r_{k,t} \in \mathcal{U}_t. \tag{6}$$

Since $\sum_{k\in\mathcal{S}_t} w_{k,t}r_{k,t} = 0$, we have $\dim\mathcal{U}_t \le |\mathcal{S}_t| - 1$, and conversely every vector in $\mathcal{U}_t$ is reachable by some sum-one affine weighting (Appendix A.2.1). Thus $\mathcal{U}_t$ characterizes where affine aggregation can move relative to FedAvg, but not which reachable direction is beneficial. This directional ambiguity motivates using federation-level label geometry (LGR) to guide the server correction.

Label-Geometric Regression (LGR). Label contrasts provide an observable covariate for composition-associated update disagreement. Rather than assuming a fixed correspondence between classes and parameter directions, LGR estimates the finite-round response of client residuals to deviations from the federation label composition. Let $H_C \in \mathbb{R}^{C\times(C-1)}$ satisfy $H_C^\top H_C = I$ and $H_C^\top \mathbf{1} = 0$, and define $x_k = H_C^\top(\pi_k - \pi_g)$, $\bar{x}_t = \sum_k w_{k,t}x_k$, and $\widetilde{x}_{k,t} = x_k - \bar{x}_t$. We use the local approximation

$$r_{k,t} = B_t\widetilde{x}_{k,t} + \varepsilon_{k,t}, \qquad B_t \in \mathbb{R}^{d_\theta\times(C-1)}, \tag{7}$$

where $\varepsilon_{k,t}$ absorbs disagreement not explained by label composition, including class-conditional heterogeneity, stochasticity, and multi-step local dynamics. This approximation motivates the regression but is not assumed to be exact.

Define $\Sigma_{x,t} = \sum_k w_{k,t}\widetilde{x}_{k,t}\widetilde{x}_{k,t}^\top$, $\Sigma_{rx,t} = \sum_k w_{k,t}r_{k,t}\widetilde{x}_{k,t}^\top$, and $r_{x,t} = \max\{\mathrm{tr}(\Sigma_{x,t})/(C-1), 10^{-12}\}$. Weighted ridge regression gives

$$\widehat{B}_t = \Sigma_{rx,t}(\Sigma_{x,t} + r_{x,t}I)^{-1}, \qquad C_t^0 = -\widehat{B}_t\bar{x}_t, \tag{8}$$

where C_t^0 is the fitted response toward the federation label center, not an estimate guaranteed to recover the full-participation update. Equivalently,

$$d_t = -(\Sigma_{x,t} + r_{x,t}I)^{-1}\bar{x}_t, \qquad \alpha_{k,t} = w_{k,t}(1 + \widetilde{x}_{k,t}^\top d_t), \tag{9}$$

with $\sum_k \alpha_{k,t} = 1$ and $C_t^0 = \sum_k (\alpha_{k,t} - w_{k,t})\Delta_{k,t} \in \mathcal{U}_t$. The coefficients may be negative, so LGR defines an affine rather than convex reweighting. Appendix A.3.1 derives the regression identity, and Appendix A.3.2 establishes its connection to regularized label matching.

Bounded realization. To limit deviation from FedAvg, FedPRS projects the candidate correction onto a relative-radius ball,

$$C_t = s_t C_t^0, \qquad s_t = \min\left\{1, \frac{\rho_s \|\bar{\Delta}_t\|_2}{\|C_t^0\|_2}\right\}, \tag{10}$$

with $s_t = 1$ and $C_t = 0$ when $C_t^0 = 0$, and updates

$$\theta_{t+1} = \theta_t + \bar{\Delta}_t + C_t. \tag{11}$$

We set $\rho_s = 0.25$. The bound controls correction magnitude but does not guarantee a favorable direction or downstream improvement. Only trainable parameters receive the correction, while buffers follow FedAvg; beyond the one-time class-count exchange, no additional client message is required.

Computation and communication. Let $M = |\mathcal{S}_t|$ and let P denote the number of trainable parameters. Once the label contrasts are available, LGR requires $O(MC^2+C^3)$ server-side arithmetic to form the label-space covariance and solve the $(C-1)$-dimensional system, together with $O(MP)$ for the additional parameter-vector combination. The coefficient form avoids constructing either a $P \times P$ covariance matrix or the $P \times (C-1)$ response matrix. FAC only adds a fixed C-dimensional logit adjustment to local training and requires no additional forward pass or network parameters. Communication adds only a one-time C-dimensional class-count vector from each client and the corresponding federation-prior broadcast, with no additional per-round client message beyond the standard model exchange. These are asymptotic operation and communication counts rather than measured wall-clock or memory improvements.

Algorithm 1 Training process of FedPRS

Input: Initial global model θ_0, training rounds T, local epochs E, ρ, ρ_s, target prior $\pi_\star$
Output: Global model θ_T

Clients report class counts; server computes and broadcasts π_g
Clients construct $\widehat{\pi}_k$ by Eq. (4); server stores $x_k = H_C^\top(\pi_k - \pi_g)$
for $t = 0$ to $T-1$ **do**
    Server randomly selects active clients $\mathcal{S}_t$ and broadcasts θ_t
// Clients execute:
    for $k \in \mathcal{S}_t$ **in parallel do**
        Initialize from θ_t and train for E local epochs using Eq. (5)
        Send the updated model $\theta_{k,t}^+$ to the server
    end for
// Server executes:
    Compute $\Delta_{k,t}$, $w_{k,t}$, the FedAvg update $\bar{\Delta}_t$, d_t and $\alpha_{k,t}$
    Form $C_t^0 = \sum_{k\in\mathcal{S}_t}(\alpha_{k,t} - w_{k,t})\Delta_{k,t}$, and project C_t^0 by Eq. (10) to obtain C_t
    Update θ_{t+1} by Eq. (11)
end for
return θ_T

4 Theoretical Analysis

4.1 Support and Prior-Space Distortion

For $q_k^\lambda = (1-\lambda)\pi_k + \lambda\pi_g$, full-support π_g, and $0 < \lambda \le 1$,

$$q_k^\lambda(c) \ge \lambda\pi_g(c) > 0, \qquad \|q_k^\lambda - \pi_k\| = \lambda\|\pi_g - \pi_k\|. \tag{12}$$

The second identity holds for every norm. Thus any positive mixture strength restores positive calibration mass to every federation-supported class, while the displacement from the Local-Only prior grows linearly with λ. Under full-support calibration, Proposition 1 further implies logit-Hessian rank $C-1$ at finite logits. This rank statement characterizes local class competition only and does not bound conditioning or imply convexity in network parameters.

With $\omega_k = N_k / \sum_j N_j$, the federation prior $\pi_g = \sum_k \omega_k \pi_k$ uniquely minimizes $\sum_k \omega_k \|\pi_k - a\|_2^2$ and $\sum_k \omega_k \mathrm{KL}(\pi_k \| a)$ over common anchors a. For any fixed $\lambda > 0$, it also minimizes the average squared prior displacement induced by the mixture, as shown in Appendix A.1.3.

4.2 Affine Correctability, Direction, and Scale

Suppress the round index. For analysis, let Δ_{all} denote the full-participation update obtained from the same round-start model and local-training rule, weighted by ω_k, and define the matched-reference discrepancy $e = \bar{\Delta} - \Delta_{\mathrm{all}}$. When the reference contains the same realized selected-client updates, e isolates the participation discrepancy for that realization. With independent realizations of these updates, e may also contain local-training randomness. In either case, Δ_{all} is an algorithmic reference rather than a globally optimal update or population gradient.

For $\mathcal{U} = \mathrm{span}\{r_k : k \in \mathcal{S}\}$ and orthogonal projector $P_{\mathcal{U}}$,

$$C^* = \arg\min_{C \in \mathcal{U}} \|e + C\|_2^2 = -P_{\mathcal{U}} e, \qquad \min_{C \in \mathcal{U}} \|e + C\|_2^2 = \|(I - P_{\mathcal{U}})e\|_2^2. \tag{13}$$

Thus C^* is the best correction reachable within the selected residual subspace relative to this matched reference. It does not impose the LGR correction budget.

For $e \neq 0$, define $\mathrm{CRF} = \|P_{\mathcal{U}} e\|_2^2 / \|e\|_2^2 \in [0, 1]$. For $C^* \neq 0$ and $C \in \mathcal{U} \setminus \{0\}$, define

$$A = \cos(C, C^*), \qquad \eta = \frac{\|C\|_2}{\|C^*\|_2}, \qquad D = \|e\|_2^2 - \|e + C\|_2^2. \tag{14}$$

Let $\mathrm{RSE} = D / \|C^*\|_2^2$ and $\mathrm{LER} = D / \|e\|_2^2$.

Theorem 2 (Alignment–scale decomposition)**.** *Under the above nondegeneracy conditions,*

$$\mathrm{RSE} = 2\eta A - \eta^2, \qquad \mathrm{LER} = \mathrm{CRF}\,(2\eta A - \eta^2). \tag{15}$$

The squared matched-reference error decreases if and only if $A > 0$ and $0 < \eta < 2A$.

Proof sketch. Since $C \in \mathcal{U}$, $\langle e, C \rangle = -\langle C^*, C \rangle$. Expanding the squared norm gives $D = 2\langle C^*, C \rangle - \|C\|_2^2$, which yields the result. The full proof is given in Appendix A.2.2.

For fixed $A > 0$, $\mathrm{RSE} = A^2 - (\eta - A)^2$ is maximized at $\eta = A$. Radial rescaling adjusts the correction scale without changing alignment, so it can prevent overshooting but cannot fix a nonpositive direction. CRF measures reachable discrepancy, while identifying a beneficial correction direction requires additional structure or information; see Appendix A.2.3.

4.3 Bounded Update Deviation

Proposition 2 (Projection and update-deviation bounds)**.** *Let $C_t^0 \in \mathcal{U}_t$ and $\bar{\Delta}_t$ be finite and let $\rho_s \geq 0$. Eq.* (10) *is the Euclidean projection of C_t^0 onto the ball $\{C : \|C\|_2 \leq \rho_s \|\bar{\Delta}_t\|_2\}$ and therefore remains in $\mathcal{U}_t$. If $0 \leq \rho_s < 1$, then*

$$(1 - \rho_s)\|\bar{\Delta}_t\|_2 \leq \|\bar{\Delta}_t + C_t\|_2 \leq (1 + \rho_s)\|\bar{\Delta}_t\|_2. \tag{16}$$

For $\bar{\Delta}_t \neq 0$,

$$\cos(\bar{\Delta}_t, \bar{\Delta}_t + C_t) \geq \sqrt{1 - \rho_s^2}. \tag{17}$$

Proof sketch. Projection onto a Euclidean ball is radial. The norm bounds follow from the triangle and reverse-triangle inequalities, while decomposing the correction into components parallel and orthogonal to $\bar{\Delta}_t$ gives the angle bound. Details are provided in Appendix A.3.3.

For $\bar{\Delta}_t \neq 0$ and $\rho_s = 0.25$, the LGR correction satisfies $\|C_t\|_2 \leq 0.25 \|\bar{\Delta}_t\|_2$, yielding an angular deviation from the FedAvg update of at most $\arcsin(0.25) \approx 14.5°$. Therefore, LGR provides a bounded perturbation to FedAvg in both magnitude and direction.

