V-HFLoRA: Target-Basis Aggregation for Heterogeneous Federated LoRA

Anonymous authors
Paper under double-blind review

Abstract

Vision foundation models (VFMs) can be efficiently adapted with LoRA, but federated aggregation becomes challenging when clients use structurally heterogeneous frozen backbones. Differences in layer correspondence, parameter dimensions, and backbone-induced coordinate geometry can make direct cross-model aggregation inconsistent or undefined. We further observe that low LoRA rank does not imply low-dimensional support in frozen-backbone coordinates, motivating update-conditioned geometry. We propose **V-HFLoRA**, which formulates cross-backbone aggregation as a target-specific compatibility problem. It identifies functionally corresponding layers, transports matched source updates into the target geometry, and combines them through support-aware masking and reliability-weighted aggregation before low-rank recovery. We establish target-side convergence by separating collaborative bias from approximation errors due to matching, transport, aggregation, and recovery. Across heterogeneous-backbone benchmarks, V-HFLoRA consistently outperforms local training, same-backbone aggregation, aligned averaging, heterogeneous federated LoRA baselines, and prediction-space distillation. Mechanism diagnostics further verify the complementary roles of transport and masking.

1 Introduction

Parameter-efficient fine-tuning has made it practical to adapt vision foundation models (VFMs) on resource-constrained devices. Among existing approaches, LoRA is particularly attractive because it freezes the pretrained backbone and learns only low-rank adaptation matrices, reducing memory, computation, and communication costs. Combined with federated learning (FL), LoRA provides a natural mechanism for privacy-preserving VFM adaptation over decentralised client data.

Most federated LoRA methods, however, rely on clients sharing a compatible backbone or adapter parameterization. This assumption becomes restrictive when clients deploy structurally heterogeneous VFMs to accommodate different computation, memory, or latency constraints. In this setting, cross-model incompatibility arises at three levels. Functionally corresponding layers may occur at different depths, matched layers may have different parameter dimensions, and even equally shaped LoRA updates may be expressed in different coordinate systems induced by their frozen backbones. Consequently, tensor-shape compatibility alone is insufficient for meaningful cross-backbone aggregation.

Figure 1 illustrates two consequences of this heterogeneity. Figure 1(A) shows that low LoRA rank does not imply equally low-dimensional support in frozen-backbone coordinates: the number of directions required to retain 90% of the projected update energy varies substantially across attention modules. A fixed retained dimension can therefore discard informative directions for some updates while preserving weakly used directions for others. Figure 1(B) further shows that even iso-dimensional backbones need not share the same parameter geometry, whereas cross-dimensional backbones additionally introduce dimension and layer correspondence mismatch. These observations motivate treating heterogeneous federated LoRA aggregation as a *target-specific compatibility problem*.

To address this problem, we propose **V-HFLoRA**, a geometric aggregation framework for federated LoRA across structurally heterogeneous backbones. V-HFLoRA preserves each client's native LoRA parameterization and establishes compatibility only at aggregation time. It uses update-conditioned

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.

backbone geometry to represent local updates, functional similarity to identify corresponding layers, and compatible transport to rewrite matched source updates in the target geometry. For cross-dimensional pairs, the method first establishes a mutually retained dimension and performs Frobenius-optimal dimension alignment before basis-overlap transport. Residual coordinate conflicts are then removed through target-anchored masking, and the surviving contributions are reliability-weighted before low-rank recovery.

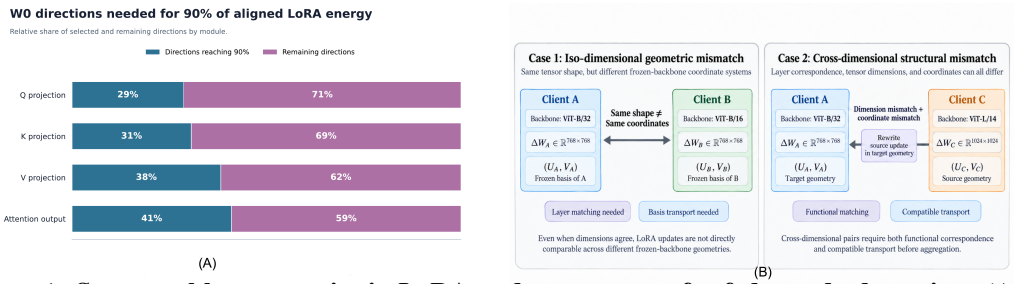


Figure 1: **Structural heterogeneity in LoRA update geometry for federated adaptation.** (A) The number of frozen-backbone directions required to retain 90% of LoRA-update energy varies across attention modules, motivating update-conditioned subspace selection. (B) Cross-backbone incompatibility arises even between iso-dimensional layers with different frozen-backbone coordinate systems, and can further involve parameter-dimension and layer-structure mismatch.

Our contributions are threefold. First, we formulate federated LoRA aggregation across structurally heterogeneous frozen backbones as a target-specific compatibility problem, highlighting layer correspondence, parameter dimensionality, and backbone-induced geometry as distinct obstacles to direct aggregation. Second, we show that low intrinsic LoRA rank does not imply low-dimensional support in backbone coordinates and develop V-HFLoRA to resolve these incompatibilities without redesigning local adapters. Third, we provide a target-side convergence analysis and extensive experiments showing consistent improvements over local training, same-backbone aggregation, aligned averaging, heterogeneous federated LoRA baselines, and prediction-space distillation, with ablations and mechanism diagnostics isolating the roles of the proposed components.

2 Related Work

FedProx Li et al. (2020) and SCAFFOLD Karimireddy et al. (2020) address statistical heterogeneity in federated learning, while recent federated PEFT methods extend LoRA to heterogeneous ranks and resource budgets, including FlexLoRA Bai et al. (2024) and FedRA Su et al. (2024). These extensions nevertheless typically assume that clients share either a compatible backbone architecture or a common parameter layout for aggregation.

A separate line of work considers collaboration across structurally heterogeneous models. HeteroTune Jia et al. (2024), Te-LoRA Li et al. (2025b), and Co-LoRA Seo et al. (2026) address cross-model heterogeneity through learned aggregation, factor alignment, or dimension-invariant adapter designs. V-HFLoRA instead preserves each client's native LoRA parameterisation and establishes cross-model compatibility with respect to the geometry induced by the corresponding frozen backbone.

FedDF Lin et al. (2020) and FedProto Tan et al. (2022) avoid direct parameter alignment by exchanging predictions or class prototypes. V-HFLoRA, by contrast, continues to communicate LoRA updates: calibration data are used only to identify correspondences across heterogeneous models rather than as the medium through which knowledge is transferred.

Related principles also appear in model merging and structure-aware federated optimisation. Git Re-Basin Ainsworth et al. (2022) and ZipIt! Stoica et al. (2024) align models before combining their parameters, while structure-aware federated methods suppress or reweight incompatible update components Zhang et al. (2026); Jhunjhunwala et al. (2025). V-HFLoRA combines these principles in a LoRA-specific setting: matched source updates are first transported into a target-specific coordinate system, after which aggregation is performed over the resulting target coordinates.

3 Method

We introduce **V-HFLoRA**, a target-specific geometric aggregation framework for federated LoRA across heterogeneous VFM backbones, as illustrated in Figure 2. Sec. 3.1 represents each local update relative to its frozen backbone through an update-conditioned coordinate support. Sec. 3.2 establishes

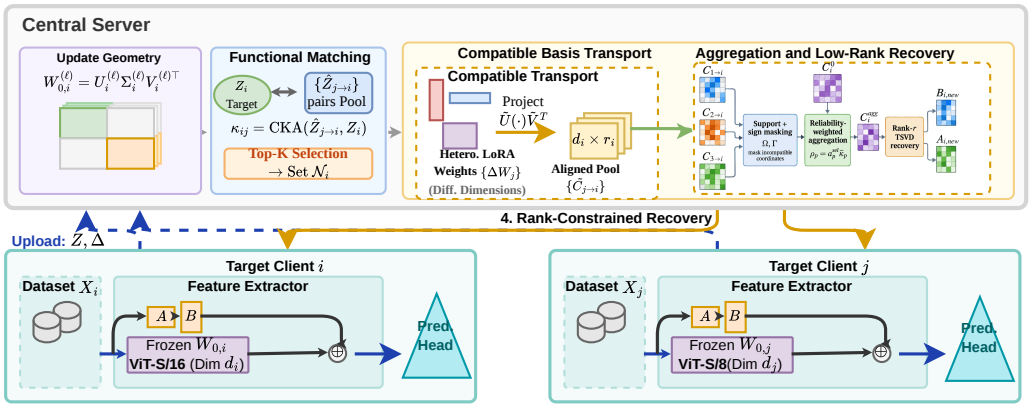


Figure 2: **Overview of the V-HFLoRA pipeline.** V-HFLoRA resolves structural LoRA-geometry heterogeneity through four stages.

layer correspondence across heterogeneous models using response-based functional matching. Given these correspondences, Sec. 3.3 transports source updates into the geometry of each target layer and performs cross-dimensional coordinate registration when necessary. Sec. 3.4 then performs support-aware masking and reliability-weighted aggregation in the target space, followed by low-rank recovery. Further implementation details are provided in Appendix B.

3.1 Update-Conditioned Backbone Geometry

A local LoRA update is defined relative to its frozen backbone weight. V-HFLoRA therefore represents the update in a fixed coordinate system induced by the singular directions of the frozen weight. For client i and layer ℓ, let $U_i^{(\ell)}$ and $V_i^{(\ell)}$ denote the output- and input-side singular bases of $W_{0,i}^{(\ell)}$. At communication round t, the local LoRA update $\Delta W_{i,t}^{(\ell)} = B_{i,t}^{(\ell)} A_{i,t}^{(\ell)}$ is represented as

$$H_{i,t}^{(\ell)} = U_i^{(\ell)\top} \Delta W_{i,t}^{(\ell)} V_i^{(\ell)}. \tag{1}$$

Equation equation 1 provides an exact change of coordinates without requiring the LoRA update to align with any particular singular directions. V-HFLoRA subsequently retains the smallest leading subspace that captures at least 90% of the projected update energy, defining an update-dependent dimension $q_{i,t}^{(\ell)}$ that can vary across clients, layers, and communication rounds. Unlike the prescribed LoRA rank $r_i^{(\ell)}$, $q_{i,t}^{(\ell)}$ determines the backbone-coordinate subspace retained for cross-model exchange. The precise selection criterion is given in Appendix B.1.

3.2 Functional Layer Matching

Layer index does not reliably indicate correspondence across structurally heterogeneous backbones, since functionally similar modules may appear at different depths. V-HFLoRA therefore establishes cross-backbone correspondence from layer responses to a common server-side unlabelled calibration set. Candidate matches are restricted to the same module family, such as Q-to-Q, K-to-K, V-to-V, and O-to-O. To account for differences in covariance structure, anisotropy, and effective dimension across backbone response spaces, V-HFLoRA applies CCA-based adjustment before measuring functional similarity with CKA.

For a candidate source layer (j, π) and target layer (i, ℓ), let $R_j^{(\pi)}$ and $R_i^{(\ell)}$ denote their responses to the same ordered calibration examples. After response reduction, regularised CCA maps the two response matrices into canonical coordinates:

$$\left(Z_j^{(\pi)}, Z_i^{(\ell)}\right) = \mathrm{CCA}\left(R_j^{(\pi)}, R_i^{(\ell)}\right). \tag{2}$$

A target-oriented ridge map $M_{\pi\to\ell}^*$ is then fitted to align the source canonical representation with the target canonical representation. Functional compatibility is measured by

$$\kappa_{\pi\to\ell} = \mathrm{CKA}\left(Z_j^{(\pi)} M_{\pi\to\ell}^*, Z_i^{(\ell)}\right). \tag{3}$$

The resulting scores are combined with module-family and monotone-order constraints to determine the matched source set $\mathcal{M}_i^{(\ell)}$. Full details are provided in Appendix B.2.

3.3 COMPATIBLE BASIS TRANSPORT

A functionally matched source update remains represented in the coordinate system induced by its own frozen backbone. Before aggregation, V-HFLoRA therefore rewrites the retained source update in the geometry of the matched target layer.

Same-dimensional transport. For same-dimensional layers, tensor shapes are compatible, but their backbone-induced coordinate systems need not coincide. Let $q_j = q_{j,t}^{(\pi)}$ and $q_i = q_{i,t}^{(\ell)}$. The retained source coordinate $C_{j,t}^{(\pi)}$ is expressed in the target retained basis through the overlaps between the corresponding source and target basis directions:

$$C_{j\to i,t}^{(\ell)} = U_{i,[q_i]}^{(\ell)\top} U_{j,[q_j]}^{(\pi)} C_{j,t}^{(\pi)} V_{j,[q_j]}^{(\pi)\top} V_{i,[q_i]}^{(\ell)}. \tag{4}$$

Different-dimensional transport. When the matched source and target layers differ in parameter dimension, their retained coordinate representations may also differ in size. V-HFLoRA first restricts both sides to the largest retained dimension shared by the pair:

$$q_{j\to i,t}^{(\pi,\ell)} = \min\left(q_{j,t}^{(\pi)}, q_{i,t}^{(\ell)}\right). \tag{5}$$

For $q = q_{j\to i,t}^{(\pi,\ell)}$, the source update is re-expressed using its first q frozen-backbone directions:

$$C_{j,t}^{(\pi,[q])} = U_{j,[q]}^{(\pi)\top} \Delta W_{j,t}^{(\pi)} V_{j,[q]}^{(\pi)} \in \mathbb{R}^{q\times q}. \tag{6}$$

The resulting $q \times q$ coordinate provides the common low-dimensional representation used for cross-dimensional transport. Because the retained source and target bases lie in different ambient parameter spaces, their basis overlap cannot be computed directly. V-HFLoRA therefore constructs Frobenius-optimal linear transforms between the retained source and target bases and uses these transforms to express the full $q \times q$ source coordinate in the target geometry. The closed-form transforms and their derivation are given in Appendix B.3.

3.4 TARGET-SPACE AGGREGATION AND LOW-RANK RECOVERY

After transport, eligible source updates are represented in a common target geometry, but their coordinate-wise contributions may still disagree. V-HFLoRA therefore applies coordinate-level masking before aggregation. Unsupported coordinates are excluded, as are transported coordinates whose signs conflict with the corresponding local target coordinate.

Let $\mathcal{S}_{i,t}^{(\ell)}[a,b]$ denote the set of sources retained at target coordinate $[a,b]$. The surviving contributions are weighted according to their matching reliability:

$$w_{i,p,t}^{(\ell)}[a,b] = \frac{a_{i,p,t}^{\mathrm{sel},(\ell)} \widetilde{\kappa}_{i,p,t}^{(\ell)}}{\sum_{p'\in\mathcal{S}_{i,t}^{(\ell)}[a,b]} a_{i,p',t}^{\mathrm{sel},(\ell)} \widetilde{\kappa}_{i,p',t}^{(\ell)}}, \qquad p \in \mathcal{S}_{i,t}^{(\ell)}[a,b]. \tag{7}$$

Here, $a_{i,p,t}^{\mathrm{sel},(\ell)}$ captures pair-selection reliability, while $\widetilde{\kappa}_{i,p,t}^{(\ell)}$ denotes the normalised CCA-adjusted CKA score.