5 Experiments

Experimental setup. We evaluate CIFAR-10/100-LT with imbalance factor $\mathrm{IF} \in \{10, 50, 100\}$ and class-wise Dirichlet concentration $\alpha \in \{0.1, 0.5\}$. IF is the largest-to-smallest training-class count

ratio, where smaller α generally yields more concentrated client label distributions. Unless varied, we use CIFAR-style ResNet32, 20 clients, 4 participants per round, 5 local epochs, batch size 64, and 200 rounds. Methods share data partitions, base initialization, and participation schedules within each comparison. We compare FedAvg (McMahan et al., 2017), CReFF (Shang et al., 2022), FedLF (Lu et al., 2025), FedYoYo (Yan et al., 2025), FedPALS (Listo Zec et al., 2026), and our FedPRS with $(\rho, \rho_s) = (0.1, 0.25)$. We report CReFF on CIFAR-10-LT only. CIFAR-100-LT results are omitted because we did not obtain a sufficiently reliable adaptation of its server-side classifier-retraining procedure under the present experimental scaffold. Appendix B.1 gives the necessary settings.

We report Last5 averages and mean $\pm$ sample standard deviation over three seeds. We adopt Macro-F1 as the primary metric because it equally weights all classes and is therefore less dominated by head classes under long-tailed distributions, while jointly reflecting per-class precision and recall. Head/medium/tail recall averages classes ordered by training frequency, split $3/4/3$ on CIFAR-10 and $33/34/33$ on CIFAR-100. All tests use balanced data, and FedPRS predicts from raw logits. For the balanced evaluation setting, we use the uniform deployment prior $\pi_\star(c) = 1/C$ for all classes.

Main results. Tables 1-2 report Macro-F1 over the twelve imbalance-heterogeneity settings. FedPRS achieves the highest mean Macro-F1 among the reported baselines in all twelve main settings. At IF$= 100$, $\alpha = 0.1$, it achieves 0.521 ± 0.069 on CIFAR-10-LT and 0.318 ± 0.011 on CIFAR-100-LT. The strongest reported baselines attain 0.437 ± 0.070 and 0.247 ± 0.015, respectively. Appendix B.2 reports the complete metrics, where FedPRS also leads mean accuracy and medium-class recall in all twelve settings, although head- and tail-recall leaders vary across reported methods and conditions.

Table 1: CIFAR-10-LT Last5 Macro-F1 across imbalance and heterogeneity settings.

	$\alpha = 0.1$			$\alpha = 0.5$		
Method	IF=100	IF=50	IF=10	IF=100	IF=50	IF=10
FedAvg	0.129±0.033	0.145±0.116	0.415±0.062	0.544±0.022	0.646±0.012	0.799±0.008
CReFF	0.437±0.070	0.471±0.049	0.643±0.022	0.620±0.005	0.666±0.020	0.772±0.006
FedLF	0.317±0.040	0.393±0.048	0.586±0.055	0.594±0.035	0.629±0.080	0.771±0.009
FedYoYo	0.339±0.074	0.396±0.068	0.633±0.034	0.511±0.021	0.616±0.028	0.781±0.005
FedPALS	0.327±0.024	0.368±0.021	0.654±0.054	0.589±0.025	0.667±0.015	0.816±0.007
FedPRS	**0.521±0.069**	**0.601±0.075**	**0.714±0.042**	**0.734±0.005**	**0.771±0.007**	**0.838±0.002**

Table 2: CIFAR-100-LT Last5 Macro-F1 across imbalance and heterogeneity settings.

	$\alpha = 0.1$			$\alpha = 0.5$		
Method	IF=100	IF=50	IF=10	IF=100	IF=50	IF=10
FedAvg	0.239±0.016	0.298±0.012	0.437±0.019	0.280±0.010	0.340±0.021	0.500±0.012
FedLF	0.166±0.009	0.206±0.008	0.308±0.017	0.226±0.006	0.279±0.011	0.405±0.003
FedYoYo	0.224±0.009	0.259±0.006	0.375±0.019	0.282±0.014	0.321±0.003	0.431±0.025
FedPALS	0.247±0.015	0.307±0.009	0.449±0.005	0.291±0.009	0.341±0.003	0.496±0.009
FedPRS	**0.318±0.011**	**0.368±0.007**	**0.475±0.010**	**0.358±0.019**	**0.413±0.013**	**0.525±0.004**

Effects of imbalance and client heterogeneity. At each fixed α, FedPRS's mean Macro-F1 increases as IF decreases from 100 to 10. Its margin over the strongest reported comparator nevertheless varies with the setting. At $\alpha = 0.1$, the margins are approximately $0.084/0.130/0.060$ on CIFAR-10-LT and $0.071/0.061/0.026$ on CIFAR-100-LT for IF$= 100/50/10$, respectively. Thus FedPRS remains first in mean Macro-F1 across the evaluated imbalance levels, although its margin over the strongest comparator is not monotonic. At IF$= 10$, $\alpha = 0.5$, the margins narrow to 0.022 and 0.025. On the other hand, increasing α from 0.1 to 0.5 improves FedPRS's mean Macro-F1 at every fixed IF. These comparisons indicate sensitivity to client label allocation, not an isolated effect of vacant classes. Changing α also changes class proportions and client sample sizes, while changing IF alters both minority-class evidence and the total training-set size.

Training dynamics and class-group trade-offs. Figure 2 reports the full evaluation trajectories. FedPRS retains a late-round advantage in the plotted aggregate metrics. Figure 3 provides descriptive single-seed class profiles. It shows that the methods exhibit distinct head-medium-tail trade-offs across the two datasets. FedPRS attains stronger medium-class recall in the plotted settings, while the head- and tail-recall leaders vary across methods. These results complement the aggregate Macro-F1 comparison.

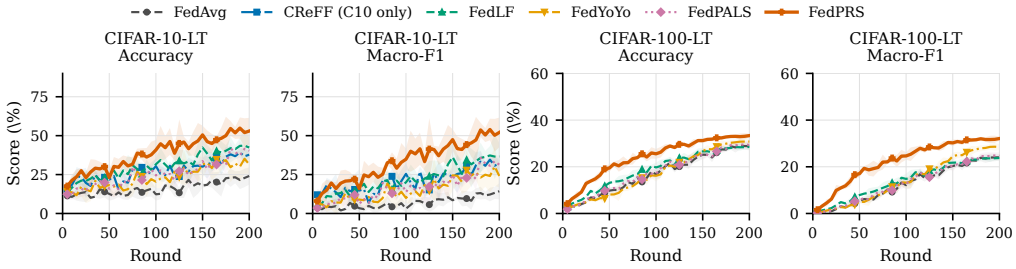


Figure 2: Accuracy and Macro-F1 trajectories on CIFAR-10/100-LT (IF=100, $\alpha = 0.1$).

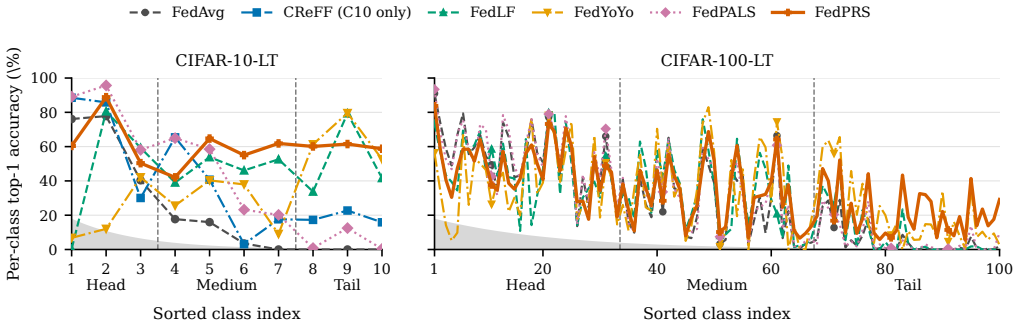


Figure 3: Last5 per-class recall on CIFAR-10/100-LT (IF=100, $\alpha = 0.1$).

Component ablation. Table 3 separates FAC and LGR on CIFAR-10-LT. FAC uses FedAvg aggregation, while LGR only uses raw cross-entropy. FAC accounts for most of the improvement, raising Macro-F1 from 0.129 to 0.476. Adding LGR further increases Macro-F1 to 0.521 and tail recall from 0.414 to 0.447, whereas LGR alone reaches only 0.204 Macro-F1. Appendix B.2.1 shows that this additional mean gain varies across seeds.

Table 3: Last5 component ablation on CIFAR-10-LT (IF=100, $\alpha = 0.1$).

Method	Accuracy	Macro-F1	Head	Medium	Tail
FedAvg	0.229 ± 0.045	0.129 ± 0.033	0.653 ± 0.073	0.084 ± 0.059	0.000 ± 0.000
FAC	0.495 ± 0.064	0.476 ± 0.074	0.645 ± 0.080	0.444 ± 0.094	0.414 ± 0.075
LGR only	0.309 ± 0.056	0.204 ± 0.057	$\mathbf{0.699 \pm 0.046}$	0.235 ± 0.104	0.017 ± 0.030
FedPRS	$\mathbf{0.533 \pm 0.064}$	$\mathbf{0.521 \pm 0.069}$	0.679 ± 0.072	$\mathbf{0.487 \pm 0.075}$	$\mathbf{0.447 \pm 0.150}$

Client-participation sensitivity. Table 4 varies the number of participating clients M with 20 total clients and a fixed 200-round budget. FedPRS's mean Macro-F1 increases from $M = 2$ to $M = 12$ and then declines, while its margin over FAC generally narrows as participation increases. Under full participation, the LGR correction becomes numerically zero, consistent with the label-center identity. Complete predictive and mechanism statistics are reported in Appendix B.3.

Table 4: Participation-rate sensitivity on CIFAR-10-LT (IF=100, $\alpha = 0.1$).

M	Rate	FedPRS Acc.	FedPRS F1	ΔF1 vs. FAC	Budget Active
2	10%	0.439 ± 0.084	0.395 ± 0.105	$+0.080$	60.8%
4	20%	0.533 ± 0.064	0.521 ± 0.069	$+0.045$	64.5%
8	40%	0.550 ± 0.066	0.538 ± 0.070	$+0.039$	56.3%
12	60%	$\mathbf{0.568 \pm 0.049}$	$\mathbf{0.557 \pm 0.056}$	$+0.019$	37.8%
16	80%	0.533 ± 0.095	0.515 ± 0.106	$+0.037$	17.2%
20	100%	0.516 ± 0.126	0.494 ± 0.149	-0.010	0.0%

Cross-backbone evaluation. Table 5 shows that, on CIFAR-10-LT at IF= 100, $\alpha = 0.1$, FedPRS has the highest mean Macro-F1 on ResNet32, ResNet18, and WRN-28-2. CReFF is the strongest external baseline on the first two backbones and FedYoYo on WRN-28-2. FedPALS leads head recall, FedYoYo leads tail recall, and FedPRS leads medium recall on all three. These results support transfer across the tested convolutional backbones, not architecture independence. Appendix B.4 reports all five metrics. The Macro-F1 margin over the strongest baseline remains 0.063-0.084 across backbones.

Table 5: Cross-backbone Last5 accuracy and Macro-F1 on CIFAR-10-LT (IF=100, $\alpha = 0.1$).

	ResNet32		ResNet18		WRN-28-2	
Method	Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
FedAvg	0.229 ± 0.045	0.129 ± 0.033	0.318 ± 0.087	0.220 ± 0.090	0.270 ± 0.038	0.169 ± 0.036
CReFF	0.462 ± 0.053	0.437 ± 0.070	0.523 ± 0.039	0.504 ± 0.048	0.497 ± 0.036	0.468 ± 0.049
FedLF	0.373 ± 0.052	0.317 ± 0.040	0.449 ± 0.033	0.410 ± 0.049	0.442 ± 0.028	0.398 ± 0.049
FedYoYo	0.395 ± 0.055	0.339 ± 0.074	0.476 ± 0.048	0.428 ± 0.051	0.523 ± 0.032	0.482 ± 0.030
FedPALS	0.401 ± 0.033	0.327 ± 0.024	0.475 ± 0.044	0.434 ± 0.056	0.461 ± 0.029	0.414 ± 0.026
FedPRS	$\mathbf{0.533 \pm 0.064}$	$\mathbf{0.521 \pm 0.069}$	$\mathbf{0.594 \pm 0.024}$	$\mathbf{0.586 \pm 0.022}$	$\mathbf{0.555 \pm 0.055}$	$\mathbf{0.545 \pm 0.058}$

Local-optimization-budget sensitivity. Figure 4 varies $E \in \{1, 3, 5, 7, 10\}$ on CIFAR-10-LT at a fixed 200-round communication budget. FedPRS achieves the highest mean accuracy and Macro-F1 at all five budgets, with Macro-F1 increasing from 0.261 to 0.560. The strongest comparator shifts from CReFF at smaller budgets to FedYoYo at $E = 10$, while mean tail-recall ordering remains method dependent. These results indicate robustness to local-computation budget without implying uniform class-group dominance. Complete CIFAR-10/100-LT metrics are provided in Appendix B.5. The CIFAR-100-LT sweep in Appendix B.5.2 also preserves FedPRS's Macro-F1 lead at every budget, although its margin over FedYoYo narrows to 0.011 at the largest budget.

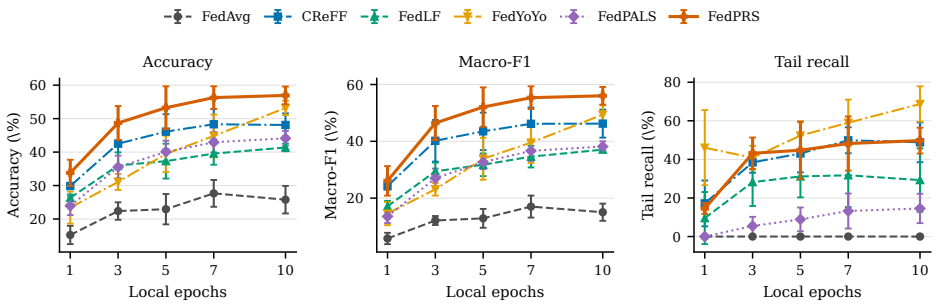


Figure 4: CIFAR-10-LT sensitivity to local epochs (IF=100, $\alpha = 0.1$; 200 rounds).

6 Conclusion

FedPRS couples Federation-Anchored Calibration (FAC) with Label-Geometric Regression (LGR) for Fed-LT. FAC restores competition over federation-supported classes while retaining local-prior information, whereas LGR uses label geometry to construct a bounded affine correction around FedAvg. Our analysis separates structural support restoration from predictive utility and decomposes server-side correction into correctability, directional alignment, and scale. Under the evaluated CIFAR-10/100-LT protocols, FedPRS achieves the highest mean Macro-F1 among the reported baselines in all twelve main settings and across the three tested CIFAR-10-LT backbones. Component ablations further show that FAC accounts for most of the improvement, while the additional benefit of LGR varies across seeds.

Two limitations remain. First, the theoretical guarantees are structural and geometric: support restoration does not provide positive supervision for locally absent classes, and the correction bounds do not imply convergence or downstream-risk improvement. Second, the current evidence is limited to CIFAR-10/100-LT, convolutional backbones, and three random seeds, so broader datasets and additional-seed validation are needed to assess generality.

AI USE STATEMENT

Generative AI tools were used to assist with the retrieval and discovery of related work and to aid and polish the writing of this manuscript. The authors verified the relevance and accuracy of the cited literature and reviewed and revised all AI-assisted text. The authors take full responsibility for the final content of this submission.

REPRODUCIBILITY STATEMENT

Appendix A provides the complete proofs and states the assumptions used by the theoretical results. Appendix B documents optimization and preprocessing settings, method-specific hyperparameters, evaluation rules, and complete experimental results. The anonymous supplementary code provides the method and baseline implementations, experiment configurations, fixed data and participation manifests, and scripts for reproducing the reported experiments. The three random seeds (42, 43, and 44) are identified in Appendix B and the supplementary configurations; Section 5 specifies the aggregation used for the summary statistics.

REFERENCES

Jintong Gao, He Zhao, Yibo Yang, and Dandan Guo. FedNPC: Stochastic noise-driven post-hoc classifier calibration method for federated long-tailed learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Findings (CVPRF)*, pp. 7737–7746, 2026.

Gihun Lee, Minchan Jeong, Yongjin Shin, Sangmin Bae, and Se-Young Yun. Preservation of the global knowledge by not-true distillation in federated learning. In *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022 (NeurIPS)*, volume 35, pp. 38461–38474, 2022.

Tian Li, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. In *Proceedings of the Third Conference on Machine Learning and Systems (MLsys)*, 2020.

Edvin Listo Zec, Adam Breitholtz, and Fredrik D. Johansson. Overcoming label shift in target-aware federated learning. *Transactions on Machine Learning Research*, 2026.

Xiuhua Lu, Peng Li, and Xuefeng Jiang. FedLF: Adaptive logit adjustment and feature optimization in federated long-tailed learning. In *Proceedings of the 16th Asian Conference on Machine Learning (ACML)*, volume 260 of *Proceedings of Machine Learning Research*, pp. 303–318, 2025.

Yinglong Ma, Huili Liu, and Yandan Wang. OFLT: One-shot federated long-tailed learning with discriminative representation learning and adaptive logit calibration. *IEEE Internet of Things Journal*, 13(10):21005–21017, 2026.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Proceedings of AISTATS*, volume 54 of *PMLR*, pp. 1273–1282, 2017.

Aditya Krishna Menon, Sadeep Jayasumana, Ankit Singh Rawat, Himanshu Jain, Andreas Veit, and Sanjiv Kumar. Long-tail learning via logit adjustment. In *9th International Conference on Learning Representations (ICLR)*, 2021.

Xinyi Shang, Yang Lu, Gang Huang, and Hanzi Wang. Federated learning on heterogeneous and long-tailed data via classifier re-training with federated features. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 2218–2224, 2022.

Xian Shuai, Yulin Shen, Siyang Jiang, Zhihe Zhao, Zhenyu Yan, and Guoliang Xing. BalanceFL: Addressing class imbalance in long-tail federated learning. In *Proceedings of the ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, pp. 271–284, 2022.

Yingchao Wang, Wenqi Niu, and Hanpo Hou. FedAKD: Federated adaptive knowledge distillation via global knowledge calibration and decoupling. In *Proceedings of the ACM Web Conference 2026 (WWW)*, pp. 5415–5426, 2026. doi: 10.1145/3774904.3792386.

Yiwen Wang, Jiaxin Li, Heye Zhang, Jingfeng Zhang, Feng Wan, Anqi Qiu, and Zhifan Gao. FedMDD: Multi-deliberation based calibration for federated long-tailed learning. *Knowledge-Based Systems*, 323:113741, 2025.

Zikai Xiao, Zihan Chen, Songshang Liu, Hualiang Wang, Yang Feng, Jin Hao, Joey Tianyi Zhou, Jian Wu, Howard Hao Yang, and Zuozhu Liu. Fed-GraB: Federated long-tailed learning with self-adjusting gradient balancer. In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023 (NeurIPS)*, volume 36, 2023.

Zikai Xiao, Zihan Chen, Liyinglan Liu, Yang Feng, Joey Tianyi Zhou, Jian Wu, Wanlu Liu, Howard Yang, and Zuozhu Liu. FedLoGe: Joint local and generic federated learning under long-tailed data. In *The Twelfth International Conference on Learning Representations (ICLR)*, 2024.

Shanshan Yan, Zexi Li, Chao Wu, Meng Pang, Yang Lu, Yan Yan, and Hanzi Wang. You are your own best teacher: Achieving centralized-level performance in federated learning under heterogeneous and long-tailed data. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 2750–2759, 2025.

Jie Zhang, Zhiqi Li, Bo Li, Jianghe Xu, Shuang Wu, Shouhong Ding, and Chao Wu. Federated learning with label distribution skew via logits calibration. In *Proceedings of the 39th International Conference on Machine Learning (ICML)*, volume 162 of *Proceedings of Machine Learning Research*, pp. 26311–26329, 2022.

Yifan Zhang, Bingyi Kang, Bryan Hooi, Shuicheng Yan, and Jiashi Feng. Deep long-tailed learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9):10795–10816, 2023.

A Proofs and Theoretical Boundaries

All update-space statements use Euclidean geometry on the trainable parameters, and parameter-space bounds do not imply deployment-risk guarantees.

A.1 Calibration and Support

A.1.1 Three-prior population optimum

Proof of Theorem 1. Let $\eta_c(x) = P_k(Y = c \mid X = x)$ and let $p_z(x)$ be the adjusted softmax. Conditional cross-entropy gives

$$\mathcal{R}_k(z; q) = \mathbb{E}_X H(\eta(X)) + \mathbb{E}_X \mathrm{KL}(\eta(X) \| p_z(X)). \tag{18}$$

The entropy term is finite, bounded by $\log C$. Positivity of η_c makes $z_c = \log \eta_c - \log q(c) + \log \pi_\star(c)$ a measurable, finite-valued realization almost everywhere; values on the exceptional null set can be chosen arbitrarily. The minimum is therefore attained. Nonnegativity of KL implies that a minimizer must satisfy $p_z = \eta$ almost everywhere, and that this condition is sufficient.

Two finite score vectors have the same softmax exactly when they differ by a class-independent shift. Bayes' rule gives

$$\eta_c(x) = \frac{p(x \mid c)\pi_k(c)}{\sum_j p(x \mid j)\pi_k(j)}.$$

Substitution yields Eq. (2), with the log normalizer absorbed into $b_k(x)$. Conversely, substituting the stated logits recovers η and attains the entropy lower bound. □

With $q = \pi_k$, the raw-logit rule is the Bayes rule for a target with prior $\pi_\star$ and the same class-conditional distributions. With $q = \pi_g$, it retains $\log(\pi_k/\pi_g)$, which is class-independent only when the two normalized priors coincide. Without label shift, replacing $p(x \mid c)$ by $p_k(x \mid c)$ shows that prior cancellation does not remove class-conditional heterogeneity.

The assumptions matter. If a posterior coordinate is zero on a set of positive marginal probability while q has full support, posterior matching requires an extended-logit limit rather than finite logits. If $q(c) = 0 < \pi_k(c)$, the risk is infinite. If $\pi_k(c) = q(c) = 0$ under exact masking, the absent-class logit is not identified by the local objective. None of these boundary cases permits direct application of the full-support theorem to an empirical zero-count class.

A.1.2 Support truncation and Hessian rank

Proof of Proposition 1. For finite logits and full-support $\pi_\star$, $p_c \propto e^{z_c} q(c)/\pi_\star(c)$ is positive exactly on the nonempty support S of q. An example with $y \in S$ has gradient $p - e_y$, whose absent-class coordinates vanish. Its logit Hessian is $H = \mathrm{diag}(p) - pp^\top$, and

$$v^\top H v = \frac{1}{2} \sum_{i,j \in S} p_i p_j (v_i - v_j)^2. \tag{19}$$

This quadratic form vanishes exactly when v is constant on S; inactive coordinates are unrestricted. Hence $\dim \ker H = C - |S| + 1$ and $\mathrm{rank}\, H = |S| - 1$. This includes $|S| = 1$, where $H = 0$, and full support, where the rank is $C - 1$. □

For independent affine classifier rows $z_c = W_c^\top h + b_c$, the chain rule also gives zero data-loss gradients for absent-class W_c, b_c under exact masking. This does not cover parameter tying, shared-feature changes, or regularization. Under a positive anchor, an absent class instead receives logit derivative $p_c > 0$: its score is suppressed in the absence of positive examples. Restoring competition is not recovery of missing class-conditional information.