Each target coordinate is then aggregated only over the surviving transported contributions:

$$C_{i,t}^{\mathrm{agg},(\ell)}[a,b] = \begin{cases} \displaystyle\sum_{p\in\mathcal{S}_{i,t}^{(\ell)}[a,b]} w_{i,p,t}^{(\ell)}[a,b]\, C_{p\to i,t}^{(\ell)}[a,b], & \mathcal{S}_{i,t}^{(\ell)}[a,b] \neq \emptyset, \\ C_{i,t}^{0,(\ell)}[a,b], & \mathcal{S}_{i,t}^{(\ell)}[a,b] = \emptyset. \end{cases} \tag{8}$$

Masking precedes weight normalisation, ensuring that unsupported or sign-conflicting coordinates contribute neither to the normalising denominator nor to the aggregate. If no compatible source remains, V-HFLoRA retains the corresponding local target coordinate. The aggregated coordinate is then projected to the prescribed LoRA rank by truncated SVD in the retained target space, yielding the LoRA factors for the next communication round. The precise support, masking, reliability, and low-rank recovery rules are given in Appendix B.4.

Algorithm 1 V-HFLoRA for Heterogeneous Federated VFM Adaptation

Input: client VFMs $\{W_{0,i}\}_{i\in\mathcal{C}}$, local data $\{\mathcal{D}_i\}$, server-side unlabelled calibration set $\mathcal{D}_{\text{cal}}$, LoRA ranks $\{r_i^{(\ell)}\}$, retained-energy threshold ρ.
Precompute: singular bases $\{U_i^{(\ell)}, V_i^{(\ell)}\}$ of the frozen backbone layers.
for $t = 0, \dots, T-1$ **do**
  Sample active clients $\mathcal{C}^t \subseteq \mathcal{C}$.
  for client $i \in \mathcal{C}^t$ **in parallel do**
    Perform local LoRA training and upload $\{B_{i,t}^{(\ell)}, A_{i,t}^{(\ell)}\}_\ell$.
  end for
  $\Delta W_{i,t}^{(\ell)} = B_{i,t}^{(\ell)} A_{i,t}^{(\ell)} \to H_{i,t}^{(\ell)} \to q_{i,t}^{(\ell)}$. ▷ *Update-Conditioned Backbone Geometry*
  $\{R_i^{(\ell)}\}_{\mathcal{D}_{\text{cal}}} \to$ CCA $\to$ ridge alignment $\to \{\kappa_{j\to i}^{(\pi,\ell)}\} \to \{\mathcal{M}_i^{(\ell)}\}$. ▷ *Functional Layer Matching*
  for target client $i \in \mathcal{C}^t$ **do**
    For each matched source–target pair, apply basis-overlap transport if dimensions agree; otherwise perform mutual-q Procrustes registration and transport, obtaining $\{C_{p\to i,t}^{(\ell)}, \Omega_{p\to i,t}^{(\ell)}\}$. ▷ *Compatible Transport*
    $\{C_{p\to i,t}^{(\ell)}, \Omega_{p\to i,t}^{(\ell)}\} \to \{\Gamma_{p,t}^{(\ell)}\} \to \{\mathcal{S}_{i,t}^{(\ell)}\} \to \{w_{i,p,t}^{(\ell)} \propto a_{i,p,t}^{\text{sel},(\ell)} \widetilde{\kappa}_{i,p,t}^{(\ell)}\} \to C_{i,t}^{\text{agg},(\ell)} \to C_{i,t}^{(r),(\ell)} \to \{B_{i,t+1}^{(\ell)}, A_{i,t+1}^{(\ell)}\}$. ▷ *Mask–Reliability Aggregation and Recovery*
    Send $\{B_{i,t+1}^{(\ell)}, A_{i,t+1}^{(\ell)}\}_\ell$ to target client i.
  end for
end for
return $\{B_{i,T}^{(\ell)}, A_{i,T}^{(\ell)}\}_{i\in\mathcal{C}}$.

4 Target-Space Stability and Convergence

We analyze V-HFLoRA from a target-wise personalized federated learning perspective, adopting the partial-model view of FedSim Pillutla et al. (2022) and PFed-NS Liu et al. (2025). For each target model, the frozen backbone defines a fixed coordinate system, while the update-conditioned retained dimension determines which coordinates participate in cross-client sharing at each round.

Let g_t° denote the ideal collaborative direction in the target space and $\widehat{g}_t$ the direction produced by the full V-HFLoRA pipeline. We separate the population-level discrepancy

$$B_t = \|\mathbb{E}_t[g_t^\circ] - \nabla F(y_t)\|_2^2$$

from the additional pipeline approximation error

$$\mathcal{E}_t = \mathbb{E}_t\left[\|\widehat{g}_t - g_t^\circ\|_2^2\right].$$

Under standard smoothness and bounded-variance assumptions, the analysis in Appendix C yields, for a sufficiently small step size,

$$\frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}\|\nabla F(y_t)\|_2^2 \le \mathcal{O}\left(\frac{1}{\eta T}\right) + \mathcal{O}(\overline{B}_T) + \mathcal{O}(\overline{\mathcal{E}}_T) + \mathcal{O}(\eta\sigma^2). \tag{9}$$

For V-HFLoRA, the pipeline error further decomposes into functional-matching failure, compatible-transport perturbation, mask–reliability aggregation error, and rank-constrained recovery:

$$\mathcal{E}_t \lesssim \delta_t^{\text{match}} + \mathcal{T}_t + \mathcal{W}_t + \tau_{r,t}^2. \tag{10}$$

The corresponding terms are controlled by matching-margin stability, cross-dimensional registration perturbation, simplex-constrained aggregation, and best rank-r approximation. Under the bounded-update conditions stated in Appendix C, their round averages admit finite bounds

$$\overline{B}_T \le \epsilon_B, \quad \overline{\delta}_T^{\text{match}} \le \epsilon_\delta, \quad \overline{\mathcal{T}}_T \le \epsilon_T, \quad \overline{\mathcal{W}}_T \le \epsilon_W, \quad \overline{\tau}_{r,T}^2 \le \epsilon_\tau,$$

where the constants are independent of the communication horizon T. Consequently, choosing $\eta = \Theta(T^{-1/2})$ gives

$$\frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}\|\nabla F(y_t)\|_2^2 \le \mathcal{O}(T^{-1/2}) + \mathcal{R}, \tag{11}$$

where

$$\mathcal{R} = C_B\epsilon_B + C_\delta\epsilon_\delta + C_T\epsilon_T + C_W\epsilon_W + C_\tau\epsilon_\tau < \infty.$$

Thus, V-HFLoRA converges to a finite first-order stationary neighborhood without requiring the approximation errors to vanish. When the residual terms additionally approach zero, $\mathcal{R} \to 0$, recovering the standard $\mathcal{O}(T^{-1/2})$ stationarity rate.

5 EXPERIMENTS

We evaluate V-HFLoRA on CIFAR-10, CIFAR-100 Alex (2009), and TinyImageNet Le et al. (2015). Each dataset is partitioned across $K = 10$ clients using a Dirichlet label split with $\alpha_{\text{dir}} = 0.5$.

Our main heterogeneous setting assigns TinyCLIP ViT-8M/16 and OpenCLIP ViT-B/32 backbones to different clients. The two vision encoders differ substantially in both width and depth: TinyCLIP ViT-8M/16 uses a hidden dimension of 256 with 10 Transformer layers, whereas ViT-B/32 uses a hidden dimension of 768 with 12 layers. This setting therefore introduces simultaneous layer-correspondence, parameter-dimension, and backbone-geometry mismatch, providing a direct test of cross-dimensional federated LoRA aggregation.

We compare V-HFLoRA against Local Only, Grouped-Backbone FedAvg, Aligned FedAvg, Co-LoRA Seo et al. (2026), and FedDF Lin et al. (2020) under the same client partitions and local training budget. Full implementation and baseline details are provided in Appendices B and F. We report mean personalized test accuracy across clients as the primary metric, together with per-backbone accuracy to account for the capacity gap between TinyCLIP and ViT-B/32. Additional experiments examine robustness to data heterogeneity, same-backbone aggregation, and the contributions of individual V-HFLoRA components.

5.1 MAIN RESULTS

Table 1 reports the main results under cross-dimensional backbone heterogeneity, where clients use either TinyCLIP ViT-8M/16 ($d = 256$) or CLIP ViT-B/32 ($d = 768$). The two backbones differ in both width and depth, introducing simultaneous mismatch in layer correspondence, parameter dimensions, and backbone-induced update geometry.

Restricting collaboration to clients with the same backbone already provides substantial gains over Local Only, but leaves information from the other backbone unused. Aligned FedAvg enables cross-backbone sharing and further improves over Grouped FedAvg by 0.05, 1.81, and 3.20 percentage points on CIFAR-10, CIFAR-100, and TinyImageNet, respectively. This indicates that cross-model information can be beneficial once a basic compatibility mechanism is introduced.

Among the heterogeneous LoRA baselines, Co-LoRA achieves the strongest performance by constructing a dimension-invariant shared parameterization. V-HFLoRA further improves over Co-LoRA by 0.27, 0.57, and 1.00 percentage points on CIFAR-10, CIFAR-100, and TinyImageNet, respectively, while preserving the native LoRA parameterization of each backbone. In contrast to Co-LoRA, V-HFLoRA establishes compatibility at aggregation time by matching functionally corresponding layers, transporting source updates into the target geometry, and resolving residual target-space conflicts before aggregation. The improvements are consistent across all three datasets, with larger gains on CIFAR-100 and TinyImageNet.

Table 1: Main results under heterogeneous OpenCLIP backbones.

	Local	Same-space	Aligned-space	Dimension-invariant	Prediction-space	Target-space	Improvement
Dataset	Local Only	Grouped FedAvg	Aligned FedAvg	Co-LoRA	FedDF	**V-HFLoRA**	Gain
CIFAR-10	86.21 ± 1.92	95.73 ± 1.14	95.78 ± 1.02	96.05 ± 1.20	94.30 ± 1.37	$\mathbf{96.32 \pm 0.50}$	$+0.27$
CIFAR-100	66.48 ± 1.86	70.00 ± 1.31	71.81 ± 1.18	75.86 ± 1.05	65.28 ± 1.72	$\mathbf{76.43 \pm 0.74}$	$+0.57$
TinyImageNet	60.87 ± 1.95	64.92 ± 1.44	68.12 ± 1.26	69.84 ± 1.10	59.08 ± 1.83	$\mathbf{70.84 \pm 0.88}$	$+1.00$

5.2 ABLATION STUDY

We conduct ablation studies on CIFAR-100 under the cross-dimensional TinyCLIP–CLIP setting, while keeping the client partition, local training budget, LoRA configuration, communication schedule, and backbone assignments fixed. We examine three aspects of V-HFLoRA: functional correspondence, update-conditioned backbone geometry, and the complementary effects of cross-dimensional transport and target-space aggregation.

Representation and matching. Table 2 first evaluates how V-HFLoRA constructs compatible representations across heterogeneous backbones. Removing any major component reduces accuracy relative to the full model. The largest drops occur when coordinate masking or CCA adjustment is removed, reducing accuracy by 3.01 and 2.62 percentage points, respectively.

For functional correspondence, fixed layer-index matching attains 72.36 accuracy, while response-based CKA improves this to 73.81. Applying CCA before CKA further increases accuracy to

75.62, and enforcing monotone layer ordering yields the full 76.43. These results indicate that both representation-level similarity and cross-model layer order are useful for establishing correspondence between backbones with different depths.

The retained-geometry ablation shows a complementary effect. Using fixed small and medium retained subspaces gives 74.66 and 75.48, respectively, while retaining the full basis reaches 75.71. Dynamic retention achieves the highest accuracy of 76.43 while using only about 44% of the backbone directions on average. Thus, performance is not explained by retaining more coordinates; rather, adapting the retained geometry to the local update provides a more effective representation for cross-backbone aggregation.

Table 2: **Representation and matching ablations on CIFAR-100 under the cross-dimensional TinyCLIP–CLIP setting.**

Setting	Auxiliary statistic	Acc.
(a) Overall component ablation		
w/o dynamic retention	–	75.71
w/o CCA adjustment	–	73.81
w/o monotone matching	–	75.62
w/o coordinate mask	–	73.42
w/o reliability weighting	–	74.18
Full V-HFLoRA	–	**76.43**
(b) Functional matching		
Layer index	–	72.36
Raw CKA	–	73.81
CCA–CKA	–	75.62
CCA–CKA + monotone matching	–	**76.43**
(c) Update-conditioned retained geometry		
Fixed small q	25% retained	74.66
Fixed medium q	50% retained	75.48
Full basis	100% retained	75.71
Dynamic q	**44% retained**	**76.43**

Cross-dimensional transport and target-space aggregation. Table 3 separates the effect of cross-dimensional transport from the post-transport aggregation rule. Starting from aligned averaging, reliability weighting alone improves accuracy from 71.81 to 73.42, while target-anchored masking alone reaches 74.18. Combining masking and reliability weighting yields 76.43, an improvement of 4.62 percentage points over aligned averaging. This suggests that donor reliability and coordinate-level directional consistency capture complementary aspects of post-transport compatibility.

The lower part of Table 3 directly examines width mismatch between the 256-dimensional TinyCLIP backbone and the 768-dimensional CLIP backbone. Simple truncation or padding achieves 66.21, while a generic linear projection improves accuracy only modestly to 67.48. Compatible transport followed by ordinary averaging reaches 70.62, showing that explicitly rewriting the source update in the target geometry provides a substantially stronger cross-dimensional correspondence than shape adjustment alone. Adding mask–reliability aggregation further raises accuracy to 76.43. These results separate the roles of the two stages: compatible transport addresses cross-backbone geometric mismatch, whereas target-space aggregation handles the residual coordinate conflicts that remain after transport.

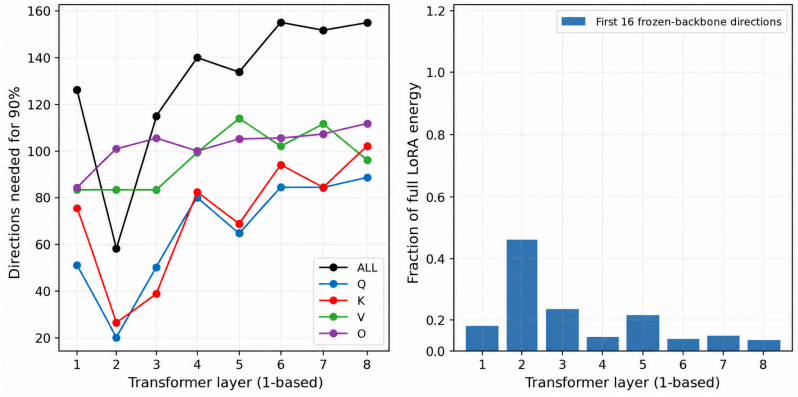
5.3 Mechanism Diagnostics

Update-conditioned retained geometry. Figure 3(A) shows that the retained dimension required to preserve 90% of the projected LoRA-update energy varies substantially across Transformer layers and attention modules. In many cases, this dimension is well above the LoRA rank $r = 16$, and a fixed set of 16 frozen-backbone directions captures only a limited fraction of the update energy. Thus, the intrinsic rank of a LoRA update does not determine how broadly its energy is distributed over the frozen-backbone geometry, motivating the update-dependent retained dimension $q_{i,t}^{(\ell)}$.