A.1.3 Prior-space barycenters

Let $\omega_k > 0$, $\sum_k \omega_k = 1$, and $\bar{\pi} = \sum_k \omega_k \pi_k$. For any anchor a in the probability simplex,

$$\sum_k \omega_k \|\pi_k - a\|_2^2 = \sum_k \omega_k \|\pi_k - \bar{\pi}\|_2^2 + \|a - \bar{\pi}\|_2^2.$$

The cross term vanishes by centering, so $\bar{\pi}$ uniquely minimizes the objective. For fixed $\lambda > 0$, the squared mixture displacement is λ^2 times this objective and has the same minimizer. At $\lambda = 0$ all anchors give zero displacement.

Using the convention $0 \log(0/a) = 0$,

$$\sum_k \omega_k \mathrm{KL}(\pi_k \| a) = \sum_k \omega_k \mathrm{KL}(\pi_k \| \bar{\pi}) + \mathrm{KL}(\bar{\pi} \| a).$$

The identity holds whenever the terms are finite; if a vanishes on positive mass of $\bar{\pi}$, the objective is infinite. Gibbs' inequality gives the unique minimizer $\bar{\pi}$, including on the boundary of the simplex. Data weights give $\bar{\pi} = \pi_g$. This KL result concerns the anchor itself, not $\sum_k \omega_k \mathrm{KL}(\pi_k \| (1-\lambda)\pi_k + \lambda a)$, and neither objective is a classification risk.

A.2 Affine Correction and Reference Geometry

A.2.1 Centering, references, and reachability

For $r_k = \Delta_k - \bar{\Delta}$ and $\bar{\Delta} = \sum_k w_k \Delta_k$,

$$\sum_k w_k r_k = 0, \qquad \sum_k w_k \|\Delta_k\|_2^2 = \|\bar{\Delta}\|_2^2 + \sum_k w_k \|r_k\|_2^2.$$

For $\sum_k \alpha_k = 1$, expanding $\Delta_k = \bar{\Delta} + r_k$ proves Eq. (6). Conversely, for $C = \sum_k \beta_k r_k \in \mathcal{U}$, set $\alpha_k = \beta_k + (1 - \sum_j \beta_j) w_k$. Then $\sum_k \alpha_k = 1$ and $\sum_k \alpha_k r_k = C$. Thus $\mathcal{U}$ is precisely the affine-reachable correction space. Weighted centering is a nontrivial column dependence, implying $\dim \mathcal{U} \le |\mathcal{S}| - 1$.

Write $\Omega_{\mathcal{S}} = \sum_{k \in \mathcal{S}} \omega_k$. If selected realized updates are reused and $0 < \Omega_{\mathcal{S}} < 1$, the reference decomposes as

$$\Delta_{\mathrm{all}} = \Omega_{\mathcal{S}} \bar{\Delta} + (1 - \Omega_{\mathcal{S}}) \bar{\Delta}_{\overline{\mathcal{S}}}, \qquad e = (1 - \Omega_{\mathcal{S}})(\bar{\Delta} - \bar{\Delta}_{\overline{\mathcal{S}}}).$$

Here the unselected average uses weights $\omega_k / (1 - \Omega_{\mathcal{S}})$. Full participation with the same realized updates gives $e = 0$. If the reference uses independent realizations of the selected-client updates, it includes an additional stochastic discrepancy; the error is no longer exclusively due to participation.

A.2.2 Oracle projection and alignment–scale decomposition

Proof of Eq. (13) *and Theorem 2.* Put $P = P_{\mathcal{U}}$. For every $C \in \mathcal{U}$, orthogonality gives

$$\|e + C\|_2^2 = \|(I - P)e\|_2^2 + \|Pe + C\|_2^2.$$

The second term has the unique minimum zero at $C^* = -Pe$, proving the projection claim. Since $\langle e, C \rangle = -\langle C^*, C \rangle$,

$$D = 2\langle C^*, C \rangle - \|C\|_2^2 = \|C^*\|_2^2 - \|C - C^*\|_2^2. \tag{20}$$

The first equality holds without nondegeneracy assumptions. When $C, C^* \neq 0$, substitution of A and η yields $D = \|C^*\|_2^2 (2\eta A - \eta^2)$; division proves the RSE and LER formulas. Since $\eta > 0$, $D > 0$ is equivalent to $A > 0$ and $0 < \eta < 2A$. Completing the square gives $2\eta A - \eta^2 = A^2 - (\eta - A)^2$. □

Equation (20) also gives $\mathrm{RSE} \le 1$, with equality only at $C = C^*$ when $C^* \neq 0$. If $C = 0$, reduction is zero and alignment is undefined. If $Pe = 0$, every nonzero $C \in \mathcal{U}$ increases error by $\|C\|^2$; exact RSE is undefined. If $e = 0$, exact CRF and LER are undefined.

For $C = sC^0$ with $s > 0$, the scale cancels in the alignment cosine and $\eta(C) = s\eta(C^0)$. For fixed positive alignment, the best scale over $s \in [0, 1]$ is $\min\{1, A/\eta(C^0)\}$ without an additional norm budget. With nonpositive alignment, every nonzero positive rescaling is harmful to reference error.

The ratios above use exact, nonzero denominators. Adding numerical stabilizers generally changes their factorization; it does not define an otherwise undefined alignment.

A.2.3 DIRECTIONAL NON-IDENTIFIABILITY

Fix selected-client observations with $\mathcal{U} \neq \{0\}$, and suppose unselected clients have positive total weight and unrestricted updates. For any nonzero $v \in \mathcal{U}$, choose their weighted contribution so that Δ_{all} equals either $\bar{\Delta} + v$ or $\bar{\Delta} - v$. Both worlds have the same selected observations but errors $-v$ and v, and hence opposite oracle corrections. A deterministic selected-update-only rule returns the same C in both, while

$$\frac{1}{2}\big(\| - v + C\|^2 + \|v + C\|^2\big) = \|v\|^2 + \|C\|^2.$$

A nonzero C therefore worsens at least one compatible world. This is a worst-case boundary with unrestricted unseen updates, not an impossibility result for useful aggregation. It does not apply under full participation; structural assumptions or additional information can restrict the compatible worlds and remove the ambiguity.

A.3 LGR AND BOUNDED REALIZATION

A.3.1 RESIDUAL-RESPONSE REGRESSION

For motivation only, fix a common model, common differentiable loss, common step size ℓ_t, pure label shift, and finite expected gradients. With G_t containing the class-conditional expected gradients, one expected step gives

$$\Delta_{k,t} = -\ell_t G_t \pi_k, \qquad r_{k,t} = -\ell_t G_t H_C \widetilde{x}_{k,t}.$$

The second equality follows because prior differences sum to zero. Client-dependent FAC changes the class-conditional gradients; unequal local work and multi-step dynamics also invalidate this exact special case. The regression interpretation requires no claim that these effects vanish.

Suppress t. Let Z, R contain columns $\widetilde{x}_k, r_k$, and $W = \operatorname{diag}(w)$. Then $\Sigma_x = ZWZ^\top$, $\Sigma_{rx} = RWZ^\top$, and $Zw = Rw = 0$. The weighted ridge objective is

$$J(B) = \sum_{k \in \mathcal{S}} w_k \|r_k - B\widetilde{x}_k\|_2^2 + r_x \|B\|_F^2,$$

where $\|\cdot\|_F$ denotes the Frobenius norm. For $A = \Sigma_x + r_x I \succ 0$, completing the square gives

$$J(B) - J(\widehat{B}) = \operatorname{tr}\big((B - \widehat{B})A(B - \widehat{B})^\top\big), \qquad \widehat{B} = \Sigma_{rx} A^{-1}.$$

Thus Eq. (8) is the unique minimizer. Setting $d = -A^{-1}\bar{x}$, $\delta = WZ^\top d$, and $\alpha = w + \delta$ gives

$$\mathbf{1}^\top \delta = (Zw)^\top d = 0, \qquad C^0 = \Sigma_{rx} d = R\delta = \sum_k (\alpha_k - w_k)\Delta_k.$$

This proves the exact affine representation. Because the regularized matrix is positive definite, its symmetric pseudoinverse coincides with the ordinary inverse in exact arithmetic. Ridge ensures a unique fitted response, not recovery of unobserved responses. If $v \in \ker \Sigma_x$, then $Z^\top v = 0$ and $\widehat{B}v = 0$. In particular, $C^0 = 0$ when all selected priors coincide, or when $\bar{x} = 0$, including full participation with empirical federation weights.

A.3.2 RELATIONSHIP TO REGULARIZED LABEL MATCHING

The same unscaled coefficients uniquely solve

$$\min_{\mathbf{1}^\top a = 1} \left\| \sum_k a_k x_k \right\|_2^2 + r_x \sum_k \frac{a_k^2}{w_k}. \tag{21}$$

To verify this, write $a = w + \delta$, $\mathbf{1}^\top \delta = 0$. The objective becomes $\|\bar{x} + Z\delta\|^2 + r_x \delta^\top W^{-1}\delta + r_x$. For $\delta = WZ^\top d$ from above, $\bar{x} + Z\delta = -r_x d$, so its gradient is $2Z^\top(-r_x d) + 2r_x W^{-1}\delta = 0$. Positive ridge makes the objective strictly convex; this stationary, feasible point is the unique solution.

Because H_C is an isometry on zero-sum contrasts, the first term in Eq. (21) equals $\|\sum_k a_k \pi_k - \pi_g\|^2$. Let $N_{\mathcal{S}} = \sum_{k \in \mathcal{S}} N_k$ denote the total selected-client sample count. Since $w_k = N_k / N_{\mathcal{S}}$, we have $r_x / w_k = N_{\mathcal{S}} r_x / N_k$. At target π_g and regularization $\mu = N_{\mathcal{S}} r_x$, adding $a \geq 0$ therefore gives the corresponding regularized simplex label-matching objective. This relationship concerns the unscaled candidate. Radial clipping projects its parameter correction onto a ball; it does not solve label matching subject to a parameter-norm constraint.

A.3.3 PROJECTION AND DEVIATION BOUNDS

Proof of Proposition 2. Put $v = C^0$ and $b = \rho_s\|\bar{\Delta}\|$. If $\|v\| \le b$, $C = v$ attains zero projection distance. Otherwise any feasible C satisfies $\|v - C\| \ge \|v\| - \|C\| \ge \|v\| - b$, with equality at $bv/\|v\|$. Strict convexity of squared distance gives uniqueness, including $b = 0$. Radial scaling preserves $\mathcal{U}$. The triangle and reverse triangle inequalities give Eq. (16).

For $\bar{\Delta} \neq 0$, normalize its norm to one and write $C = a\widehat{\Delta} + c_\perp$, with $a^2 + \|c_\perp\|^2 \le \rho_s^2$. Since $\rho_s < 1$, $1 + a > 0$ and

$$\cos(\bar{\Delta}, \bar{\Delta} + C) = \frac{1 + a}{\sqrt{(1 + a)^2 + \|c_\perp\|^2}}.$$

The difference $(1 + a)^2 - (1 - \rho_s^2)((1 + a)^2 + \|c_\perp\|^2)$ is at least $(a + \rho_s^2)^2 \ge 0$. Taking the positive square root proves Eq. (17). The bound is attained in dimension at least two by $a = -\rho_s^2$ and $\|c_\perp\| = \rho_s\sqrt{1 - \rho_s^2}$. □

When $\bar{\Delta} = 0$, the projection gives zero correction, and the angle is undefined. Finally, the budget bounds η by $\rho_s\|\bar{\Delta}\|/\|C^*\|$ when $C^* \neq 0$, but supplies no information about alignment A.

B ADDITIONAL EXPERIMENTAL EVIDENCE

This appendix presents the settings needed to interpret the comparisons, complete predictive metrics, and detailed cross-backbone and local-epoch comparisons.

B.1 EXPERIMENTAL SETTINGS

Data and metrics. FAC/LGR priors use training counts only. Evaluation follows the balanced test sets and Last5 protocol in Section 5. Group recall averages per-class recall; Macro-F1 averages per-class F1 and cannot be reconstructed from group recall. Table scores are fractions. All three-seed results use seeds 42, 43, and 44.

Optimization and backbones. FedAvg, FedPALS, and FedPRS use SGD with initial learning rate 0.1, momentum 0.9, weight decay 5×10^{-4}, and no Nesterov acceleration. At zero-based round t, their learning rate is $0.05[1 + \cos(\pi t/200)]$, fixed within the round. A fresh optimizer is constructed for each local call. For local model training, CReFF on CIFAR-10-LT, FedLF, and FedYoYo use SGD with learning rate 0.1, zero momentum and weight decay, and no decay schedule. CReFF's server-side optimization settings are specified separately in Table 6. Changing local epochs does not change either schedule. Cross-backbone comparisons use a CIFAR-style ResNet18 with a 3×3 stride-one stem and no initial max-pooling, and WRN-28-2 with zero dropout. Methods share base initialization within a backbone–seed pair; method-specific modules have their own initialization.

Preprocessing and scope. The base augmentation is a random 32×32 crop with padding four and horizontal flipping; evaluation has no stochastic augmentation. The ResNet32 implementations of FedAvg, FedPALS, and FedPRS apply dataset-specific normalization: CIFAR-10 uses $(0.4914, 0.4822, 0.4465)/(0.2470, 0.2435, 0.2616)$, while CIFAR-100 uses $(0.5071, 0.4867, 0.4408)/(0.2675, 0.2565, 0.2761)$, with each pair written as mean/standard deviation. CReFF on CIFAR-10-LT, FedLF, and FedYoYo use $(0.4914, 0.4822, 0.4465)/(0.2023, 0.1994, 0.2010)$; the CIFAR-10 cross-backbone scaffold also uses this convention. FedYoYo uses its weak/strong views, including rotation, CIFAR AutoAugment, and Cutout. Thus shared partitions and budgets do not imply identical optimization or preprocessing. The comparisons use the settings specified here, which may differ from those in each method's original paper.

Table 6 gives method-specific settings. Parameter and buffer handling follows each method-specific implementation; FedPRS corrects trainable parameters only and leaves buffers to FedAvg. The fixed FAC/LGR parameters are used across datasets, backbones, and epoch budgets.

Table 6: Method-specific hyperparameters used in the comparisons.

Method	Settings
FedAvg (McMahan et al., 2017)	Raw cross-entropy; selected-client sample-count weighting.
CReFF (Shang et al., 2022)	100 synthetic features per class; 100 feature-matching steps per round; feature SGD learning rate 0.1. Server classifier retraining: 300 epochs per round, batch size 64, SGD learning rate 0.01. Both server optimizers use zero momentum and weight decay. Real classifier-gradient probes use up to 32 examples per observed class.
FedLF (Lu et al., 2025)	Adaptive logit-scaling floor 0.25; class-centred and feature-decorrelation losses each weighted by 0.01. Class-centre distance margin capped at 100; singleton-batch decorrelation loss set to zero.
FedYoYo (Yan et al., 2025)	DLA exponent $\tau = 1.5$; local-prior mixing coefficient $\gamma = 0.1$; distillation temperature $T = 1.5$ and weight 4. Linear distillation warm-up over 50 rounds. Federation effective-count smoothing uses old/new weights $0.9/0.1$. Weak/strong views as described in the protocol.
FedPALS (Listo Zec et al., 2026)	Raw cross-entropy; balanced target $1/C$. Training-only validation selects ESS fraction $0.10/0.75$ for CIFAR-10/100.
FedPRS	FAC pseudo-count ratio $\rho = 0.1$, equivalent to $\lambda = 1/11$; prior floor 10^{-12} followed by normalization; raw-logit inference. LGR ridge scale $r_{x,t} = \max\{\mathrm{tr}(\Sigma_{x,t})/(C-1), 10^{-12}\}$ and relative budget $\rho_s = 0.25$.

B.2 Complete CIFAR-LT Results

Tables 7 and 8 give the five metrics underlying the main comparison. Higher Macro-F1 does not imply higher recall in every class group. In particular, on CIFAR-100-LT, FedPRS trades head recall for medium/tail recall relative to FedAvg, and FedYoYo has higher tail recall at $\alpha = 0.1$. Changes in IF and α alter several properties of the data simultaneously and do not isolate support fragmentation causally.

CIFAR-10-LT trends. FedPRS has the highest mean accuracy, Macro-F1, and medium-class recall in all six CIFAR-10-LT settings. Its advantage is therefore not confined to the most imbalanced or most heterogeneous configuration. The class-group columns qualify this aggregate mean ordering. FedPALS generally favors head classes, while FedYoYo obtains the highest tail recall in five of the six settings; FedPRS leads tail recall only at $(\alpha, \mathrm{IF}) = (0.1, 10)$. At $\alpha = 0.1$, the tail leader changes without changing the Macro-F1 leader, showing that improved minority recall can accompany weaker performance elsewhere. At $\alpha = 0.5$, FedPRS remains second in tail recall and first in medium recall throughout. For example, at $(\alpha, \mathrm{IF}) = (0.1, 50)$, FedPRS has lower tail recall than FedYoYo (0.598 versus 0.721), but higher head/medium recall ($0.668/0.569$ versus $0.178/0.472$) and Macro-F1 (0.601 versus 0.396). The principal empirical pattern is thus broader class coverage rather than uniform dominance of either endpoint of the class-frequency spectrum.

CIFAR-100-LT trends. FedPRS again leads mean accuracy, Macro-F1, and medium recall in all six settings. At $\alpha = 0.1$, FedYoYo has the highest tail recall for every IF, whereas FedPRS is second; at $\alpha = 0.5$, FedPRS has the highest tail recall for every IF. Head recall is instead led by FedAvg or FedPALS. At IF$= 100$, increasing α from 0.1 to 0.5 raises FedPRS's tail recall from 0.205 to 0.225, while FedYoYo's decreases from 0.225 to 0.199; FedPRS's medium recall also increases from 0.316 to 0.352. This reversal across α indicates that the observed class-group allocation depends on client heterogeneity and cannot be summarized by an invariant head–tail shift. The smaller Macro-F1 margins at IF$= 10$ also show that FedPRS's relative advantage is not a monotone function of global imbalance. Because IF and α change multiple aspects of the sampled training problem, these comparisons establish empirical coverage across settings rather than a causal support-fragmentation law.

Table 7: Complete CIFAR-10-LT Last5 results for all compared methods.

α	IF	Method	Accuracy	Macro-F1	Head	Medium	Tail
0.1	10	FedAvg	0.469 ± 0.054	0.415 ± 0.062	0.670 ± 0.067	0.429 ± 0.067	0.320 ± 0.144
		CReFF	0.650 ± 0.022	0.643 ± 0.022	0.685 ± 0.031	0.575 ± 0.040	0.716 ± 0.074
		FedLF	0.600 ± 0.044	0.586 ± 0.055	0.637 ± 0.023	0.574 ± 0.066	0.599 ± 0.156
		FedYoYo	0.649 ± 0.026	0.633 ± 0.034	0.584 ± 0.065	0.633 ± 0.046	0.737 ± 0.084
		FedPALS	0.669 ± 0.041	0.654 ± 0.054	$\mathbf{0.819 \pm 0.041}$	0.604 ± 0.120	0.605 ± 0.080
		FedPRS	$\mathbf{0.718 \pm 0.042}$	$\mathbf{0.714 \pm 0.042}$	0.749 ± 0.054	$\mathbf{0.673 \pm 0.064}$	$\mathbf{0.747 \pm 0.072}$
	50	FedAvg	0.249 ± 0.116	0.145 ± 0.116	0.527 ± 0.163	0.225 ± 0.167	0.001 ± 0.003
		CReFF	0.493 ± 0.042	0.471 ± 0.049	0.518 ± 0.037	0.436 ± 0.093	0.545 ± 0.022
		FedLF	0.426 ± 0.044	0.393 ± 0.048	0.460 ± 0.060	0.434 ± 0.103	0.383 ± 0.018
		FedYoYo	0.458 ± 0.045	0.396 ± 0.068	0.178 ± 0.080	0.472 ± 0.079	$\mathbf{0.721 \pm 0.056}$
		FedPALS	0.428 ± 0.034	0.368 ± 0.021	$\mathbf{0.668 \pm 0.088}$	0.429 ± 0.060	0.185 ± 0.052
		FedPRS	$\mathbf{0.607 \pm 0.073}$	$\mathbf{0.601 \pm 0.075}$	0.668 ± 0.097	$\mathbf{0.569 \pm 0.060}$	0.598 ± 0.069
	100	FedAvg	0.229 ± 0.045	0.129 ± 0.033	0.653 ± 0.073	0.084 ± 0.059	0.000 ± 0.000
		CReFF	0.462 ± 0.053	0.437 ± 0.070	0.590 ± 0.054	0.382 ± 0.056	0.440 ± 0.114
		FedLF	0.373 ± 0.052	0.317 ± 0.040	0.539 ± 0.176	0.295 ± 0.069	0.312 ± 0.109
		FedYoYo	0.395 ± 0.055	0.339 ± 0.074	0.256 ± 0.134	0.402 ± 0.066	$\mathbf{0.523 \pm 0.069}$
		FedPALS	0.401 ± 0.033	0.327 ± 0.024	$\mathbf{0.730 \pm 0.069}$	0.388 ± 0.107	0.089 ± 0.061
		FedPRS	$\mathbf{0.533 \pm 0.064}$	$\mathbf{0.521 \pm 0.069}$	0.679 ± 0.072	$\mathbf{0.487 \pm 0.075}$	0.447 ± 0.150
0.5	10	FedAvg	0.798 ± 0.009	0.799 ± 0.008	0.903 ± 0.012	0.752 ± 0.025	0.756 ± 0.011
		CReFF	0.776 ± 0.005	0.772 ± 0.006	0.817 ± 0.013	0.704 ± 0.017	0.830 ± 0.020
		FedLF	0.773 ± 0.010	0.771 ± 0.009	0.779 ± 0.014	0.720 ± 0.012	0.837 ± 0.017
		FedYoYo	0.784 ± 0.005	0.781 ± 0.005	0.768 ± 0.005	0.725 ± 0.021	$\mathbf{0.880 \pm 0.009}$
		FedPALS	0.816 ± 0.007	0.816 ± 0.007	$\mathbf{0.907 \pm 0.005}$	0.766 ± 0.011	0.791 ± 0.016
		FedPRS	$\mathbf{0.838 \pm 0.001}$	$\mathbf{0.838 \pm 0.002}$	0.890 ± 0.009	$\mathbf{0.784 \pm 0.004}$	0.857 ± 0.010
	50	FedAvg	0.652 ± 0.013	0.646 ± 0.012	0.899 ± 0.005	0.606 ± 0.015	0.464 ± 0.022
		CReFF	0.673 ± 0.019	0.666 ± 0.020	0.768 ± 0.041	0.579 ± 0.048	0.704 ± 0.030
		FedLF	0.629 ± 0.080	0.629 ± 0.080	0.703 ± 0.083	0.589 ± 0.075	0.608 ± 0.085
		FedYoYo	0.633 ± 0.018	0.616 ± 0.028	0.474 ± 0.083	0.619 ± 0.023	$\mathbf{0.811 \pm 0.022}$
		FedPALS	0.671 ± 0.015	0.667 ± 0.015	$\mathbf{0.904 \pm 0.009}$	0.631 ± 0.008	0.490 ± 0.032
		FedPRS	$\mathbf{0.771 \pm 0.007}$	$\mathbf{0.771 \pm 0.007}$	0.873 ± 0.017	$\mathbf{0.724 \pm 0.004}$	0.732 ± 0.014
	100	FedAvg	0.570 ± 0.018	0.544 ± 0.022	0.882 ± 0.015	0.564 ± 0.036	0.266 ± 0.025
		CReFF	0.629 ± 0.004	0.620 ± 0.005	0.722 ± 0.008	0.561 ± 0.016	0.626 ± 0.002
		FedLF	0.603 ± 0.030	0.594 ± 0.035	0.694 ± 0.028	0.566 ± 0.064	0.563 ± 0.022
		FedYoYo	0.541 ± 0.017	0.511 ± 0.021	0.353 ± 0.015	0.508 ± 0.048	$\mathbf{0.774 \pm 0.025}$
		FedPALS	0.606 ± 0.023	0.589 ± 0.025	$\mathbf{0.888 \pm 0.012}$	0.596 ± 0.034	0.337 ± 0.034
		FedPRS	$\mathbf{0.734 \pm 0.005}$	$\mathbf{0.734 \pm 0.005}$	0.877 ± 0.010	$\mathbf{0.688 \pm 0.015}$	0.652 ± 0.026