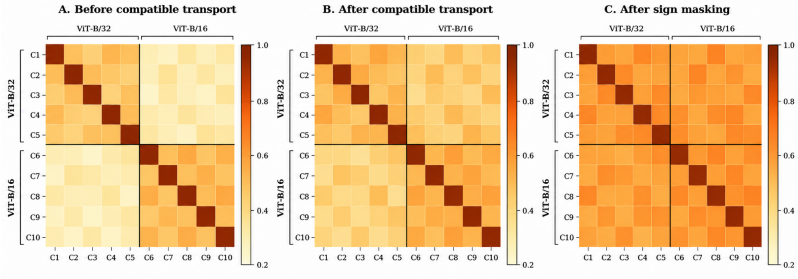
Transport and masking diagnostics. Figure 3(B) compares pairwise client-update similarity at three stages. Before compatible transport, same-backbone clients are moderately similar, whereas cross-backbone pairs are generally less compatible. After transport, the same-backbone blocks

Table 3: **Cross-dimensional transport and target-space aggregation ablations on CIFAR-100.**

Setting	Mask	Reliability weight	Acc.
(a) Target-space aggregation			
Aligned average	×	×	71.81
Reliability weight only	×	✓	73.42
Mask only	✓	×	74.18
Full mask–weight aggregation	✓	✓	**76.43**
(b) Cross-dimensional transport			
Local only	No cross-model transfer		66.48
Truncate / pad	Shape adjustment only		66.21
Generic linear projection	Dimension alignment only		67.48
Compatible transport + average	Geometry transport		70.62
Compatible transport + mask–weight	Transport + aggregation		**76.43**



(a) **Update-conditioned retained geometry.** The number of frozen-backbone directions required to preserve 90% of the projected LoRA-update energy varies substantially across layers and Q/K/V/O modules, while a fixed 16-direction basis captures only a limited fraction of the update energy.



(b) **Effect of transport and masking on cross-client compatibility.** Pairwise update similarity is shown before compatible transport, after transport into the target geometry, and after sign masking. Cross-backbone similarities increase selectively after transport, while masking further improves the retained target-space agreement.

Figure 3: **Mechanism diagnostics of V-HFLoRA.**

remain largely unchanged, while many cross-backbone pairs become more similar, indicating that transport mainly resolves backbone-induced geometric mismatch rather than uniformly increasing all pairwise similarities. Sign masking then produces a further, more moderate increase by removing coordinates that remain directionally inconsistent in the target space. After masking, cross-backbone similarities become comparable to those within the same backbone, and some heterogeneous pairs are as compatible as, or more compatible than, homogeneous pairs. Together, these results support the sequence of compatible transport followed by conflict-aware target-space aggregation.

5.4 Scalability with the Number of Clients

We evaluate scalability by increasing the number of heterogeneous clients from $K = 5$ to $K = 35$ on CIFAR-100 while keeping the local training configuration and backbone-assignment protocol un-

changed. Table 4 compares V-HFLoRA with Aligned FedAvg, which uses cross-backbone alignment but directly averages the transported updates.

Table 4: **Scalability with the number of heterogeneous clients on CIFAR-100.**

K	5	10	15	20	25	30	35
Aligned FedAvg	72.0	71.8	71.4	71.0	70.6	70.2	69.8
V-HFLoRA	**75.3**	**75.7**	**76.1**	**76.1**	**76.0**	**75.8**	**75.5**
Gain	+3.3	+3.9	+4.7	+5.1	+5.4	+5.6	+5.7

As K increases from 5 to 35, Aligned FedAvg decreases from 72.0% to 69.8%, whereas V-HFLoRA remains between 75.3% and 76.1%. Accordingly, the advantage of V-HFLoRA over direct aligned averaging widens from 3.3 to 5.7 percentage points. This pattern is consistent with the motivation for conflict-aware aggregation: as the donor pool grows, increasing the number of transported source updates does not necessarily make direct averaging more reliable, whereas selective masking and weighting can prevent incompatible source coordinates from dominating the target update.

5.5 Robustness to Data Heterogeneity

We examine whether V-HFLoRA remains effective under different degrees of statistical heterogeneity. We vary the Dirichlet concentration parameter α_{dir}, where a smaller value produces a more strongly non-IID client label distribution, while keeping all other experimental settings unchanged. Following the personalized evaluation protocol used throughout the paper, each client is evaluated on its corresponding local test partition, and we report the average personalized accuracy across clients.

Table 5: **Robustness to label heterogeneity on CIFAR-100.**

Method	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 1.0$
Local Only	69.14	67.82	66.48	64.91
Grouped FedAvg	72.36	71.18	70.00	68.54
Aligned FedAvg	74.08	72.94	71.81	70.37
V-HFLoRA	**78.12**	**77.21**	**76.43**	**75.18**

V-HFLoRA achieves the highest personalized accuracy at every tested Dirichlet concentration. Relative to the strongest baseline, Aligned FedAvg, V-HFLoRA improves accuracy by 4.04, 4.27, 4.62, and 4.81 percentage points for $\alpha_{\text{dir}} = 0.1$, 0.3, 0.5, and 1.0, respectively. Thus, its advantage is preserved across the tested range of label heterogeneity. This pattern is consistent with the proposed target-space aggregation remaining effective under substantial variation in client label distributions.

6 Conclusion

We studied federated LoRA aggregation across structurally heterogeneous VFM backbones, where layer correspondence, parameter dimensions, and backbone-induced geometry prevent direct cross-model aggregation. We introduced **V-HFLoRA** to establish target-specific compatibility at aggregation time while preserving native local LoRA parameterizations. Our target-side analysis separates collaborative bias from approximation errors introduced by matching, transport, aggregation, and low-rank recovery. Experiments across heterogeneous-backbone settings show consistent improvements over existing aggregation and cross-model collaboration baselines, while ablations and mechanism diagnostics support the complementary roles of update-conditioned geometry, compatible transport, and conflict-aware aggregation.

Reproducibility Statement

We provide the information required to reproduce V-HFLoRA and its experimental evaluation throughout the paper and appendices. The complete algorithmic specification, including functional matching, compatible transport, mask–reliability aggregation, and low-rank recovery, is given in the method section and accompanying appendices. Implementation details, training configurations, hyperparameters, dataset partitioning, and calibration settings are reported in Appendix B, while baseline implementations and evaluation protocols are described in Appendix F. The assumptions and complete proofs supporting the theoretical results are provided in the convergence-analysis appendix.

AI USE STATEMENT

In this work, we did not use generative AI tools for tasks requiring disclosure, including theoretical development, mathematical claims or proofs, research methodology or experimental design, implementation, data processing, or interpretation of results. We used generative AI tools only for language editing and improving the readability of the manuscript, which falls under the recommended disclosure category. We reviewed all AI-assisted edits and take responsibility for the final content of this work.

REFERENCES

Samuel K Ainsworth, Jonathan Hayase, and Siddhartha Srinivasa. Git re-basin: Merging models modulo permutation symmetries. *arXiv preprint arXiv:2209.04836*, 2022.

Krizhevsky Alex. Learning multiple layers of features from tiny images. *https://www. cs. toronto. edu/kriz/learning-features-2009-TR. pdf*, 2009.

Jiamu Bai, Daoyuan Chen, Bingchen Qian, Liuyi Yao, and Yaliang Li. Federated fine-tuning of large language models under heterogeneous tasks and client resources. *Advances in Neural Information Processing Systems*, 37:14457–14483, 2024.

Jieming Bian, Lei Wang, Letian Zhang, and Jie Xu. Lora-fair: Federated lora fine-tuning with aggregation and initialization refinement. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3737–3746. IEEE, 2025.

Jieming Bian, Lei Wang, Letian Zhang, and Jie Xu. Fedtreelora: Reconciling statistical and functional heterogeneity in federated lora fine-tuning. *arXiv preprint arXiv:2603.13282*, 2026.

Peva Blanchard, El Mahdi El Mhamdi, Rachid Guerraoui, and Julien Stainer. Machine learning with adversaries: Byzantine tolerant gradient descent. *Advances in neural information processing systems*, 30, 2017.

Yae Jee Cho, Luyang Liu, Zheng Xu, Aldi Fahrezi, and Gauri Joshi. Heterogeneous lora for federated fine-tuning of on-device foundation models. In *Proceedings of the 2024 conference on empirical methods in natural language processing*, pages 12903–12913, 2024.

Corinna Cortes, Mehryar Mohri, and Afshin Rostamizadeh. Algorithms for learning kernels based on centered alignment. *The Journal of Machine Learning Research*, 13:795–828, 2012.

Moming Duan, Duo Liu, Xinyuan Ji, Renping Liu, Liang Liang, Xianzhang Chen, and Yujuan Tan. Fedgroup: Efficient federated learning via decomposed similarity-based clustering. In *2021 IEEE Intl Conf on parallel & distributed processing with applications, big data & cloud computing, sustainable computing & communications, social computing & networking (ISPA/BDCloud/SocialCom/SustainCom)*, pages 228–237. IEEE, 2021.

Boyu Fan, Xiang Su, Sasu Tarkoma, and Pan Hui. Helora: Lora-heterogeneous federated fine-tuning for foundation models. *ACM Transactions on Internet Technology*, 25(2):1–22, 2025.

Farzad Farhadzadeh, Debasmit Das, Shubhankar Borse, and Fatih Porikli. Lora-x: Bridging foundation models with training-free cross-model adaptation. In *International Conference on Learning Representations*, volume 2025, pages 9835–9858, 2025.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021.

Gabriel Ilharco, Marco Tulio Ribeiro, Mitchell Wortsman, Suchin Gururangan, Ludwig Schmidt, Hannaneh Hajishirzi, and Ali Farhadi. Editing models with task arithmetic. *arXiv preprint arXiv:2212.04089*, 2022.

Divyansh Jhunjhunwala, Arian Raje, Madan Ravi Ganesh, Chaithanya Kumar Mummadi, Chaoqun Dong, Jiawei Zhou, Wan-Yi Lin, Gauri Joshi, and Zhenzhen Li. Fedrpca: Enhancing federated lora aggregation using robust pca. *arXiv preprint arXiv:2506.01194*, 2025.

Ruofan Jia, Weiying Xie, Jie Lei, Haonan Qin, Jitao Ma, and Leyuan Fang. Towards efficient model-heterogeneity federated learning for large models. *arXiv preprint arXiv:2411.16796*, 2024.

Sai Praneeth Karimireddy, Satyen Kale, Mehryar Mohri, Sashank Reddi, Sebastian Stich, and Ananda Theertha Suresh. Scaffold: Stochastic controlled averaging for federated learning. In *International conference on machine learning*, pages 5132–5143. PMLR, 2020.

Yann Le, Xuan Yang, et al. Tiny imagenet visual recognition challenge. *CS 231N*, 7(7):3, 2015.

Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. *Proceedings of Machine learning and systems*, 2:429–450, 2020.

Yanan Li, Fanxu Meng, Muhan Zhang, Shiai Zhu, Shangguang Wang, and Mengwei Xu. Lorasuite: Efficient lora adaptation across large language model upgrades. *arXiv preprint arXiv:2505.13515*, 2025a.

Yanan Li, Fanxu Meng, Muhan Zhang, Shiai Zhu, Shangguang Wang, and Mengwei Xu. Lorasuite: Efficient lora adaptation across large language model upgrades. *Advances in Neural Information Processing Systems*, 38:170149–170176, 2026.

Zhixuan Li, Binqian Xu, Xiangbo Shu, Jiachao Zhang, Yazhou Yao, Guo-Sen Xie, and Jinhui Tang. Tensor-aggregated lora in federated fine-tuning. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1058–1067. IEEE, 2025b.

Tao Lin, Lingjing Kong, Sebastian U Stich, and Martin Jaggi. Ensemble distillation for robust model fusion in federated learning. *Advances in neural information processing systems*, 33:2351–2363, 2020.

Yuchen Liu, Ligang He, Zhigao Zhang, and Shenyuan Ren. Pfed-ns: An adaptive personalized federated learning scheme through neural network segmentation. *IEEE Transactions on Computers*, 74(6):1936–1948, 2025.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.

Krishna Pillutla, Kshitiz Malik, Abdel-Rahman Mohamed, Mike Rabbat, Maziar Sanjabi, and Lin Xiao. Federated learning with partial model personalization. In *International Conference on Machine Learning*, pages 17716–17758. PMLR, 2022.

Peter H Schönemann. A generalized solution of the orthogonal procrustes problem. *Psychometrika*, 31(1):1–10, 1966.

Minhyuk Seo, Taeheon Kim, Hankook Lee, Jonghyun Choi, and Tinne Tuytelaars. Co-lora: Collaborative model personalization on heterogeneous multi-modal clients. In *International Conference on Learning Representations*, volume 2026, pages 16450–16505, 2026.

George Stoica, Daniel Bolya, Jakob Bjorner, Pratik Ramesh, Taylor Hearn, and Judy Hoffman. Zipit! merging models from different tasks without training. In *International Conference on Learning Representations*, volume 2024, pages 29215–29237, 2024.

Shangchao Su, Bin Li, and Xiangyang Xue. Fedra: A random allocation strategy for federated tuning to unleash the power of heterogeneous clients. In *European Conference on Computer Vision*, pages 342–358. Springer, 2024.

Youbang Sun, Zitao Li, Yaliang Li, and Bolin Ding. Improving lora in privacy-preserving federated learning. In *International conference on learning representations*, volume 2024, pages 17978–17994, 2024.

Yue Tan, Guodong Long, Lu Liu, Tianyi Zhou, Qinghua Lu, Jing Jiang, and Chengqi Zhang. Fedproto: Federated prototype learning across heterogeneous clients. In *Proceedings of the AAAI conference on artificial intelligence*, pages 8432–8440, 2022.

Ziyao Wang, Zheyu Shen, Yexiao He, Guoheng Sun, Hongyi Wang, Lingjuan Lyu, and Ang Li. Flora: Federated fine-tuning large language models with heterogeneous low-rank adaptations. *Advances in Neural Information Processing Systems*, 37:22513–22533, 2024.

Feifan Xia, Mingyang Liao, Yuyang Fang, Defang Li, Yantong Xie, Weikang Li, Yang Li, Deguo Xia, and Jizhou Huang. Cross-lora: A data-free lora transfer framework across heterogeneous llms. *arXiv preprint arXiv:2508.05232*, 2025.

Lishan Yang, Nam Kha Nguygen, Po Hu, Wei Emma Zhang, Yanjun Shu, Mong Yuan Sim, and Weitong Chen. Fedilora: Heterogeneous lora for federated multimodal fine-tuning under missing modalities. *arXiv e-prints*, pages arXiv–2509, 2025.

Liping Yi, Han Yu, Gang Wang, Xiaoguang Liu, and Xiaoxiao Li. pfedlora: Model-heterogeneous personalized federated learning with lora tuning. *arXiv preprint arXiv:2310.13283*, 2023.

Dong Yin, Yudong Chen, Ramchandran Kannan, and Peter Bartlett. Byzantine-robust distributed learning: Towards optimal statistical rates. In *International conference on machine learning*, pages 5650–5659. Pmlr, 2018.

Weiyao Zhang, Jinyang Li, Qi Song, Miao Wang, Chungang Lin, Haitong Luo, Xuying Meng, and Yujun Zhang. Heterogeneity-oblivious robust federated learning. In *IEEE INFOCOM 2026-IEEE Conference on Computer Communications*, pages 1–10. IEEE, 2026.