B.2.1 Component variability

Table 3 shows that FAC raises medium/tail recall from $0.084/0.000$ to $0.444/0.414$, while head recall changes from 0.653 to 0.645. LGR alone reaches only 0.017 tail recall, despite the highest head recall (0.699), suggesting that server correction alone does not reproduce FAC's minority-class gains in this setting. Adding LGR to FAC raises mean Macro-F1 by 0.045 and tail recall by 0.033, but increases the tail-recall standard deviation from 0.075 to 0.150. The additional benefit of LGR is not uniform across seeds: accuracy, Macro-F1, and tail recall improve for seeds 42 and 43 but decrease by $0.0050/0.0049/0.0529$ for seed 44. This variability limits the predictive conclusion; it does not contradict the alignment–scale identity, which concerns parameter-space reference error.

B.3 Client-Participation Sensitivity

Table 9 gives the complete three-seed Last5 results for $M \in \{2, 4, 8, 12, 16, 20\}$. The $M = 4$ setting corresponds to the main experimental configuration; at every participation rate, all four methods share a fixed participation schedule for each seed and M. FedPRS exceeds FAC in mean Macro-F1 through $M = 16$, with margins $0.080/0.045/0.039/0.019/0.037$ for $M = 2/4/8/12/16$. It also exceeds FedPALS at every evaluated rate. These differences are descriptive sensitivity measurements under a fixed communication budget: larger M processes more clients and data per round.

Table 8: Complete CIFAR-100-LT Last5 results for all compared methods.

α	IF	Method	Accuracy	Macro-F1	Head	Medium	Tail
0.1	10	FedAvg	0.450 ± 0.019	0.437 ± 0.019	0.585 ± 0.026	0.444 ± 0.024	0.320 ± 0.011
		FedLF	0.324 ± 0.016	0.308 ± 0.017	0.335 ± 0.047	0.322 ± 0.012	0.315 ± 0.023
		FedYoYo	0.393 ± 0.017	0.375 ± 0.019	0.349 ± 0.038	0.394 ± 0.022	$\mathbf{0.434 \pm 0.018}$
		FedPALS	0.462 ± 0.004	0.449 ± 0.005	$\mathbf{0.588 \pm 0.007}$	0.467 ± 0.010	0.332 ± 0.017
		FedPRS	$\mathbf{0.479 \pm 0.010}$	$\mathbf{0.475 \pm 0.010}$	0.531 ± 0.033	$\mathbf{0.483 \pm 0.016}$	0.424 ± 0.019
	50	FedAvg	0.337 ± 0.014	0.298 ± 0.012	$\mathbf{0.570 \pm 0.003}$	0.338 ± 0.039	0.103 ± 0.004
		FedLF	0.239 ± 0.004	0.206 ± 0.008	0.331 ± 0.011	0.268 ± 0.025	0.116 ± 0.010
		FedYoYo	0.279 ± 0.006	0.259 ± 0.006	0.258 ± 0.020	0.307 ± 0.009	$\mathbf{0.272 \pm 0.008}$
		FedPALS	0.345 ± 0.010	0.307 ± 0.009	0.566 ± 0.015	0.353 ± 0.012	0.115 ± 0.006
		FedPRS	$\mathbf{0.376 \pm 0.007}$	$\mathbf{0.368 \pm 0.007}$	0.487 ± 0.015	$\mathbf{0.375 \pm 0.015}$	0.265 ± 0.007
	100	FedAvg	0.288 ± 0.017	0.239 ± 0.016	0.553 ± 0.015	0.260 ± 0.032	0.052 ± 0.004
		FedLF	0.207 ± 0.010	0.166 ± 0.009	0.326 ± 0.030	0.237 ± 0.010	0.058 ± 0.008
		FedYoYo	0.245 ± 0.007	0.224 ± 0.009	0.247 ± 0.006	0.263 ± 0.037	$\mathbf{0.225 \pm 0.022}$
		FedPALS	0.294 ± 0.015	0.247 ± 0.015	$\mathbf{0.553 \pm 0.021}$	0.272 ± 0.022	0.057 ± 0.004
		FedPRS	$\mathbf{0.331 \pm 0.010}$	$\mathbf{0.318 \pm 0.011}$	0.472 ± 0.012	$\mathbf{0.316 \pm 0.010}$	0.205 ± 0.009
0.5	10	FedAvg	0.509 ± 0.011	0.500 ± 0.012	$\mathbf{0.644 \pm 0.004}$	0.512 ± 0.014	0.370 ± 0.017
		FedLF	0.412 ± 0.003	0.405 ± 0.003	0.411 ± 0.007	0.435 ± 0.008	0.388 ± 0.007
		FedYoYo	0.441 ± 0.024	0.431 ± 0.025	0.450 ± 0.027	0.438 ± 0.028	0.435 ± 0.017
		FedPALS	0.505 ± 0.010	0.496 ± 0.009	0.640 ± 0.015	0.512 ± 0.006	0.363 ± 0.009
		FedPRS	$\mathbf{0.529 \pm 0.003}$	$\mathbf{0.525 \pm 0.004}$	0.591 ± 0.005	$\mathbf{0.534 \pm 0.008}$	$\mathbf{0.462 \pm 0.006}$
	50	FedAvg	0.375 ± 0.021	0.340 ± 0.021	$\mathbf{0.615 \pm 0.020}$	0.373 ± 0.031	0.137 ± 0.013
		FedLF	0.306 ± 0.007	0.279 ± 0.011	0.411 ± 0.009	0.363 ± 0.015	0.142 ± 0.016
		FedYoYo	0.335 ± 0.005	0.321 ± 0.003	0.367 ± 0.024	0.353 ± 0.010	0.284 ± 0.015
		FedPALS	0.374 ± 0.003	0.341 ± 0.003	0.604 ± 0.002	0.377 ± 0.011	0.142 ± 0.004
		FedPRS	$\mathbf{0.420 \pm 0.011}$	$\mathbf{0.413 \pm 0.013}$	0.550 ± 0.011	$\mathbf{0.412 \pm 0.010}$	$\mathbf{0.298 \pm 0.013}$
	100	FedAvg	0.327 ± 0.009	0.280 ± 0.010	0.598 ± 0.015	0.314 ± 0.006	0.069 ± 0.006
		FedLF	0.269 ± 0.009	0.226 ± 0.006	0.422 ± 0.014	0.334 ± 0.018	0.048 ± 0.006
		FedYoYo	0.302 ± 0.012	0.282 ± 0.014	0.370 ± 0.013	0.336 ± 0.025	0.199 ± 0.015
		FedPALS	0.336 ± 0.008	0.291 ± 0.009	$\mathbf{0.608 \pm 0.012}$	0.325 ± 0.005	0.077 ± 0.011
		FedPRS	$\mathbf{0.368 \pm 0.020}$	$\mathbf{0.358 \pm 0.019}$	0.526 ± 0.023	$\mathbf{0.352 \pm 0.021}$	$\mathbf{0.225 \pm 0.016}$

Table 9: Complete Last5 participation-rate results on CIFAR-10-LT (IF=100, $\alpha = 0.1$).