A Related Work

Federated fine-tuning under statistical and resource heterogeneity. Heterogeneity has long been a central concern in federated learning. FedProx Li et al. (2020) and SCAFFOLD Karimireddy et al. (2020) address client drift under non-IID data and system variability. With the emergence of foundation models, LoRA Hu et al. (2021) has enabled parameter-efficient federated fine-tuning, including privacy-preserving LoRA adaptation Sun et al. (2024). More recent methods relax homogeneous-client assumptions from the perspectives of task, resource, capacity, or LoRA-rank heterogeneity. FlexLoRA adapts global LoRA parameters to heterogeneous client ranks Bai et al. (2024); FedRA accommodates heterogeneous resources through layer allocation Su et al. (2024); FediLoRA considers rank and modality heterogeneity Yang et al. (2025); and FedTreeLoRA studies statistical and functional heterogeneity in federated LoRA fine-tuning Bian et al. (2026). These methods substantially broaden the scope of federated PEFT, but mainly address heterogeneity while preserving a sufficiently compatible architecture or aggregation layout.

Model-heterogeneous federated PEFT. Recent work has begun to address parameter-efficient collaboration across heterogeneous client models more directly. HeteroTune proposes FedAdapter, a lightweight multi-branch cross-model aggregator for federated fine-tuning with different client model scales Jia et al. (2024). Te-LoRA addresses aggregation noise and LoRA misalignment through pre-aggregation alignment and tensor-to-matrix decomposition of heterogeneous LoRA factors Li et al. (2025b). Co-LoRA introduces a dimension-invariant LoRA module together with task-relevance-aware aggregation to support personalized collaboration across heterogeneous client architectures Seo et al. (2026). These methods are the closest to our setting, but they address structural heterogeneity through adapter design, factor alignment, or dimension-invariant sharing. V-HFLoRA instead keeps the native client adapters and explicitly uses the frozen backbone as the geometric reference: source layers are matched by function, their updates are rewritten in each target layer's retained backbone geometry, and aggregation is performed only after coordinate-level compatibility filtering.

Prediction- and representation-space collaboration. An alternative to parameter-space aggregation is to exchange knowledge in prediction or representation space. FedDF performs server-side ensemble distillation and therefore permits collaboration without directly averaging heterogeneous model parameters Lin et al. (2020). FedProto instead communicates class prototypes, allowing clients with heterogeneous data distributions and model architectures to share class-level representations Tan et al. (2022). These approaches avoid explicit parameter-coordinate alignment, but require knowledge transfer through auxiliary predictions or representations. V-HFLoRA retains parameter-efficient LoRA communication and uses a small unlabelled calibration set only to establish functional layer correspondence rather than as the medium through which client knowledge is distilled.

Cross-model transport and structure-aware aggregation. V-HFLoRA is also related to model merging and cross-model adaptation. Task arithmetic Ilharco et al. (2022), Git Re-Basin Ainsworth et al. (2022), and ZipIt! Stoica et al. (2024) show that parameters or task updates can be combined after suitable alignment. LoRA-X, Cross-LoRA, LoRASuite, and Theseus further investigate adapter or task-vector transfer across model versions and architectures Farhadzadeh et al. (2025); Xia et al. (2025); Li et al. (2025a; 2026). Robust federated aggregation instead studies unreliable or conflicting updates in a common aggregation space, from classical Byzantine-robust methods Blanchard et al. (2017); Yin et al. (2018) to structure-aware LoRA aggregation such as Horus and FedRPCA Zhang et al. (2026); Jhunjhunwala et al. (2025). V-HFLoRA combines these two perspectives in a different way: cross-model compatibility is first established through functional matching and backbone-relative transport, after which coordinate-wise masking and reliability weighting control which transported contributions enter each target update.

B Implementation Details of V-HFLoRA

This section describes the end-to-end implementation of V-HFLoRA. The procedure follows four stages: *update-conditioned backbone geometry*, *functional layer matching*, *compatible basis transport*, and *target-space mask–weight aggregation with low-rank recovery*. For compactness, we write $p = (j, \pi)$ for a matched source layer when no ambiguity arises.

B.1 Update-Conditioned Backbone Geometry

A LoRA adapter is not an isolated parameter object, but a low-rank update to a particular frozen backbone weight. Before establishing cross-model correspondence or aggregation, we therefore characterise each local LoRA update relative to the geometry of the backbone on which it was trained.

For client i and layer ℓ, let

$$W_{0,i}^{(\ell)} = U_i^{(\ell)} \Sigma_i^{(\ell)} V_i^{(\ell)\top} \tag{12}$$

denote the singular value decomposition of the frozen backbone weight. The purpose of this decomposition is not to truncate or compress $W_{0,i}^{(\ell)}$. Instead, $U_i^{(\ell)}$ and $V_i^{(\ell)}$ provide output-side and input-side reference bases defined by the frozen backbone itself. Because the backbone remains fixed during LoRA adaptation, these bases are computed once before federated training and reused across communication rounds.

At round t, the local LoRA update is

$$\Delta W_{i,t}^{(\ell)} = B_{i,t}^{(\ell)} A_{i,t}^{(\ell)}. \tag{13}$$

We express this update in the frozen-backbone coordinates as

$$H_{i,t}^{(\ell)} = U_i^{(\ell)\top} \Delta W_{i,t}^{(\ell)} V_i^{(\ell)}. \tag{14}$$

Hence, $H_{i,t}^{(\ell)}$ describes how the current LoRA update is distributed over the singular directions of its own frozen backbone. The backbone geometry is fixed, whereas the update energy carried by these directions can change with local training and across communication rounds.

We next identify the smallest backbone subspace that captures the dominant projected LoRA update energy. For the subspace spanned by the first q frozen-backbone singular directions, define

$$\eta_{i,t}^{(\ell)}(q) = \frac{\left\| U_{i,[q]}^{(\ell)\top} \Delta W_{i,t}^{(\ell)} V_{i,[q]}^{(\ell)} \right\|_F^2}{\left\| H_{i,t}^{(\ell)} \right\|_F^2}. \tag{15}$$

Equivalently, the numerator measures the energy of the leading $q \times q$ block of $H_{i,t}^{(\ell)}$ under the frozen-backbone ordering.

We fix the retained-energy threshold to

$$\rho = 0.9 \tag{16}$$

throughout all experiments and select

$$q_{i,t}^{(\ell)} = \min \left\{ q : \eta_{i,t}^{(\ell)}(q) \geq \rho \right\}. \tag{17}$$

Thus, $W_{0,i}^{(\ell)}$ determines a fixed reference coordinate system, whereas the current LoRA update determines how much of this geometry is retained for cross-client knowledge exchange. The resulting dimension $q_{i,t}^{(\ell)}$ can vary across clients, layers, and communication rounds.

Importantly, $q_{i,t}^{(\ell)}$ is distinct from the LoRA rank $r_i^{(\ell)}$. The LoRA rank constrains

$$\operatorname{rank}\left(\Delta W_{i,t}^{(\ell)} \right) \leq r_i^{(\ell)},$$

whereas $q_{i,t}^{(\ell)}$ characterises how broadly this low-rank update is distributed over the frozen-backbone coordinate system. A rank-r update can therefore have non-zero projections over substantially more than r backbone singular directions.

B.2 Functional Layer Matching

The backbone-coordinate representation determines how an update is represented, but not which layers should exchange information. V-HFLoRA therefore establishes source–target correspondence from functional responses rather than architectural indices.

Matching responses. At each communication round, each client performs local LoRA training and uploads its layerwise LoRA factors

$$\{B_{i,t}^{(\ell)}, A_{i,t}^{(\ell)}\},$$

The server maintains public copies of the frozen pretrained backbones used by the clients and reconstructs each adapted model by inserting the uploaded LoRA modules into the corresponding backbone. These adapted models are evaluated on a common server-side unlabelled calibration set to obtain layerwise responses. Thus, functional matching is performed entirely on the server and does not require private client data or client-side activations.

For a candidate source layer u and target layer v belonging to the same module family, let

$$R_u \in \mathbb{R}^{n_{\mathrm{cal}} \times d_u}, \qquad R_v \in \mathbb{R}^{n_{\mathrm{cal}} \times d_v} \tag{18}$$

denote their centred response matrices on the same ordered calibration examples. These response matrices are used only to establish functional layer correspondence. They are distinct from the frozen-backbone parameter coordinates used for LoRA transport.

Adjustment before CKA. Directly computing CKA on R_u and R_v can make the matching score sensitive to backbone-specific response geometry. Structurally heterogeneous backbones can exhibit different marginal covariance spectra, anisotropy, and effective response dimensions, such that raw CKA may reflect both functional correspondence and backbone-induced representation structure. We therefore apply dimensionality reduction and canonical adjustment before computing the final similarity score.

The source and target responses are first reduced independently using PCA/SVD. For the source response,

$$R_u = \widetilde{U}_u \widetilde{\Sigma}_u \widetilde{V}_u^\top, \tag{19}$$

we retain the smallest k_u satisfying

$$\frac{\sum_{a=1}^{k_u} \widetilde{\Sigma}_{u,aa}^2}{\sum_a \widetilde{\Sigma}_{u,aa}^2} \geq \tau_{\mathrm{pca}}, \qquad k_u \leq d_{\mathrm{pca}}^{\max}, \tag{20}$$

and define

$$A_u = R_u \widetilde{V}_{u,[:,1:k_u]}. \tag{21}$$

The target response is reduced analogously to obtain A_v.

We then apply regularised CCA. Define

$$\Sigma_{xx} = \frac{1}{n_{\mathrm{cal}} - 1} A_u^\top A_u + \varepsilon_{\mathrm{reg}} I, \tag{22}$$

$$\Sigma_{yy} = \frac{1}{n_{\mathrm{cal}} - 1} A_v^\top A_v + \varepsilon_{\mathrm{reg}} I, \tag{23}$$

and

$$\Sigma_{xy} = \frac{1}{n_{\mathrm{cal}} - 1} A_u^\top A_v. \tag{24}$$

Let

$$\Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1/2} = U_{\mathrm{c}} \Lambda V_{\mathrm{c}}^\top. \tag{25}$$

We retain canonical directions satisfying the implemented correlation threshold,

$$k_{\mathrm{cca}} = \#\{a : \lambda_a > \tau_{\mathrm{cca}}\}, \tag{26}$$

and discard the candidate pair if $k_{\mathrm{cca}} = 0$. Otherwise, the canonical response coordinates are

$$Z_u = A_u \Sigma_{xx}^{-1/2} U_{\mathrm{c},[:,1:k_{\mathrm{cca}}]}, \tag{27}$$

$$Z_v = A_v \Sigma_{yy}^{-1/2} V_{\mathrm{c},[:,1:k_{\mathrm{cca}}]}. \tag{28}$$

This adjustment reduces sensitivity to backbone-specific marginal response geometry and places the two layers in coordinates that emphasise their shared functional variation.

Directional functional matching. Because matching is evaluated from a source layer towards a particular target layer, we further fit a target-oriented ridge alignment in the canonical response space:

$$M^*_{u\to v} = \arg\min_M \|Z_u M - Z_v\|_F^2 + \lambda_{\mathrm{m}}\|M\|_F^2. \tag{29}$$

The corresponding directional functional compatibility score is

$$\kappa_{u\to v} = \mathrm{CKA}\left(Z_u M^*_{u\to v}, Z_v\right), \tag{30}$$

where linear CKA is

$$\mathrm{CKA}(X,Y) = \frac{\|X^\top Y\|_F^2}{\|X^\top X\|_F \|Y^\top Y\|_F}. \tag{31}$$

The map $M^*_{u\to v}$ is used only to evaluate functional layer correspondence in response space; it does not transport LoRA parameters.

The server evaluates candidate pairs only within the same module family, such as Q-to-Q, K-to-K, V-to-V, and O-to-O. For source client j, target client i, and module family m, let

$$\{\ell^{(i)}_{m,a}\}_{a=1}^{T^{(i)}_m} \quad \text{and} \quad \{\ell^{(j)}_{m,b}\}_{b=1}^{S^{(j)}_m}$$

denote the ordered target and source layer sequences. Define

$$\widetilde{S}_{ab} = \begin{cases} \kappa_{\ell^{(j)}_{m,b}\to\ell^{(i)}_{m,a}}, & \text{if the pair is admissible,} \\ -\infty, & \text{otherwise.} \end{cases} \tag{32}$$

Let $\Pi_{j\to i,m}$ denote the set of monotone partial matchings between the two ordered sequences. The selected alignment is

$$\widehat{\pi}_{j\to i,m} \in \arg\max_{\pi\in\Pi_{j\to i,m}} \sum_{(a,b)\in\pi} \left(\widetilde{S}_{ab} - \lambda_{\mathrm{pair}}\right). \tag{33}$$

The penalty λ_{pair} discourages the alignment from accepting low-confidence pairs merely because CKA scores are non-negative. For each target layer (i,ℓ), the resulting matched source set is

$$\mathcal{M}_i^{(\ell)} = \{(j, \pi_{j\to i}(\ell)) : \pi_{j\to i}(\ell) \text{ exists}\}. \tag{34}$$

This stage determines *which layers exchange information*. The actual source-to-target update transport is performed next in the frozen-backbone parameter geometry.

B.3 Compatible Basis Transport

For a matched source $p = (j,\pi) \in \mathcal{M}_i^{(\ell)}$, let

$$q_j = q^{(\pi)}_{j,t}, \qquad q_i = q^{(\ell)}_{i,t},$$

and define the retained source coordinate

$$C^{(\pi)}_{j,t} = U^{(\pi)\top}_{j,[q_j]} \Delta W^{(\pi)}_{j,t} V^{(\pi)}_{j,[q_j]}. \tag{35}$$

Its retained update is

$$\Delta W^{(\pi),\mathrm{ret}}_{j,t} = U^{(\pi)}_{j,[q_j]} C^{(\pi)}_{j,t} V^{(\pi)\top}_{j,[q_j]}. \tag{36}$$

Transport acts on this retained update as a whole. It does not assume a semantic one-to-one correspondence between individual source and target singular vectors. The purpose of the transport is instead to express the complete retained source update in a coordinate system associated with the matched target layer.

Same-dimensional transport. When source and target parameter matrices have compatible input and output dimensions, source–target basis overlaps are directly defined. Let

$$P^{U,(\pi,\ell)}_{j\to i,t} = U^{(\ell)\top}_{i,[q_i]} U^{(\pi)}_{j,[q_j]}, \tag{37}$$

and

$$P^{V,(\pi,\ell)}_{j\to i,t} = V^{(\ell)\top}_{i,[q_i]} V^{(\pi)}_{j,[q_j]}. \tag{38}$$

The full retained source coordinate is then rewritten in target coordinates as

$$C^{(\ell)}_{j\to i,t} = P^{U,(\pi,\ell)}_{j\to i,t} C^{(\pi)}_{j,t} P^{V,(\pi,\ell)\top}_{j\to i,t}. \tag{39}$$

Equivalently,

$$C^{(\ell)}_{j\to i,t} = U^{(\ell)\top}_{i,[q_i]} U^{(\pi)}_{j,[q_j]} C^{(\pi)}_{j,t} V^{(\pi)\top}_{j,[q_j]} V^{(\ell)}_{i,[q_i]}.$$

Thus, the entire source coordinate is transformed jointly according to the source–target basis overlap; individual singular directions are not matched or aggregated independently.

Different-dimensional transport. When the matched source and target parameter matrices have different input or output dimensions, the direct overlaps above are not defined. We therefore restrict transport to the largest retained dimension supported by both updates,

$$q^{(\pi,\ell)}_{j\to i,t} = \min\left(q^{(\pi)}_{j,t}, q^{(\ell)}_{i,t}\right). \tag{40}$$

For compactness, write

$$q = q^{(\pi,\ell)}_{j\to i,t}.$$

The mutually retained source coordinate is

$$C^{(\pi,[q])}_{j,t} = U^{(\pi)\top}_{j,[q]} \Delta W^{(\pi)}_{j,t} V^{(\pi)}_{j,[q]}. \tag{41}$$

We construct output- and input-side linear transforms by

$$\Theta^{U}_{j\to i,t} \in \arg\min_{\Theta} \left\| \Theta U^{(\pi)}_{j,[q]} - U^{(\ell)}_{i,[q]} \right\|^2_F, \tag{42}$$

$$\Theta^{V}_{j\to i,t} \in \arg\min_{\Theta} \left\| \Theta V^{(\pi)}_{j,[q]} - V^{(\ell)}_{i,[q]} \right\|^2_F. \tag{43}$$

For orthonormal retained bases, the minimum-Frobenius-norm solutions are

$$\Theta^{U}_{j\to i,t} = U^{(\ell)}_{i,[q]} U^{(\pi)\top}_{j,[q]}, \qquad \Theta^{V}_{j\to i,t} = V^{(\ell)}_{i,[q]} V^{(\pi)\top}_{j,[q]}. \tag{44}$$

The retained update is transported jointly as

$$\widetilde{\Delta W}^{(\ell)}_{j\to i,t} = \Theta^{U}_{j\to i,t} \left(U^{(\pi)}_{j,[q]} C^{(\pi,[q])}_{j,t} V^{(\pi)\top}_{j,[q]} \right) \Theta^{V\top}_{j\to i,t}, \tag{45}$$

and represented in the target retained coordinates as

$$C^{(\ell,[q])}_{j\to i,t} = U^{(\ell)\top}_{i,[q]} \widetilde{\Delta W}^{(\ell)}_{j\to i,t} V^{(\ell)}_{i,[q]}. \tag{46}$$

Equations equation 45–equation 46 emphasise that transport is applied to the retained update as a whole rather than through independent sign or semantic decisions for individual singular directions. Coordinate-wise compatibility is evaluated only after the source update has been expressed in the common target geometry.

In implementation, the full ambient transport matrices need not be materialised; the operations above are evaluated through the retained bases and low-dimensional coordinates.

B.4 TARGET-SPACE MASK AND WEIGHT AGGREGATION

After transport, source updates participating at a given coordinate are expressed relative to the same target retained geometry. The remaining incompatibility is therefore numerical rather than structural: different transported sources can disagree in sign or magnitude.

For target layer (i, ℓ), its local retained coordinate is

$$C_{i,t}^{0,(\ell)} = U_{i,[q_i]}^{(\ell)\top} \Delta W_{i,t}^{(\ell)} V_{i,[q_i]}^{(\ell)}. \tag{47}$$

For coordinate $[a, b]$, define the geometrically supported donor set

$$\mathcal{E}_{i,t}^{(\ell)}[a,b] = \left\{ p \in \mathcal{M}_i^{(\ell)} : (a,b) \in \Omega_{p\to i,t}^{(\ell)} \right\}. \tag{48}$$

Masking before weighting. For each supported donor $p \in \mathcal{E}_{i,t}^{(\ell)}[a,b]$, we define the coordinate mask

$$\Gamma_{p,t}^{(\ell)}[a,b] = \begin{cases} 1, & C_{p\to i,t}^{(\ell)}[a,b]\, C_{i,t}^{0,(\ell)}[a,b] \geq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{49}$$

The surviving donor set is therefore

$$\mathcal{S}_{i,t}^{(\ell)}[a,b] = \left\{ p \in \mathcal{E}_{i,t}^{(\ell)}[a,b] : \Gamma_{p,t}^{(\ell)}[a,b] = 1 \right\}. \tag{50}$$

A masked pair does not participate in either weight normalisation or aggregation.

The mask is applied only after compatible transport. Consequently, opposite signs at this stage refer to disagreement between updates represented in the same target coordinate system, rather than to untransported source coordinates.

Reliability weighting. After coordinate-level conflicts have been removed, the surviving pairs are weighted according to the reliability of their source–target correspondence. For a matched source pair $p = (j, \pi)$, let $a_{i,p,t}^{\mathrm{sel},(\ell)} \geq 0$ denote the reliability assigned during pair selection, and let $\widetilde{\kappa}_{i,p,t}^{(\ell)} \in [0,1]$ denote the normalised CCA-adjusted CKA compatibility obtained from functional matching. We define the source reliability as

$$\rho_{i,p,t}^{(\ell)} = a_{i,p,t}^{\mathrm{sel},(\ell)} \widetilde{\kappa}_{i,p,t}^{(\ell)}. \tag{51}$$

For target coordinate $[a, b]$, let $\mathcal{S}_{i,t}^{(\ell)}[a,b]$ denote the pairs that both support this coordinate and survive the coordinate mask. The normalised aggregation weight is

$$w_{i,p,t}^{(\ell)}[a,b] = \frac{\rho_{i,p,t}^{(\ell)}}{\displaystyle\sum_{p' \in \mathcal{S}_{i,t}^{(\ell)}[a,b]} \rho_{i,p',t}^{(\ell)}}, \qquad p \in \mathcal{S}_{i,t}^{(\ell)}[a,b], \tag{52}$$

whenever $\mathcal{S}_{i,t}^{(\ell)}[a,b] \neq \emptyset$.

Equivalently, using the binary coordinate mask explicitly,

$$w_{i,p,t}^{(\ell)}[a,b] = \frac{\Gamma_{p,t}^{(\ell)}[a,b]\, \rho_{i,p,t}^{(\ell)}}{\displaystyle\sum_{p' \in \mathcal{E}_{i,t}^{(\ell)}[a,b]} \Gamma_{p',t}^{(\ell)}[a,b]\, \rho_{i,p',t}^{(\ell)}}. \tag{53}$$

Thus, masking is applied before weighting: a conflicting source receives zero weight and is excluded from both normalisation and aggregation. Among the surviving pairs, functional matching reliability determines the relative contribution to the target coordinate.

Masked weighted aggregation. For each target coordinate, the surviving source contributions are aggregated as

$$C_{i,t}^{\mathrm{agg},(\ell)}[a,b] = \begin{cases} \displaystyle\sum_{p \in \mathcal{S}_{i,t}^{(\ell)}[a,b]} w_{i,p,t}^{(\ell)}[a,b]\, C_{p\to i,t}^{(\ell)}[a,b], & \mathcal{S}_{i,t}^{(\ell)}[a,b] \neq \emptyset, \\ C_{i,t}^{0,(\ell)}[a,b], & \mathcal{S}_{i,t}^{(\ell)}[a,b] = \emptyset. \end{cases} \tag{54}$$

C Convergence Analysis

This appendix establishes target-side convergence of V-HFLoRA. The analysis separates two sources of discrepancy. The first is the intrinsic discrepancy between an ideal population-level V-HFLoRA collaborative direction and the exact target gradient. The second is the additional approximation error introduced when the ideal pipeline is replaced by finite-sample functional matching, compatible transport, mask–reliability aggregation, and rank-constrained recovery.

We first formulate the optimization problem in a fixed target coordinate space and derive a general convergence result for an inexact target-space update. We then show that the approximation error appearing in this result can be controlled by the individual components of V-HFLoRA.

Throughout the appendix, we fix a target client i and suppress the target index whenever no ambiguity arises. Let $\mathcal{F}_t$ denote the history available at the beginning of communication round t, and write

$$\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot \mid \mathcal{F}_t].$$

C.1 Target-Space Formulation

Consider a frozen target layer

$$W_0 \in \mathbb{R}^{d_o \times d_i},$$

with full singular value decomposition

$$W_0 = U\Sigma V^\top,$$

where

$$U^\top U = UU^\top = I_{d_o}, \qquad V^\top V = VV^\top = I_{d_i}.$$

For any LoRA-induced update ΔW, define its coordinate representation in the frozen target basis by

$$X = U^\top \Delta W V. \tag{55}$$

Because U and V are orthogonal,

$$\Delta W = UXV^\top.$$

The coordinate transformation preserves the Frobenius geometry. Indeed,

$$\begin{aligned}
\|UXV^\top\|_F^2 &= \operatorname{tr}\left[(UXV^\top)^\top (UXV^\top)\right] && (56)\\
&= \operatorname{tr}\left[VX^\top U^\top UXV^\top\right] && (57)\\
&= \operatorname{tr}\left[VX^\top XV^\top\right] && (58)\\
&= \operatorname{tr}\left[X^\top XV^\top V\right] && (59)\\
&= \operatorname{tr}\left[X^\top X\right] && (60)\\
&= \|X\|_F^2. && (61)
\end{aligned}$$

Hence

$$\|UXV^\top\|_F = \|X\|_F. \tag{62}$$

Let the target client's objective be

$$f(W, v),$$

where v collects target-local parameters such as the prediction head. The equivalent objective in the frozen target coordinates is

$$F(X, v) = f(W_0 + UXV^\top, v). \tag{63}$$

The corresponding coordinate gradient follows directly from the chain rule. For an arbitrary perturbation dX,

$$dW = U(dX)V^\top.$$

Therefore

$$dF = \langle \nabla_W f, dW \rangle_F \tag{64}$$
$$= \left\langle \nabla_W f, U(dX)V^\top \right\rangle_F \tag{65}$$
$$= \mathrm{tr}\left[(\nabla_W f)^\top U(dX)V^\top \right] \tag{66}$$
$$= \mathrm{tr}\left[V^\top (\nabla_W f)^\top U(dX) \right] \tag{67}$$
$$= \mathrm{tr}\left[\left(U^\top \nabla_W f V \right)^\top dX \right] \tag{68}$$
$$= \left\langle U^\top \nabla_W f V, dX \right\rangle_F . \tag{69}$$

Since, by definition,

$$dF = \langle \nabla_X F, dX \rangle_F$$

for every dX, we obtain

$$\nabla_X F = U^\top \nabla_W f V. \tag{70}$$

Thus the frozen-backbone representation is an orthogonal reparameterization of the target update space rather than an auxiliary optimization space.

Across all adapted target layers, we vectorize the corresponding coordinate matrices and denote the complete adapter variable by x. Together with the target-local parameters v, define

$$y = (x, v).$$

We henceforth write the target objective simply as $F(y)$.

At round t, the update-conditioned retention rule selects a retained dimension q_t. Importantly, q_t does not redefine the dimension of x. Instead, it determines a coordinate projector

$$P_t = P(q_t)$$

that specifies which target coordinates are eligible for federated sharing at that round. Thus $P_t x_t$ is shared, whereas $(I - P_t)x_t$ remains local. Changing q_t to q_{t+1} changes the sharing support but not the current parameter vector or the objective F. Consequently, standard telescoping of the objective values remains valid.

We now introduce the two update directions used throughout the analysis. Let

$$g_t^\circ$$

denote the *ideal target-space collaborative direction*: the direction obtained using population-level functional matching, population-level compatible transport, and ideal mask–reliability aggregation, before the finite-rank recovery imposed by the implementation. Let

$$\widehat{g}_t$$

denote the actual V-HFLoRA direction.

Their difference is

$$e_t = \widehat{g}_t - g_t^\circ. \tag{71}$$

The remainder of the appendix will first establish convergence in terms of $\mathbb{E}_t\|e_t\|^2$, and will then derive an explicit V-HFLoRA-specific bound for this quantity.

C.2 A General Convergence Result for Inexact Target-Space Updates

A complete federated round may contain several local optimization steps. We express the resulting target-side displacement in the equivalent form

$$y_{t+1} = y_t - \eta \widehat{g}_t. \tag{72}$$

Given y_t and y_{t+1}, this simply defines the effective round direction as

$$\widehat{g}_t = -\frac{y_{t+1} - y_t}{\eta}.$$

Let

$$G_t = \nabla F(y_t).$$

The ideal collaborative direction g_t° need not be an unbiased estimate of G_t. We therefore define

$$m_t = \mathbb{E}_t[g_t^\circ] \tag{73}$$

and explicitly retain its target-side bias,

$$b_t = m_t - G_t, \qquad B_t = \|b_t\|^2. \tag{74}$$

The term B_t captures population-level disagreement between the ideal V-HFLoRA collaborative direction and the exact target gradient. In particular, it may include persistent client heterogeneity and structural approximation induced by restricting information exchange to target-compatible subspaces.

We require only the standard smoothness and bounded-variance conditions at this stage.

Smoothness. Assume that F is L-smooth:

$$F(y') \le F(y) + \langle \nabla F(y), y' - y \rangle + \frac{L}{2}\|y' - y\|^2, \tag{75}$$

and lower bounded by

$$F(y) \ge F_\star > -\infty.$$

Variance of the ideal collaborative direction. Assume

$$\mathbb{E}_t\left[\|g_t^\circ - m_t\|^2\right] \le \sigma^2. \tag{76}$$

Finally, define the actual pipeline approximation error by

$$\mathcal{E}_t = \mathbb{E}_t\left[\|\widehat{g}_t - g_t^\circ\|^2\right]. \tag{77}$$

No particular form of $\mathcal{E}_t$ is assumed in this subsection.

Theorem C.1 (Convergence under an inexact target-space update)**.** *Suppose that equation 75 and equation 76 hold, and let*

$$0 < \eta < \frac{1}{4L}.$$

Define

$$\Gamma = \frac{1}{2} - 2L\eta > 0.$$

Then

$$\begin{aligned}\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\|\nabla F(y_t)\|^2 \le{}& \frac{F(y_0) - F_\star}{\eta\Gamma T} \\ &+ \frac{1 + 2L\eta}{\Gamma}\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}B_t \\ &+ \frac{1 + L\eta}{\Gamma}\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\mathcal{E}_t \\ &+ \frac{L\eta}{\Gamma}\sigma^2.\end{aligned} \tag{78}$$

Proof. We first derive a lower bound for the expected descent inner product.