M	Rate	Method	Accuracy	Macro-F1	Head	Medium	Tail
2	10%	FedAvg	0.159 ± 0.024	0.063 ± 0.019	0.469 ± 0.139	0.046 ± 0.044	0.000 ± 0.000
		FAC	0.374 ± 0.064	0.314 ± 0.090	0.596 ± 0.106	0.315 ± 0.044	0.229 ± 0.152
		FedPALS	0.245 ± 0.027	0.133 ± 0.021	$\mathbf{0.633 \pm 0.065}$	0.138 ± 0.069	0.000 ± 0.000
		FedPRS	$\mathbf{0.439 \pm 0.084}$	$\mathbf{0.395 \pm 0.105}$	0.597 ± 0.118	$\mathbf{0.387 \pm 0.098}$	$\mathbf{0.350 \pm 0.134}$
4	20%	FedAvg	0.229 ± 0.045	0.129 ± 0.033	0.653 ± 0.073	0.084 ± 0.059	0.000 ± 0.000
		FAC	0.495 ± 0.064	0.476 ± 0.074	0.645 ± 0.080	0.444 ± 0.094	0.414 ± 0.075
		FedPALS	0.401 ± 0.033	0.327 ± 0.024	$\mathbf{0.730 \pm 0.069}$	0.388 ± 0.107	0.089 ± 0.061
		FedPRS	$\mathbf{0.533 \pm 0.064}$	$\mathbf{0.521 \pm 0.069}$	0.679 ± 0.072	$\mathbf{0.487 \pm 0.075}$	$\mathbf{0.447 \pm 0.150}$
8	40%	FedAvg	0.357 ± 0.094	0.266 ± 0.110	$\mathbf{0.761 \pm 0.103}$	0.293 ± 0.146	0.038 ± 0.034
		FAC	0.523 ± 0.086	0.499 ± 0.099	0.734 ± 0.086	0.482 ± 0.116	0.368 ± 0.101
		FedPALS	0.457 ± 0.033	0.410 ± 0.040	0.742 ± 0.021	0.464 ± 0.057	0.162 ± 0.017
		FedPRS	$\mathbf{0.550 \pm 0.066}$	$\mathbf{0.538 \pm 0.070}$	0.741 ± 0.049	$\mathbf{0.494 \pm 0.102}$	$\mathbf{0.435 \pm 0.071}$
12	60%	FedAvg	0.369 ± 0.055	0.293 ± 0.062	$\mathbf{0.772 \pm 0.090}$	0.262 ± 0.136	0.109 ± 0.014
		FAC	0.554 ± 0.066	0.538 ± 0.076	0.756 ± 0.079	0.478 ± 0.084	0.455 ± 0.138
		FedPALS	0.480 ± 0.015	0.447 ± 0.014	0.764 ± 0.083	0.459 ± 0.091	0.224 ± 0.063
		FedPRS	$\mathbf{0.568 \pm 0.049}$	$\mathbf{0.557 \pm 0.056}$	0.745 ± 0.073	$\mathbf{0.504 \pm 0.080}$	$\mathbf{0.477 \pm 0.068}$
16	80%	FedAvg	0.365 ± 0.108	0.298 ± 0.126	$\mathbf{0.760 \pm 0.125}$	0.258 ± 0.182	0.112 ± 0.074
		FAC	0.504 ± 0.105	0.478 ± 0.120	0.756 ± 0.117	0.432 ± 0.140	0.348 ± 0.136
		FedPALS	0.460 ± 0.025	0.423 ± 0.040	0.728 ± 0.093	0.458 ± 0.053	0.193 ± 0.099
		FedPRS	$\mathbf{0.533 \pm 0.095}$	$\mathbf{0.515 \pm 0.106}$	0.746 ± 0.092	$\mathbf{0.477 \pm 0.123}$	$\mathbf{0.395 \pm 0.102}$
20	100%	FedAvg	0.372 ± 0.105	0.304 ± 0.126	$\mathbf{0.769 \pm 0.135}$	0.260 ± 0.169	0.123 ± 0.090
		FAC	$\mathbf{0.525 \pm 0.122}$	$\mathbf{0.504 \pm 0.140}$	0.755 ± 0.104	$\mathbf{0.449 \pm 0.169}$	$\mathbf{0.398 \pm 0.121}$
		FedPALS	0.436 ± 0.060	0.400 ± 0.053	0.715 ± 0.106	0.412 ± 0.101	0.189 ± 0.075
		FedPRS	0.516 ± 0.126	0.494 ± 0.149	0.763 ± 0.099	0.425 ± 0.169	0.392 ± 0.129

The mechanism statistics in Table 10 separate predictive behavior from the server correction. The mean sampled label-center norm decreases from 0.551 at $M = 2$ to 0.097 at $M = 16$. The correction budget is active in $60.8/64.5/56.3/37.8/17.2/0.0\%$ of rounds for $M = 2/4/8/12/16/20$, respectively. At $M = 20$, federation data weights give a maximum label-center residual of 6.51×10^{-17}, a maximum correction ratio of zero, and maximum $|\alpha_k - w_k|$ of 2.78×10^{-17}. Thus LGR numerically degenerates to FedAvg aggregation under full participation. At $M = 20$, the LGR

correction is zero in every round, so the observed -0.010 mean Macro-F1 difference between FAC and FedPRS is not attributable to residual correction and should not be interpreted as a method effect. The corresponding mean label-center norm for $M = 4$ is 0.404, computed from the fixed client partitions, federation prior, and participation schedules.

Table 10: FedPRS mechanism statistics across participation rates.

M	Rate	Mean $\|\bar{x}_t\|_2$	Budget-active rounds
2	10%	0.551	60.8%
4	20%	0.404	64.5%
8	40%	0.243	56.3%
12	60%	0.161	37.8%
16	80%	0.097	17.2%
20	100%	5.58×10^{-17}	0.0%

B.4 Complete Cross-Backbone Results

Table 11 expands the main comparison. FedPRS has the highest mean Macro-F1 and medium recall on all three backbones, while FedPALS leads head recall and FedYoYo leads tail recall. The tested set supports transfer among these convolutional architectures, not model-agnostic superiority.

Table 11: Complete cross-backbone CIFAR-10-LT Last5 results (IF=100, $\alpha = 0.1$).

Backbone	Method	Accuracy	Macro-F1	Head	Medium	Tail
ResNet32	FedAvg	0.229 ± 0.045	0.129 ± 0.033	0.653 ± 0.073	0.084 ± 0.059	0.000 ± 0.000
	CReFF	0.462 ± 0.053	0.437 ± 0.070	0.590 ± 0.054	0.382 ± 0.056	0.440 ± 0.114
	FedLF	0.373 ± 0.052	0.317 ± 0.040	0.539 ± 0.176	0.295 ± 0.069	0.312 ± 0.109
	FedYoYo	0.395 ± 0.055	0.339 ± 0.074	0.256 ± 0.134	0.402 ± 0.066	$\mathbf{0.523 \pm 0.069}$
	FedPALS	0.401 ± 0.033	0.327 ± 0.024	$\mathbf{0.730 \pm 0.069}$	0.388 ± 0.107	0.089 ± 0.061
	FedPRS	$\mathbf{0.533 \pm 0.064}$	$\mathbf{0.521 \pm 0.069}$	0.679 ± 0.072	$\mathbf{0.487 \pm 0.075}$	0.447 ± 0.150
ResNet18	FedAvg	0.318 ± 0.087	0.220 ± 0.090	0.709 ± 0.111	0.258 ± 0.139	0.007 ± 0.006
	CReFF	0.523 ± 0.039	0.504 ± 0.048	0.653 ± 0.039	0.440 ± 0.049	0.505 ± 0.154
	FedLF	0.449 ± 0.033	0.410 ± 0.049	0.636 ± 0.189	0.399 ± 0.063	0.330 ± 0.232
	FedYoYo	0.476 ± 0.048	0.428 ± 0.051	0.388 ± 0.042	0.436 ± 0.073	$\mathbf{0.619 \pm 0.132}$
	FedPALS	0.475 ± 0.044	0.434 ± 0.056	$\mathbf{0.739 \pm 0.067}$	0.474 ± 0.086	0.212 ± 0.111
	FedPRS	$\mathbf{0.594 \pm 0.024}$	$\mathbf{0.586 \pm 0.022}$	0.690 ± 0.072	$\mathbf{0.557 \pm 0.016}$	0.546 ± 0.116
WRN-28-2	FedAvg	0.270 ± 0.038	0.169 ± 0.036	0.677 ± 0.076	0.166 ± 0.053	0.000 ± 0.000
	CReFF	0.497 ± 0.036	0.468 ± 0.049	0.575 ± 0.060	0.416 ± 0.045	0.527 ± 0.120
	FedLF	0.442 ± 0.028	0.398 ± 0.049	0.691 ± 0.101	0.374 ± 0.017	0.284 ± 0.124
	FedYoYo	0.523 ± 0.032	0.482 ± 0.030	0.394 ± 0.051	0.504 ± 0.030	$\mathbf{0.676 \pm 0.101}$
	FedPALS	0.461 ± 0.029	0.414 ± 0.026	$\mathbf{0.748 \pm 0.070}$	0.458 ± 0.090	0.177 ± 0.098
	FedPRS	$\mathbf{0.555 \pm 0.055}$	$\mathbf{0.545 \pm 0.058}$	0.679 ± 0.102	$\mathbf{0.509 \pm 0.067}$	0.494 ± 0.132

The Macro-F1 margin of FedPRS over the strongest external method is 0.084 on ResNet32, 0.082 on ResNet18, and 0.063 on WRN-28-2. Its corresponding accuracy margin decreases from 0.071 on ResNet32 and ResNet18 to 0.032 on WRN-28-2. The backbone change therefore preserves the aggregate mean ordering while narrowing it on WRN-28-2. The class-group mean ordering is more stable across architectures: FedPALS emphasizes head recall, FedYoYo emphasizes tail recall, and FedPRS has the highest medium recall. FedPRS's medium-recall margin over the next-best method is $0.085/0.083/0.005$ on ResNet32/ResNet18/WRN-28-2, respectively, so the same mean ordering can correspond to substantially different margins. Since each method uses its specified optimization and preprocessing choices, this experiment tests whether the observed mean ordering transfers within the common data and communication scaffold; it is not a controlled decomposition of architectural effects.

B.5 Local-Optimization-Budget Sensitivity

The sweeps use IF$= 100$, $\alpha = 0.1$, ResNet32, and 200 rounds with $E \in \{1, 3, 5, 7, 10\}$. They vary computation at fixed communication budget and do not retune FedPRS.

B.5.1 CIFAR-10-LT

Table 12 reports all metrics for all compared methods. FedPRS leads mean accuracy and Macro-F1 at every budget, but FedYoYo leads tail recall except at $E = 3$. At $E = 10$, FedYoYo's tail recall is 0.687 versus FedPRS's 0.497, despite lower Macro-F1 (0.494 versus 0.560). Thus aggregate and minority-class mean orderings need not agree. At $E = 5$, the sensitivity sweep gives a Macro-F1 of 0.435 for CReFF, closely matching the 0.437 reported in the main comparison.

Table 12: CIFAR-10-LT local-epoch sensitivity (Last5 mean±sample standard deviation).