Define the zero-mean stochastic component of the ideal direction as

$$\xi_t = g_t^\circ - m_t.$$

By equation 73,

$$\mathbb{E}_t[\xi_t] = \mathbb{E}_t[g_t^\circ] - m_t = m_t - m_t = 0.$$

Using equation 74,

$$m_t = G_t + b_t,$$

and therefore

$$g_t^\circ = G_t + b_t + \xi_t.$$

Using equation 71,

$$\widehat{g}_t = g_t^\circ + e_t = G_t + b_t + \xi_t + e_t.$$

Hence

$$\mathbb{E}_t\langle G_t, \widehat{g}_t\rangle = \mathbb{E}_t\left\langle G_t, G_t + b_t + \xi_t + e_t\right\rangle \tag{79}$$

$$= \|G_t\|^2 + \langle G_t, b_t\rangle + \mathbb{E}_t\langle G_t, \xi_t\rangle + \mathbb{E}_t\langle G_t, e_t\rangle. \tag{80}$$

Because G_t is $\mathcal{F}_t$-measurable,

$$\mathbb{E}_t\langle G_t, \xi_t\rangle = \langle G_t, \mathbb{E}_t[\xi_t]\rangle \tag{81}$$

$$= 0. \tag{82}$$

For the bias term, Cauchy–Schwarz gives

$$\langle G_t, b_t\rangle \geq -\|G_t\|\|b_t\|.$$

Using

$$ab \leq \frac{1}{4}a^2 + b^2,$$

which follows from

$$\left(\frac{a}{2} - b\right)^2 \geq 0,$$

with

$$a = \|G_t\|, \qquad b = \|b_t\|,$$

we obtain

$$\langle G_t, b_t\rangle \geq -\frac{1}{4}\|G_t\|^2 - \|b_t\|^2 \tag{83}$$

$$= -\frac{1}{4}\|G_t\|^2 - B_t. \tag{84}$$

For the algorithmic error,

$$\mathbb{E}_t\langle G_t, e_t\rangle \geq -\mathbb{E}_t\left[|\langle G_t, e_t\rangle|\right] \tag{85}$$

$$\geq -\|G_t\|\mathbb{E}_t\|e_t\|. \tag{86}$$

By Jensen's inequality,

$$\mathbb{E}_t\|e_t\| \leq \sqrt{\mathbb{E}_t\|e_t\|^2} = \sqrt{\mathcal{E}_t}.$$

Hence

$$\mathbb{E}_t\langle G_t, e_t\rangle \geq -\|G_t\|\sqrt{\mathcal{E}_t}. \tag{87}$$

Applying again

$$ab \leq \frac{1}{4}a^2 + b^2$$

with

$$a = \|G_t\|, \qquad b = \sqrt{\mathcal{E}_t},$$

gives

$$\mathbb{E}_t\langle G_t, e_t\rangle \geq -\frac{1}{4}\|G_t\|^2 - \mathcal{E}_t. \tag{88}$$

Substituting equation 82, equation 84, and equation 88 into equation 80,

$$\mathbb{E}_t\langle G_t, \widehat{g}_t\rangle \geq \|G_t\|^2 - \frac{1}{4}\|G_t\|^2 - B_t - \frac{1}{4}\|G_t\|^2 - \mathcal{E}_t \tag{89}$$

$$= \frac{1}{2}\|G_t\|^2 - B_t - \mathcal{E}_t. \tag{90}$$

We next bound the second moment of the actual direction. Since

$$\widehat{g}_t = g_t^\circ + e_t,$$

the inequality

$$\|a+b\|^2 \leq 2\|a\|^2 + 2\|b\|^2$$

gives

$$\mathbb{E}_t\|\widehat{g}_t\|^2 \leq 2\mathbb{E}_t\|g_t^\circ\|^2 + 2\mathbb{E}_t\|e_t\|^2 \tag{91}$$

$$= 2\mathbb{E}_t\|g_t^\circ\|^2 + 2\mathcal{E}_t. \tag{92}$$

Since

$$g_t^\circ = m_t + \xi_t,$$

we have

$$\mathbb{E}_t\|g_t^\circ\|^2 = \mathbb{E}_t\|m_t + \xi_t\|^2 \tag{93}$$

$$= \|m_t\|^2 + 2\langle m_t, \mathbb{E}_t[\xi_t]\rangle + \mathbb{E}_t\|\xi_t\|^2 \tag{94}$$

$$= \|m_t\|^2 + \mathbb{E}_t\|\xi_t\|^2 \tag{95}$$

$$\leq \|m_t\|^2 + \sigma^2. \tag{96}$$

Using

$$m_t = G_t + b_t,$$

we further have

$$\|m_t\|^2 = \|G_t + b_t\|^2 \tag{97}$$

$$\leq 2\|G_t\|^2 + 2\|b_t\|^2 \tag{98}$$

$$= 2\|G_t\|^2 + 2B_t. \tag{99}$$

Substituting equation 99 into equation 96,

$$\mathbb{E}_t\|g_t^\circ\|^2 \leq 2\|G_t\|^2 + 2B_t + \sigma^2.$$

Substituting this result into equation 92,

$$\mathbb{E}_t\|\widehat{g}_t\|^2 \leq 4\|G_t\|^2 + 4B_t + 2\sigma^2 + 2\mathcal{E}_t. \tag{100}$$

We now apply smoothness. From equation 72,

$$y_{t+1} - y_t = -\eta\widehat{g}_t.$$

Using equation 75,

$$F(y_{t+1}) \leq F(y_t) - \eta\langle G_t, \widehat{g}_t\rangle + \frac{L\eta^2}{2}\|\widehat{g}_t\|^2. \tag{101}$$

Taking conditional expectations,

$$\mathbb{E}_t F(y_{t+1}) \leq F(y_t) - \eta\mathbb{E}_t\langle G_t, \widehat{g}_t\rangle + \frac{L\eta^2}{2}\mathbb{E}_t\|\widehat{g}_t\|^2. \tag{102}$$

Using equation 90,

$$-\eta\mathbb{E}_t\langle G_t, \widehat{g}_t\rangle \leq -\frac{\eta}{2}\|G_t\|^2 + \eta B_t + \eta\mathcal{E}_t. \tag{103}$$

Using equation 100,

$$\begin{aligned}\frac{L\eta^2}{2}\mathbb{E}_t\|\widehat{g}_t\|^2 \leq\ & 2L\eta^2\|G_t\|^2 + 2L\eta^2 B_t \\ & + L\eta^2\sigma^2 + L\eta^2\mathcal{E}_t.\end{aligned} \tag{104}$$

Substituting equation 103 and equation 104 into equation 102,

$$\begin{aligned}\mathbb{E}_t F(y_{t+1}) \leq\ & F(y_t) - \left(\frac{\eta}{2} - 2L\eta^2\right)\|G_t\|^2 \\ & + \left(\eta + 2L\eta^2\right) B_t \\ & + \left(\eta + L\eta^2\right)\mathcal{E}_t + L\eta^2\sigma^2.\end{aligned} \tag{105}$$

Factoring out η,

$$\begin{aligned}\mathbb{E}_t F(y_{t+1}) \leq\ & F(y_t) - \eta\left(\frac{1}{2} - 2L\eta\right)\|G_t\|^2 \\ & + \eta(1+2L\eta)B_t + \eta(1+L\eta)\mathcal{E}_t \\ & + L\eta^2\sigma^2.\end{aligned} \tag{106}$$

With

$$\Gamma = \frac{1}{2} - 2L\eta,$$

the assumed step-size condition

$$\eta < \frac{1}{4L}$$

ensures $\Gamma > 0$.

Rearranging equation 106,

$$\begin{aligned}\eta\Gamma\|G_t\|^2 \leq\ & F(y_t) - \mathbb{E}_t F(y_{t+1}) \\ & + \eta(1+2L\eta)B_t + \eta(1+L\eta)\mathcal{E}_t \\ & + L\eta^2\sigma^2.\end{aligned} \tag{107}$$

Taking total expectations and using the tower property,

$$\begin{aligned}\eta\Gamma\mathbb{E}\|G_t\|^2 \leq\ & \mathbb{E}F(y_t) - \mathbb{E}F(y_{t+1}) \\ & + \eta(1+2L\eta)\mathbb{E}B_t \\ & + \eta(1+L\eta)\mathbb{E}\mathcal{E}_t + L\eta^2\sigma^2.\end{aligned} \tag{108}$$

Summing from $t = 0$ to $T-1$ gives

$$\begin{aligned}\eta\Gamma\sum_{t=0}^{T-1}\mathbb{E}\|G_t\|^2 \leq\ & \sum_{t=0}^{T-1}\left[\mathbb{E}F(y_t) - \mathbb{E}F(y_{t+1})\right] \\ & + \eta(1+2L\eta)\sum_{t=0}^{T-1}\mathbb{E}B_t \\ & + \eta(1+L\eta)\sum_{t=0}^{T-1}\mathbb{E}\mathcal{E}_t \\ & + TL\eta^2\sigma^2.\end{aligned} \tag{109}$$

The loss differences telescope:

$$\sum_{t=0}^{T-1} [F(y_t) - F(y_{t+1})] = F(y_0) - F(y_T). \tag{110}$$

Since

$$F(y_T) \geq F_\star,$$

we have

$$F(y_0) - \mathbb{E}F(y_T) \leq F(y_0) - F_\star.$$

Therefore

$$\begin{aligned} \eta\Gamma \sum_{t=0}^{T-1} \mathbb{E}\|G_t\|^2 \leq F(y_0) - F_\star \\ + \eta(1 + 2L\eta) \sum_{t=0}^{T-1} \mathbb{E}B_t \\ + \eta(1 + L\eta) \sum_{t=0}^{T-1} \mathbb{E}\mathcal{E}_t \\ + TL\eta^2\sigma^2. \end{aligned} \tag{111}$$

Dividing by $\eta\Gamma T$ and using $G_t = \nabla F(y_t)$ yields equation 78. □

Theorem C.1 reduces the remaining analysis to bounding $\mathcal{E}_t$. We now show how the approximation error decomposes across the components of V-HFLoRA.

C.3 Bounding the V-HFLoRA Approximation Error

C.3.1 Functional-Matching Stability

Functional matching is used only to identify representationally compatible source layers. We do not assume that a large CKA score implies equality of source and target gradients.

For target layer ℓ and candidate source layer π, let

$$\kappa_{\ell,\pi} = \mathrm{CKA}\left(Z_j^{(\pi)} M^\star_{\pi\to\ell}, Z_i^{(\ell)}\right)$$

denote the population CCA–ridge–CKA score, and let

$$\widehat{\kappa}_{\ell,\pi}$$

denote its empirical counterpart computed from the calibration set.

Centered alignment admits finite-sample concentration for fixed bounded representations (Cortes et al., 2012). Because our PCA, CCA, and ridge maps may themselves be estimated from the same calibration sample, we use the following uniform score-accuracy event rather than directly identifying the entire pipeline with the fixed-kernel concentration theorem:

$$\mathcal{A}_t = \left\{ \max_{\ell,\pi} |\widehat{\kappa}_{\ell,\pi} - \kappa_{\ell,\pi}| \leq \varepsilon_{\mathrm{match},t} \right\}. \tag{112}$$

Assume

$$\Pr_t(\mathcal{A}_t) \geq 1 - \delta_t^{\mathrm{score}}.$$

Let

$$J(\pi) = \sum_{\ell=1}^{L} \omega_\ell \kappa_{\ell,\pi(\ell)} - \lambda_{\mathrm{pair}} \mathcal{P}(\pi), \tag{113}$$

where

$$\omega_\ell \geq 0, \qquad \sum_{\ell=1}^{L} \omega_\ell = 1,$$

and $\mathcal{P}(\pi)$ denotes the same layer-order restriction or penalty used by the implementation. Define

$$\pi^\star = \arg\max_\pi J(\pi).$$

Similarly,

$$\widehat{J}(\pi) = \sum_{\ell=1}^{L} \omega_\ell \widehat{\kappa}_{\ell,\pi(\ell)} - \lambda_{\text{pair}} \mathcal{P}(\pi), \tag{114}$$

and

$$\widehat{\pi} = \arg\max_\pi \widehat{J}(\pi).$$

Lemma C.2 (Stability of functional matching)**.** *On $\mathcal{A}_t$,*

$$J(\pi^\star) - J(\widehat{\pi}) \leq 2\varepsilon_{\text{match},t}.$$

Moreover, if

$$\Gamma_t^{\text{match}} = J(\pi^\star) - \max_{\pi \neq \pi^\star} J(\pi) > 2\varepsilon_{\text{match},t}, \tag{115}$$

then

$$\widehat{\pi} = \pi^\star.$$

Proof. We first show that the empirical and population objectives are uniformly close.

Subtracting equation 113 from equation 114,

$$\widehat{J}(\pi) - J(\pi) = \sum_{\ell=1}^{L} \omega_\ell \left[\widehat{\kappa}_{\ell,\pi(\ell)} - \kappa_{\ell,\pi(\ell)}\right], \tag{116}$$

because the order-penalty terms cancel exactly.

Hence, on $\mathcal{A}_t$,

$$|\widehat{J}(\pi) - J(\pi)| \leq \sum_{\ell=1}^{L} \omega_\ell \left|\widehat{\kappa}_{\ell,\pi(\ell)} - \kappa_{\ell,\pi(\ell)}\right| \tag{117}$$

$$\leq \sum_{\ell=1}^{L} \omega_\ell \varepsilon_{\text{match},t} \tag{118}$$

$$= \varepsilon_{\text{match},t} \sum_{\ell=1}^{L} \omega_\ell \tag{119}$$

$$= \varepsilon_{\text{match},t}. \tag{120}$$

We now bound the population regret of the empirical optimizer. Add and subtract the two empirical objective values:

$$\begin{aligned} J(\pi^\star) - J(\widehat{\pi}) = J(\pi^\star) - \widehat{J}(\pi^\star) \\ + \widehat{J}(\pi^\star) - \widehat{J}(\widehat{\pi}) \\ + \widehat{J}(\widehat{\pi}) - J(\widehat{\pi}). \end{aligned} \tag{121}$$

Since $\widehat{\pi}$ maximizes $\widehat{J}$,

$$\widehat{J}(\pi^\star) - \widehat{J}(\widehat{\pi}) \leq 0.$$

Therefore

$$\begin{aligned} J(\pi^\star) - J(\widehat{\pi}) \leq |J(\pi^\star) - \widehat{J}(\pi^\star)| \\ + |\widehat{J}(\widehat{\pi}) - J(\widehat{\pi})|. \end{aligned} \tag{122}$$

Applying equation 120 to both terms,

$$J(\pi^\star) - J(\widehat{\pi}) \le 2\varepsilon_{\mathrm{match},t}.$$

For the exact-recovery statement, take arbitrary $\pi \neq \pi^\star$. From equation 120,

$$\widehat{J}(\pi^\star) \ge J(\pi^\star) - \varepsilon_{\mathrm{match},t},$$

while

$$\widehat{J}(\pi) \le J(\pi) + \varepsilon_{\mathrm{match},t}.$$

Subtracting,

$$\widehat{J}(\pi^\star) - \widehat{J}(\pi) \ge J(\pi^\star) - J(\pi) - 2\varepsilon_{\mathrm{match},t} \tag{123}$$

$$\ge \Gamma_t^{\mathrm{match}} - 2\varepsilon_{\mathrm{match},t}. \tag{124}$$

Under equation 115, the right-hand side is strictly positive. Thus

$$\widehat{J}(\pi^\star) > \widehat{J}(\pi)$$

for every $\pi \neq \pi^\star$, proving

$$\widehat{\pi} = \pi^\star.$$

□

After layer correspondence, the implementation retains at most k_{pair} pairs. Let s_p and $\widehat{s}_p$ denote the population and empirical ranking scores actually used by the implementation, and let

$$s_{(1)} \ge s_{(2)} \ge \cdots$$

denote the population ordering.

The same margin argument gives

$$s_{(k)} - s_{(k+1)} > 2\varepsilon_s \quad \Longrightarrow \quad \widehat{\mathcal{M}}_t = \mathcal{M}_t^\star,$$

whenever

$$\max_p |\widehat{s}_p - s_p| \le \varepsilon_s.$$

Indeed, for every population top-k pair p,

$$\widehat{s}_p \ge s_{(k)} - \varepsilon_s,$$

whereas for every pair q outside the population top-k set,

$$\widehat{s}_q \le s_{(k+1)} + \varepsilon_s.$$

The assumed margin implies

$$s_{(k)} - \varepsilon_s > s_{(k+1)} + \varepsilon_s,$$

and therefore

$$\widehat{s}_p > \widehat{s}_q.$$

We collect layer-matching and pair-pruning failures into the event

$$\mathcal{G}_t = \{\widehat{\mathcal{M}}_t = \mathcal{M}_t^\star\},$$

and denote

$$\Pr_t(\mathcal{G}_t^c) \le \delta_t^{\mathrm{match}}. \tag{125}$$

C.3.2 Compatible Basis Transport

Condition on correct pair selection. We next control the discrepancy introduced when a source update is represented in the target geometry.

Same-dimensional layers. Let

$$S_j = U_j C_j V_j^\top$$

be a retained source update. The target retained row and column bases are U_i and V_i.

The target representation used by V-HFLoRA is the solution of

$$C^\star_{j\to i} = \arg\min_C \|S_j - U_i C V_i^\top\|_F^2. \tag{126}$$

Expanding the objective,

$$\begin{aligned}\|S_j - U_i C V_i^\top\|_F^2 = \|S_j\|_F^2 + \|U_i C V_i^\top\|_F^2 \\ - 2\langle S_j, U_i C V_i^\top\rangle_F.\end{aligned} \tag{127}$$

Because U_i and V_i have orthonormal columns,

$$\|U_i C V_i^\top\|_F^2 = \operatorname{tr}\left[V_i C^\top U_i^\top U_i C V_i^\top\right] \tag{128}$$

$$= \operatorname{tr}\left[V_i C^\top C V_i^\top\right] \tag{129}$$

$$= \operatorname{tr}\left[C^\top C V_i^\top V_i\right] \tag{130}$$

$$= \|C\|_F^2. \tag{131}$$

Similarly,

$$\langle S_j, U_i C V_i^\top\rangle_F = \operatorname{tr}\left[S_j^\top U_i C V_i^\top\right] \tag{132}$$

$$= \operatorname{tr}\left[V_i^\top S_j^\top U_i C\right] \tag{133}$$

$$= \left\langle U_i^\top S_j V_i, C\right\rangle_F. \tag{134}$$

Define

$$D = U_i^\top S_j V_i.$$

Then

$$\|S_j - U_i C V_i^\top\|_F^2 = \|S_j\|_F^2 + \|C\|_F^2 - 2\langle D, C\rangle_F \tag{135}$$

$$= \|C - D\|_F^2 + \|S_j\|_F^2 - \|D\|_F^2. \tag{136}$$

The last two terms do not depend on C, so the minimizer is

$$C = D.$$

Therefore

$$C^\star_{j\to i} = U_i^\top S_j V_i = U_i^\top U_j C_j V_j^\top V_i. \tag{137}$$

Moreover,

$$\|C^\star_{j\to i}\|_F \le \|U_i^\top U_j\|_2 \|C_j\|_F \|V_j^\top V_i\|_2 \tag{138}$$

$$\le \|C_j\|_F, \tag{139}$$

because both basis-overlap matrices have operator norm at most one. Thus same-dimensional transport does not amplify the source-update norm.

The corresponding target reconstruction is

$$U_i C^\star_{j\to i} V_i^\top = P_i^U S_j P_i^V,$$

where

$$P_i^U = U_i U_i^\top, \qquad P_i^V = V_i V_i^\top.$$

The residual

$$S_j - P_i^U S_j P_i^V$$

is the component of the source retained update that cannot be represented in the target retained row and column subspaces. This is a structural projection residual, not an estimation error of the transport map, and is therefore absorbed later into the ideal-direction bias B_t.

Cross-dimensional layers. When source and target ambient dimensions differ, the overlaps $U_i^\top U_j$ and $V_j^\top V_i$ are not defined. Let

$$q = \min(q_i, q_j),$$

and consider the mutual-q source coordinate

$$C_j^{[q]} = U_{j,[q]}^\top \Delta W_j V_{j,[q]}.$$

Equalizing the coordinate dimension does not itself register the source and target latent axes. We therefore use calibration-induced orthogonal registrations

$$R_U^\star, \qquad R_V^\star \in \mathbb{R}^{q\times q}.$$

They are obtained from the standard orthogonal Procrustes problems

$$R_U^\star = \arg\min_{R^\top R=I} \|Z_j^U R - Z_i^U\|_F^2$$

and

$$R_V^\star = \arg\min_{R^\top R=I} \|Z_j^V R - Z_i^V\|_F^2.$$

The closed-form SVD solution is standard (Schönemann, 1966).

The ideal registered update is

$$C_{j\to i}^{\star,[q]} = R_U^{\star\top} C_j^{[q]} R_V^\star. \tag{140}$$

Let finite calibration data yield

$$\widehat{R}_U, \qquad \widehat{R}_V,$$

and define

$$\varepsilon_U^{\mathrm{reg}} = \|\widehat{R}_U - R_U^\star\|_2, \qquad \varepsilon_V^{\mathrm{reg}} = \|\widehat{R}_V - R_V^\star\|_2.$$

The implemented transport is

$$\widehat{C}_{j\to i}^{[q]} = \widehat{R}_U^\top C_j^{[q]} \widehat{R}_V.$$

Lemma C.3 (Cross-dimensional transport stability)**.** *The registration error satisfies*

$$\|\widehat{C}_{j\to i}^{[q]} - C_{j\to i}^{\star,[q]}\|_F \le (\varepsilon_U^{\mathrm{reg}} + \varepsilon_V^{\mathrm{reg}})\, \|C_j^{[q]}\|_F. \tag{141}$$

Proof. Starting from the two transported coordinates,

$$\begin{aligned}
&\widehat{C}_{j\to i}^{[q]} - C_{j\to i}^{\star,[q]} \\
&= \widehat{R}_U^\top C_j^{[q]} \widehat{R}_V - R_U^{\star\top} C_j^{[q]} R_V^\star.
\end{aligned} \tag{142}$$

Add and subtract

$$R_U^{\star\top} C_j^{[q]} \widehat{R}_V :$$

$$\begin{aligned}
&\widehat{C}_{j\to i}^{[q]} - C_{j\to i}^{\star,[q]} \\
&= \widehat{R}_U^\top C_j^{[q]} \widehat{R}_V - R_U^{\star\top} C_j^{[q]} \widehat{R}_V \\
&\quad + R_U^{\star\top} C_j^{[q]} \widehat{R}_V - R_U^{\star\top} C_j^{[q]} R_V^\star
\end{aligned} \tag{143}$$

$$\begin{aligned}
&= (\widehat{R}_U - R_U^\star)^\top C_j^{[q]} \widehat{R}_V \\
&\quad + R_U^{\star\top} C_j^{[q]} (\widehat{R}_V - R_V^\star).
\end{aligned} \tag{144}$$

Using the triangle inequality,

$$\begin{aligned}
&\|\widehat{C}_{j\to i}^{[q]} - C_{j\to i}^{\star,[q]}\|_F \\
&\le \|(\widehat{R}_U - R_U^\star)^\top C_j^{[q]} \widehat{R}_V\|_F \\
&\quad + \|R_U^{\star\top} C_j^{[q]} (\widehat{R}_V - R_V^\star)\|_F.
\end{aligned} \tag{145}$$

For the first term,

$$\|(\widehat{R}_U - R_U^\star)^\top C_j^{[q]} \widehat{R}_V\|_F \\ \le \|\widehat{R}_U - R_U^\star\|_2 \|C_j^{[q]}\|_F \|\widehat{R}_V\|_2. \tag{146}$$

Since $\widehat{R}_V$ is orthogonal,

$$\|\widehat{R}_V\|_2 = 1,$$

so

$$\|(\widehat{R}_U - R_U^\star)^\top C_j^{[q]} \widehat{R}_V\|_F \le \varepsilon_U^{\mathrm{reg}} \|C_j^{[q]}\|_F.$$

For the second term,

$$\|R_U^{\star\top} C_j^{[q]} (\widehat{R}_V - R_V^\star)\|_F \\ \le \|R_U^\star\|_2 \|C_j^{[q]}\|_F \|\widehat{R}_V - R_V^\star\|_2. \tag{147}$$

Since $R_U^\star$ is orthogonal,

$$\|R_U^\star\|_2 = 1,$$

and therefore

$$\|R_U^{\star\top} C_j^{[q]} (\widehat{R}_V - R_V^\star)\|_F \le \varepsilon_V^{\mathrm{reg}} \|C_j^{[q]}\|_F.$$

Adding the two terms proves equation 141. □

For either dimensional regime, let

$$z_{p,t}^\star$$

denote the ideally transported pair value expressed in target coordinates, and let

$$\widehat{z}_{p,t} = z_{p,t}^\star + e_{p,t}^{\mathrm{tr}}$$

denote the actual transported value. We use the resulting $e_{p,t}^{\mathrm{tr}}$ in the aggregation analysis below.

C.3.3 Mask–Reliability Aggregation

Fix a target coordinate k. Extend all candidate weights to a common pairs index set by assigning zero weight to unsupported or masked candidates and unit local weight when the fallback rule is activated.

Let

$$w_{p,t}^\star[k]$$

denote the ideal extended weight and

$$\widehat{w}_{p,t}[k]$$

the actual extended weight. Both satisfy

$$w_{p,t}^\star[k] \ge 0, \qquad \widehat{w}_{p,t}[k] \ge 0,$$

and

$$\sum_p w_{p,t}^\star[k] = \sum_p \widehat{w}_{p,t}[k] = 1.$$

The ideal and actual aggregates are

$$a_t^\star[k] = \sum_p w_{p,t}^\star[k] z_{p,t}^\star[k],$$

and

$$\widehat{a}_t[k] = \sum_p \widehat{w}_{p,t}[k] \widehat{z}_{p,t}[k].$$

Lemma C.4 (Perturbation of mask–reliability aggregation)**.** *Suppose*

$$|z_{p,t}^\star[k]| \le G_k.$$

Then

$$|\widehat{a}_t[k] - a_t^\star[k]|^2 \le 2 \sum_p \widehat{w}_{p,t}[k] |e_{p,t}^{\mathrm{tr}}[k]|^2 \\ + 2G_k^2 \|\widehat{w}_t[k] - w_t^\star[k]\|_1^2. \tag{148}$$

Proof. Using

$$\widehat{z}_{p,t}[k] = z^\star_{p,t}[k] + e^{\mathrm{tr}}_{p,t}[k],$$

we obtain

$$\begin{aligned}
&\widehat{a}_t[k] - a^\star_t[k] \\
&= \sum_p \widehat{w}_{p,t}[k]\widehat{z}_{p,t}[k] - \sum_p w^\star_{p,t}[k] z^\star_{p,t}[k] && (149)\\
&= \sum_p \widehat{w}_{p,t}[k]\left[z^\star_{p,t}[k] + e^{\mathrm{tr}}_{p,t}[k]\right] - \sum_p w^\star_{p,t}[k] z^\star_{p,t}[k] && (150)\\
&= \sum_p \widehat{w}_{p,t}[k] e^{\mathrm{tr}}_{p,t}[k] \\
&\quad + \sum_p \left[\widehat{w}_{p,t}[k] - w^\star_{p,t}[k]\right] z^\star_{p,t}[k]. && (151)
\end{aligned}$$

Let

$$A_k = \sum_p \widehat{w}_{p,t}[k] e^{\mathrm{tr}}_{p,t}[k],$$

and

$$B_k = \sum_p \left[\widehat{w}_{p,t}[k] - w^\star_{p,t}[k]\right] z^\star_{p,t}[k].$$

Equation equation 151 becomes

$$\widehat{a}_t[k] - a^\star_t[k] = A_k + B_k.$$

Using

$$|A_k + B_k|^2 \le 2|A_k|^2 + 2|B_k|^2,$$

we obtain

$$|\widehat{a}_t[k] - a^\star_t[k]|^2 \le 2|A_k|^2 + 2|B_k|^2. \quad (152)$$

Since the actual weights are non-negative and sum to one, Jensen's inequality gives

$$\begin{aligned}
|A_k|^2 &= \left|\sum_p \widehat{w}_{p,t}[k] e^{\mathrm{tr}}_{p,t}[k]\right|^2 && (153)\\
&\le \sum_p \widehat{w}_{p,t}[k] |e^{\mathrm{tr}}_{p,t}[k]|^2. && (154)
\end{aligned}$$

For B_k, the triangle inequality gives

$$\begin{aligned}
|B_k| &\le \sum_p |\widehat{w}_{p,t}[k] - w^\star_{p,t}[k]|\, |z^\star_{p,t}[k]| && (155)\\
&\le G_k \sum_p |\widehat{w}_{p,t}[k] - w^\star_{p,t}[k]| && (156)\\
&= G_k \|\widehat{w}_t[k] - w^\star_t[k]\|_1. && (157)
\end{aligned}$$

Hence

$$|B_k|^2 \le G_k^2 \|\widehat{w}_t[k] - w^\star_t[k]\|_1^2. \quad (158)$$

Substituting equation 154 and equation 158 into equation 152 proves equation 148. □

Define

$$\mathcal{T}_t = \sum_k \sum_p \widehat{w}_{p,t}[k] |e^{\mathrm{tr}}_{p,t}[k]|^2, \quad (159)$$

and

$$\mathcal{W}_t = \sum_k \|\widehat{w}_t[k] - w^\star_t[k]\|_1^2. \quad (160)$$

With

$$G_\infty = \max_k G_k,$$

summing equation 148 over k yields

$$\|\widehat{a}_t - a_t^\star\|^2 \le 2\mathcal{T}_t + 2G_\infty^2 \mathcal{W}_t. \tag{161}$$

The second term captures changes in support, mask decisions, and reliability normalization. For example, conditional on the same surviving set, let

$$w_p^\star = \frac{\rho_p^\star}{Z^\star}, \qquad \widehat{w}_p = \frac{\widehat{\rho}_p}{\widehat{Z}},$$

where

$$Z^\star = \sum_p \rho_p^\star, \qquad \widehat{Z} = \sum_p \widehat{\rho}_p.$$

If

$$\varepsilon_\rho = \sum_p |\widehat{\rho}_p - \rho_p^\star|,$$

then

$$|\widehat{Z} - Z^\star| \le \varepsilon_\rho.$$

Moreover,

$$|\widehat{w}_p - w_p^\star| = \left| \frac{\widehat{\rho}_p}{\widehat{Z}} - \frac{\rho_p^\star}{Z^\star} \right| \tag{162}$$

$$= \left| \frac{\widehat{\rho}_p - \rho_p^\star}{\widehat{Z}} + \rho_p^\star \left(\frac{1}{\widehat{Z}} - \frac{1}{Z^\star} \right) \right| \tag{163}$$

$$\le \frac{|\widehat{\rho}_p - \rho_p^\star|}{\widehat{Z}} + \rho_p^\star \frac{|\widehat{Z} - Z^\star|}{\widehat{Z} Z^\star}. \tag{164}$$

Summing over p and using $\sum_p \rho_p^\star = Z^\star$,

$$\|\widehat{w} - w^\star\|_1 \le \frac{\varepsilon_\rho}{\widehat{Z}} + \frac{|\widehat{Z} - Z^\star|}{\widehat{Z}} \tag{165}$$

$$\le \frac{2\varepsilon_\rho}{\widehat{Z}}. \tag{166}$$

Thus, whenever

$$\widehat{Z} \ge z_{\min} > 0,$$

$$\|\widehat{w} - w^\star\|_1 \le \frac{2\varepsilon_\rho}{z_{\min}}. \tag{167}$$

The hard sign mask itself is discontinuous at zero, so we do not treat it as a globally Lipschitz operator. Its decision is nevertheless stable away from the sign boundary. If

$$\widehat{z}_p = z_p^\star + e_p, \qquad \widehat{z}_0 = z_0^\star + e_0,$$

then

$$\widehat{z}_p \widehat{z}_0 - z_p^\star z_0^\star$$

$$= (z_p^\star + e_p)(z_0^\star + e_0) - z_p^\star z_0^\star \tag{168}$$

$$= e_p z_0^\star + e_0 z_p^\star + e_p e_0. \tag{169}$$

Hence

$$|\widehat{z}_p \widehat{z}_0 - z_p^\star z_0^\star|$$

$$\le |e_p||z_0^\star| + |e_0||z_p^\star| + |e_p||e_0|. \tag{170}$$

Therefore the mask decision is unchanged whenever

$$|z_p^\star z_0^\star| > |e_p||z_0^\star| + |e_0||z_p^\star| + |e_p||e_0|. \tag{171}$$

When equation 171 fails, the resulting support change is already accounted for by $\mathcal{W}_t$ in equation 161.