E	Method	Accuracy	Macro-F1	Head	Medium	Tail
1	FedAvg	0.152 ± 0.027	0.058 ± 0.020	0.498 ± 0.105	0.008 ± 0.011	0.000 ± 0.000
	CReFF	0.299 ± 0.004	0.242 ± 0.012	0.507 ± 0.036	0.237 ± 0.072	0.172 ± 0.119
	FedLF	0.263 ± 0.010	0.171 ± 0.019	0.629 ± 0.140	0.115 ± 0.101	0.096 ± 0.135
	FedYoYo	0.234 ± 0.048	0.146 ± 0.041	0.127 ± 0.066	0.144 ± 0.150	$\mathbf{0.461 \pm 0.195}$
	FedPALS	0.239 ± 0.028	0.136 ± 0.023	$\mathbf{0.630 \pm 0.064}$	0.126 ± 0.100	0.000 ± 0.000
	FedPRS	$\mathbf{0.338 \pm 0.039}$	$\mathbf{0.261 \pm 0.052}$	0.618 ± 0.086	$\mathbf{0.272 \pm 0.073}$	0.145 ± 0.028
3	FedAvg	0.224 ± 0.026	0.121 ± 0.016	0.642 ± 0.087	0.078 ± 0.016	0.000 ± 0.000
	CReFF	0.425 ± 0.065	0.402 ± 0.073	0.537 ± 0.062	0.372 ± 0.129	0.384 ± 0.054
	FedLF	0.360 ± 0.039	0.295 ± 0.031	0.543 ± 0.140	0.281 ± 0.091	0.282 ± 0.124
	FedYoYo	0.311 ± 0.024	0.231 ± 0.022	0.207 ± 0.035	0.316 ± 0.092	0.408 ± 0.062
	FedPALS	0.355 ± 0.034	0.270 ± 0.034	$\mathbf{0.686 \pm 0.045}$	0.332 ± 0.085	0.055 ± 0.047
	FedPRS	$\mathbf{0.487 \pm 0.051}$	$\mathbf{0.464 \pm 0.060}$	0.632 ± 0.088	$\mathbf{0.420 \pm 0.092}$	$\mathbf{0.432 \pm 0.082}$
5	FedAvg	0.229 ± 0.045	0.129 ± 0.033	0.653 ± 0.073	0.084 ± 0.059	0.000 ± 0.000
	CReFF	0.461 ± 0.052	0.435 ± 0.066	0.600 ± 0.072	0.380 ± 0.056	0.431 ± 0.098
	FedLF	0.373 ± 0.052	0.317 ± 0.040	0.539 ± 0.176	0.295 ± 0.069	0.312 ± 0.109
	FedYoYo	0.395 ± 0.055	0.339 ± 0.074	0.256 ± 0.134	0.402 ± 0.066	$\mathbf{0.523 \pm 0.069}$
	FedPALS	0.401 ± 0.033	0.327 ± 0.024	$\mathbf{0.730 \pm 0.069}$	0.388 ± 0.107	0.089 ± 0.061
	FedPRS	$\mathbf{0.533 \pm 0.064}$	$\mathbf{0.521 \pm 0.069}$	0.679 ± 0.072	$\mathbf{0.487 \pm 0.075}$	0.447 ± 0.150
7	FedAvg	0.277 ± 0.040	0.170 ± 0.039	0.670 ± 0.024	0.189 ± 0.082	0.000 ± 0.000
	CReFF	0.483 ± 0.046	0.462 ± 0.057	0.589 ± 0.056	0.392 ± 0.066	0.499 ± 0.067
	FedLF	0.395 ± 0.033	0.346 ± 0.038	0.584 ± 0.153	0.313 ± 0.095	0.318 ± 0.165
	FedYoYo	0.448 ± 0.064	0.395 ± 0.072	0.280 ± 0.105	0.468 ± 0.064	$\mathbf{0.590 \pm 0.120}$
	FedPALS	0.429 ± 0.031	0.367 ± 0.036	$\mathbf{0.741 \pm 0.073}$	0.419 ± 0.103	0.132 ± 0.091
	FedPRS	$\mathbf{0.563 \pm 0.034}$	$\mathbf{0.554 \pm 0.040}$	0.677 ± 0.038	$\mathbf{0.538 \pm 0.051}$	0.482 ± 0.141
10	FedAvg	0.258 ± 0.041	0.150 ± 0.030	0.628 ± 0.111	0.173 ± 0.021	0.000 ± 0.000
	CReFF	0.481 ± 0.035	0.462 ± 0.049	0.615 ± 0.063	0.375 ± 0.072	0.488 ± 0.103
	FedLF	0.414 ± 0.011	0.371 ± 0.007	0.612 ± 0.120	0.356 ± 0.061	0.293 ± 0.164
	FedYoYo	0.532 ± 0.021	0.494 ± 0.032	0.433 ± 0.021	0.489 ± 0.091	$\mathbf{0.687 \pm 0.092}$
	FedPALS	0.441 ± 0.022	0.382 ± 0.019	$\mathbf{0.734 \pm 0.067}$	0.443 ± 0.092	0.146 ± 0.076
	FedPRS	$\mathbf{0.569 \pm 0.027}$	$\mathbf{0.560 \pm 0.031}$	0.706 ± 0.024	$\mathbf{0.521 \pm 0.040}$	0.497 ± 0.067

Increasing E improves the mean Macro-F1 of every specialized method between the two endpoints, but the trajectories need not be monotone at every intermediate budget or metric. FedPRS rises by 0.299 from $E = 1$ to $E = 10$, while FedYoYo rises by 0.348 and becomes the closest comparator at $E = 10$. CReFF changes only slightly from $E = 7$ to $E = 10$ (0.462 at both displayed precision). FedPRS's Macro-F1 gains also diminish: the increase from $E = 1$ to $E = 3$ is 0.203, compared with 0.006 from $E = 7$ to $E = 10$. These are comparisons at fixed communication rounds, with more local work at larger E, and do not establish an advantage at equal total computation.

The size of the advantage varies with the local budget. Using the displayed means in Table 12, FedPRS's Macro-F1 margins over the strongest comparator are $0.019/0.062/0.086/0.092/0.066$ for $E = 1/3/5/7/10$, respectively. CReFF is the closest comparator through $E = 7$, after which FedYoYo narrows the gap. FedPRS also has the highest medium recall at every budget. Its substantial early improvement from $E = 1$ to $E = 3$ accompanies increases of 0.148 in medium recall and 0.287 in tail recall, compared with only 0.014 in head recall. By contrast, FedAvg's displayed tail recall remains 0.000 throughout the sweep, indicating that additional local training alone does not resolve its tail-class weakness under this configuration. At $E = 10$, FedPRS's higher head/medium recall than FedYoYo ($0.706/0.521$ versus $0.433/0.489$) helps contextualize its aggregate lead despite lower tail recall. These group averages describe the allocation of recall; they do not decompose Macro-F1, which also depends on per-class precision.

B.5.2 CIFAR-100-LT

Table 13 completes the corresponding CIFAR-100-LT comparison among the reported methods. FedLF and FedYoYo use the optimization settings specified in Appendix B.1, while FedAvg, FedPALS, and FedPRS use their momentum-SGD settings; changing E does not otherwise alter a method's configuration. FedPRS has the highest Macro-F1 at every budget, increasing from 0.145 at $E = 1$ to 0.331 at $E = 10$. Class-group mean orderings differ: FedYoYo has the highest tail recall at $E = 1, 5, 7, 10$, while FedPRS leads at $E = 3$. FedPALS has the highest head recall at $E = 3, 5, 10$, while FedAvg leads at $E = 7$ (0.573 versus 0.569 for FedPALS).

Table 13: CIFAR-100-LT local-epoch sensitivity (three-seed Last5 mean±sample standard deviation).

E	Method	Accuracy	Macro-F1	Head	Medium	Tail
1	FedAvg	0.089 ± 0.019	0.043 ± 0.016	0.241 ± 0.039	0.027 ± 0.020	0.000 ± 0.000
	FedLF	0.089 ± 0.008	0.043 ± 0.006	0.205 ± 0.016	0.062 ± 0.018	0.000 ± 0.000
	FedYoYo	0.067 ± 0.005	0.035 ± 0.004	0.029 ± 0.002	0.080 ± 0.006	$\mathbf{0.092 \pm 0.018}$
	FedPALS	0.116 ± 0.009	0.062 ± 0.008	0.301 ± 0.019	0.050 ± 0.017	0.000 ± 0.000
	FedPRS	$\mathbf{0.190 \pm 0.013}$	$\mathbf{0.145 \pm 0.013}$	$\mathbf{0.332 \pm 0.024}$	$\mathbf{0.183 \pm 0.020}$	0.056 ± 0.008
3	FedAvg	0.219 ± 0.020	0.163 ± 0.018	0.467 ± 0.034	0.174 ± 0.028	0.016 ± 0.003
	FedLF	0.179 ± 0.003	0.130 ± 0.001	0.310 ± 0.017	0.198 ± 0.007	0.030 ± 0.005
	FedYoYo	0.172 ± 0.005	0.145 ± 0.004	0.178 ± 0.013	0.181 ± 0.007	0.157 ± 0.002
	FedPALS	0.246 ± 0.011	0.188 ± 0.010	$\mathbf{0.507 \pm 0.016}$	0.212 ± 0.015	0.021 ± 0.004
	FedPRS	$\mathbf{0.310 \pm 0.019}$	$\mathbf{0.293 \pm 0.020}$	0.442 ± 0.020	$\mathbf{0.309 \pm 0.022}$	$\mathbf{0.178 \pm 0.016}$
5	FedAvg	0.288 ± 0.017	0.239 ± 0.016	0.553 ± 0.015	0.260 ± 0.032	0.052 ± 0.004
	FedLF	0.207 ± 0.010	0.166 ± 0.009	0.326 ± 0.030	0.237 ± 0.010	0.058 ± 0.008
	FedYoYo	0.245 ± 0.007	0.224 ± 0.009	0.247 ± 0.006	0.263 ± 0.037	$\mathbf{0.225 \pm 0.022}$
	FedPALS	0.294 ± 0.015	0.247 ± 0.015	$\mathbf{0.553 \pm 0.021}$	0.272 ± 0.022	0.057 ± 0.004
	FedPRS	$\mathbf{0.331 \pm 0.010}$	$\mathbf{0.318 \pm 0.011}$	0.472 ± 0.012	$\mathbf{0.316 \pm 0.010}$	0.205 ± 0.009
7	FedAvg	0.309 ± 0.014	0.265 ± 0.014	$\mathbf{0.573 \pm 0.025}$	0.282 ± 0.020	0.073 ± 0.004
	FedLF	0.228 ± 0.017	0.196 ± 0.018	0.349 ± 0.021	0.252 ± 0.021	0.084 ± 0.010
	FedYoYo	0.290 ± 0.003	0.272 ± 0.009	0.302 ± 0.006	0.312 ± 0.022	$\mathbf{0.256 \pm 0.015}$
	FedPALS	0.310 ± 0.013	0.267 ± 0.015	0.569 ± 0.017	0.289 ± 0.015	0.074 ± 0.009
	FedPRS	$\mathbf{0.340 \pm 0.004}$	$\mathbf{0.326 \pm 0.005}$	0.493 ± 0.008	$\mathbf{0.324 \pm 0.003}$	0.203 ± 0.007
10	FedAvg	0.310 ± 0.012	0.268 ± 0.011	0.570 ± 0.021	0.288 ± 0.015	0.073 ± 0.003
	FedLF	0.242 ± 0.020	0.219 ± 0.023	0.362 ± 0.019	0.255 ± 0.028	0.109 ± 0.013
	FedYoYo	0.330 ± 0.010	0.320 ± 0.012	0.363 ± 0.012	$\mathbf{0.340 \pm 0.023}$	$\mathbf{0.289 \pm 0.024}$
	FedPALS	0.322 ± 0.017	0.280 ± 0.017	$\mathbf{0.582 \pm 0.021}$	0.305 ± 0.022	0.081 ± 0.008
	FedPRS	$\mathbf{0.347 \pm 0.002}$	$\mathbf{0.331 \pm 0.000}$	0.509 ± 0.008	0.330 ± 0.001	0.202 ± 0.002

FedPRS improves by 0.186 Macro-F1 from $E = 1$ to $E = 10$; FedYoYo improves by 0.285 and nearly closes the aggregate gap at the largest budget. FedPRS's tail recall peaks at $E = 5$ (0.205) and remains near 0.20 thereafter even as its Macro-F1 increases. From $E = 5$ to $E = 10$, its head/medium recall rises from $0.472/0.316$ to $0.509/0.330$, whereas tail recall changes from 0.205 to 0.202. The accompanying Macro-F1 gain is modest (0.013), and a rising Macro-F1 curve should not be interpreted as monotone improvement for every frequency group.

The full budget sequence further qualifies the endpoint comparison. FedPRS's Macro-F1 margin is 0.083 at $E = 1$, increases to 0.105 at $E = 3$, and then decreases to 0.071, 0.054, and 0.011 at $E = 5, 7, 10$. The strongest reported comparator is FedPALS at $E = 1, 3, 5$ and FedYoYo at $E = 7, 10$. FedPRS also leads mean accuracy throughout, rising from 0.190 to 0.347, with its accuracy margin narrowing from 0.074 to 0.017 between the endpoints. Most of its Macro-F1 improvement occurs early: $E = 1$ to $E = 3$ yields 0.148, whereas $E = 3$ to $E = 10$ adds 0.038. At $E = 10$, FedYoYo exceeds FedPRS in both medium and tail recall ($0.340/0.289$ versus $0.330/0.202$), while FedPRS retains higher head recall (0.509 versus 0.363). Thus the remaining aggregate advantage at the largest budget coexists with a different class-group balance. The small final Macro-F1 margin is a descriptive three-seed mean difference, not evidence of statistical significance.