C.3.4 RANK-CONSTRAINED RECOVERY

Let

$$\mathcal{P}_r(A)$$

denote a best rank-r approximation of A in Frobenius norm. By the Eckart–Young–Mirsky theorem,

$$\mathcal{P}_r(A) \in \arg\min_{\mathrm{rank}(B)\le r} \|A - B\|_F.$$

For the ideal pre-recovery aggregate $a_t^\star$, define its unavoidable rank-r residual by

$$\tau_{r,t} = \|a_t^\star - \mathcal{P}_r(a_t^\star)\|. \tag{172}$$

Lemma C.5 (Stability of rank-constrained recovery)**.** *For an arbitrary actual aggregate $\widehat{a}_t$,*

$$\|\mathcal{P}_r(\widehat{a}_t) - a_t^\star\| \le 2\|\widehat{a}_t - a_t^\star\| + \tau_{r,t}. \tag{173}$$

Consequently,

$$\|\mathcal{P}_r(\widehat{a}_t) - a_t^\star\|^2 \le 8\|\widehat{a}_t - a_t^\star\|^2 + 2\tau_{r,t}^2. \tag{174}$$

Proof. By the triangle inequality,

$$\|\mathcal{P}_r(\widehat{a}_t) - a_t^\star\| \le \|\mathcal{P}_r(\widehat{a}_t) - \widehat{a}_t\| + \|\widehat{a}_t - a_t^\star\|. \tag{175}$$

Because $\mathcal{P}_r(\widehat{a}_t)$ is a best rank-r approximation of $\widehat{a}_t$, and $\mathcal{P}_r(a_t^\star)$ is a feasible rank-r comparator,

$$\|\mathcal{P}_r(\widehat{a}_t) - \widehat{a}_t\| \le \|\mathcal{P}_r(a_t^\star) - \widehat{a}_t\|. \tag{176}$$

Apply the triangle inequality to the right-hand side:

$$\begin{aligned} \|\mathcal{P}_r(a_t^\star) - \widehat{a}_t\| &\le \|\mathcal{P}_r(a_t^\star) - a_t^\star\| \\ &\quad + \|a_t^\star - \widehat{a}_t\| && (177)\\ &= \tau_{r,t} + \|\widehat{a}_t - a_t^\star\|. && (178) \end{aligned}$$

Combining equation 175, equation 176, and equation 178,

$$\|\mathcal{P}_r(\widehat{a}_t) - a_t^\star\| \le \tau_{r,t} + 2\|\widehat{a}_t - a_t^\star\|, \tag{179}$$

which proves equation 173.

For the squared version, let

$$d_t = \|\widehat{a}_t - a_t^\star\|.$$

Then

$$\|\mathcal{P}_r(\widehat{a}_t) - a_t^\star\| \le 2d_t + \tau_{r,t}.$$

Using

$$(a+b)^2 \le 2a^2 + 2b^2$$

with

$$a = 2d_t, \qquad b = \tau_{r,t},$$

we obtain

$$\begin{aligned} (2d_t + \tau_{r,t})^2 &\le 2(2d_t)^2 + 2\tau_{r,t}^2 && (180)\\ &= 8d_t^2 + 2\tau_{r,t}^2. && (181) \end{aligned}$$

This proves equation 174. □

C.3.5 COMBINED V-HFLORA ERROR BOUND

We now combine the previous results.

On the correct-matching event $\mathcal{G}_t$, the ideal pre-recovery direction is

$$g_t^\circ = a_t^\star,$$

whereas the actual final direction is

$$\widehat{g}_t = \mathcal{P}_r(\widehat{a}_t).$$

Using equation 174,

$$\|\widehat{g}_t - g_t^\circ\|^2 \le 8\|\widehat{a}_t - a_t^\star\|^2 + 2\tau_{r,t}^2. \tag{182}$$

Substituting equation 161,

$$\|\widehat{g}_t - g_t^\circ\|^2 \le 8\left(2\mathcal{T}_t + 2G_\infty^2\mathcal{W}_t\right) + 2\tau_{r,t}^2 \tag{183}$$

$$= 16\mathcal{T}_t + 16G_\infty^2\mathcal{W}_t + 2\tau_{r,t}^2. \tag{184}$$

To control matching failure, assume only that

$$\|\widehat{g}_t\| \le G, \qquad \|g_t^\circ\| \le G.$$

Then, on $\mathcal{G}_t^c$,

$$\|\widehat{g}_t - g_t^\circ\| \le \|\widehat{g}_t\| + \|g_t^\circ\| \tag{185}$$

$$\le 2G, \tag{186}$$

and hence

$$\|\widehat{g}_t - g_t^\circ\|^2 \le 4G^2. \tag{187}$$

Theorem C.6 (V-HFLoRA approximation error)**.** *Under equation 125,*

$$\begin{aligned}\mathcal{E}_t = \mathbb{E}_t\|\widehat{g}_t - g_t^\circ\|^2 \le\ & 4G^2\delta_t^{\mathrm{match}} \\ & + 16\mathbb{E}_t\mathcal{T}_t \\ & + 16G_\infty^2\mathbb{E}_t\mathcal{W}_t \\ & + 2\mathbb{E}_t\tau_{r,t}^2.\end{aligned} \tag{188}$$

Proof. Split the conditional expectation according to $\mathcal{G}_t$:

$$\begin{aligned}\mathcal{E}_t = \ & \mathbb{E}_t\left[\|\widehat{g}_t - g_t^\circ\|^2\mathbf{1}_{\mathcal{G}_t}\right] \\ & + \mathbb{E}_t\left[\|\widehat{g}_t - g_t^\circ\|^2\mathbf{1}_{\mathcal{G}_t^c}\right].\end{aligned} \tag{189}$$

On $\mathcal{G}_t$, equation 184 gives

$$\begin{aligned}&\mathbb{E}_t\left[\|\widehat{g}_t - g_t^\circ\|^2\mathbf{1}_{\mathcal{G}_t}\right] \\ &\le \mathbb{E}_t\left[\left(16\mathcal{T}_t + 16G_\infty^2\mathcal{W}_t + 2\tau_{r,t}^2\right)\mathbf{1}_{\mathcal{G}_t}\right] \qquad (190)\\ &\le 16\mathbb{E}_t\mathcal{T}_t + 16G_\infty^2\mathbb{E}_t\mathcal{W}_t + 2\mathbb{E}_t\tau_{r,t}^2, \qquad (191)\end{aligned}$$

where the last inequality uses $\mathbf{1}_{\mathcal{G}_t} \le 1$.

On $\mathcal{G}_t^c$, equation 187 gives

$$\begin{aligned}&\mathbb{E}_t\left[\|\widehat{g}_t - g_t^\circ\|^2\mathbf{1}_{\mathcal{G}_t^c}\right] \\ &\le 4G^2\mathbb{E}_t[\mathbf{1}_{\mathcal{G}_t^c}] \qquad (192)\\ &= 4G^2\Pr_t(\mathcal{G}_t^c) \qquad (193)\\ &\le 4G^2\delta_t^{\mathrm{match}}. \qquad (194)\end{aligned}$$

Substituting equation 191 and equation 194 into equation 189 yields equation 188. □

C.4 Convergence Consequences

We now combine the general convergence result with the V-HFLoRA-specific approximation bound. The purpose of this subsection is twofold. First, we show that the residual terms appearing in the analysis remain uniformly bounded under standard bounded-update assumptions. Second, we use these bounds to establish convergence to a finite first-order stationary neighborhood.

Uniform boundedness assumptions. In addition to the assumptions used in Theorem C.1, suppose that there exist finite constants

$$G_F, \qquad G, \qquad G_\Delta, \qquad G_\infty$$

such that, for every communication round t,

$$\|\nabla F(y_t)\| \le G_F, \tag{195}$$

$$\|g_t^\circ\| \le G, \qquad \|\widehat{g}_t\| \le G, \tag{196}$$

and, for every participating source layer,

$$\|\Delta W_{j,t}^{(\ell)}\|_F \le G_\Delta. \tag{197}$$

We further assume that every ideally transported value, including the target-local fallback value, satisfies

$$|z_{p,t}^\star[k]| \le G_\infty. \tag{198}$$

Let

$$\mathcal{K}_t$$

denote the set of target coordinates participating in aggregation at round t, across all adapted layers, and define

$$D_{\max} = \sup_t |\mathcal{K}_t|. \tag{199}$$

Because every retained dimension is bounded by the corresponding frozen layer dimension and the number of adapted target layers is finite, $D_{\max} < \infty$.

We next show that each residual term appearing in Theorem C.6 admits a finite round-uniform bound.

Lemma C.7 (Uniform boundedness of V-HFLoRA residuals)**.** *Under equation 195–equation 199, there exist finite constants*

$$\epsilon_B, \qquad \epsilon_\delta, \qquad \epsilon_T, \qquad \epsilon_W, \qquad \epsilon_\tau$$

independent of the communication horizon T such that, for every round t,

$$B_t \le \epsilon_B, \tag{200}$$

$$\delta_t^{\mathrm{match}} \le \epsilon_\delta, \tag{201}$$

$$\mathbb{E}_t \mathcal{T}_t \le \epsilon_T, \tag{202}$$

$$\mathbb{E}_t \mathcal{W}_t \le \epsilon_W, \tag{203}$$

$$\mathbb{E}_t \tau_{r,t}^2 \le \epsilon_\tau. \tag{204}$$

One valid choice is

$$\epsilon_B = (G + G_F)^2, \tag{205}$$

$$\epsilon_\delta = 1, \tag{206}$$

$$\epsilon_T = 16 D_{\max} G_\Delta^2, \tag{207}$$

$$\epsilon_W = 4 D_{\max}, \tag{208}$$

$$\epsilon_\tau = D_{\max} G_\infty^2. \tag{209}$$

Proof. We establish the five bounds separately.

Population collaborative bias. Recall that

$$B_t = \|m_t - \nabla F(y_t)\|^2, \qquad m_t = \mathbb{E}_t[g_t^\circ].$$

By Jensen's inequality and equation 196,

$$\begin{aligned} \|m_t\| &= \|\mathbb{E}_t[g_t^\circ]\| && (210)\\ &\le \mathbb{E}_t\|g_t^\circ\| && (211)\\ &\le G. && (212) \end{aligned}$$

Using equation 195,

$$\begin{aligned} B_t &= \|m_t - \nabla F(y_t)\|^2 && (213)\\ &\le \left(\|m_t\| + \|\nabla F(y_t)\|\right)^2 && (214)\\ &\le (G + G_F)^2. && (215) \end{aligned}$$

This proves equation 200 with $\epsilon_B = (G + G_F)^2$.

Functional-matching failure. By definition,

$$\delta_t^{\mathrm{match}}$$

upper bounds a conditional probability. Therefore

$$0 \le \delta_t^{\mathrm{match}} \le 1.$$

Hence equation 201 holds with

$$\epsilon_\delta = 1.$$

This bound is deliberately conservative. The matching-stability analysis in Section C.3.1 provides sharper control whenever the population matching margin dominates the finite-sample score error.

Compatible-transport error. For same-dimensional matched layers, the basis-overlap transport is computed directly from the retained frozen-backbone bases and therefore introduces no finite-calibration registration error.

For a cross-dimensional pair, Lemma C.3 gives

$$\|e_{p,t}^{\mathrm{tr}}\|_F \le (\varepsilon_U^{\mathrm{reg}} + \varepsilon_V^{\mathrm{reg}})\, \|C_j^{[q]}\|_F. \tag{216}$$

Both the population and empirical Procrustes solutions are orthogonal. Consequently,

$$\begin{aligned} \varepsilon_U^{\mathrm{reg}} &= \|\widehat{R}_U - R_U^\star\|_2 && (217)\\ &\le \|\widehat{R}_U\|_2 + \|R_U^\star\|_2 && (218)\\ &= 2, && (219) \end{aligned}$$

and similarly

$$\varepsilon_V^{\mathrm{reg}} \le 2.$$

Therefore

$$\|e_{p,t}^{\mathrm{tr}}\|_F \le 4\|C_j^{[q]}\|_F. \tag{220}$$

Because

$$C_j^{[q]} = U_{j,[q]}^\top \Delta W_j V_{j,[q]},$$

and the retained basis matrices have orthonormal columns,

$$\begin{aligned} \|C_j^{[q]}\|_F &\le \|U_{j,[q]}^\top\|_2 \|\Delta W_j\|_F \|V_{j,[q]}\|_2 && (221)\\ &\le \|\Delta W_j\|_F && (222)\\ &\le G_\Delta. && (223) \end{aligned}$$

Thus

$$\|e_{p,t}^{\mathrm{tr}}\|_F^2 \le 16 G_\Delta^2. \tag{224}$$

Recall

$$\mathcal{T}_t = \sum_{k \in \mathcal{K}_t} \sum_p \widehat{w}_{p,t}[k] |e^{\mathrm{tr}}_{p,t}[k]|^2.$$

For every target coordinate k, the nonnegative aggregation weights satisfy

$$\sum_p \widehat{w}_{p,t}[k] = 1.$$

Equation equation 224 therefore gives

$$\sum_p \widehat{w}_{p,t}[k] |e^{\mathrm{tr}}_{p,t}[k]|^2 \le 16 G_\Delta^2. \tag{225}$$

Summing over at most $D_{\max}$ target coordinates,

$$\mathcal{T}_t \le 16 D_{\max} G_\Delta^2. \tag{226}$$

Taking conditional expectations proves equation 202.

Mask–reliability aggregation error. For every target coordinate k, both

$$w_t^\star[k] \qquad \text{and} \qquad \widehat{w}_t[k]$$

are nonnegative vectors with unit sum, including the local fallback rule. Hence both belong to the probability simplex, and therefore

$$\|\widehat{w}_t[k] - w_t^\star[k]\|_1 \le 2. \tag{227}$$

Squaring,

$$\|\widehat{w}_t[k] - w_t^\star[k]\|_1^2 \le 4.$$

Using the definition

$$\mathcal{W}_t = \sum_{k \in \mathcal{K}_t} \|\widehat{w}_t[k] - w_t^\star[k]\|_1^2,$$

we obtain

$$\mathcal{W}_t \le 4 D_{\max}. \tag{228}$$

Taking conditional expectations proves equation 203.

The sharper reliability-normalization and mask-margin bounds derived in Section C.3.3 may yield substantially smaller values, but are not required for uniform boundedness.

Rank-constrained recovery. By definition,

$$\tau_{r,t} = \|a_t^\star - \mathcal{P}_r(a_t^\star)\|.$$

Because the zero matrix is feasible in the rank-r approximation problem,

$$\tau_{r,t} \le \|a_t^\star\|.$$

For every target coordinate, $a_t^\star[k]$ is a convex combination of ideally transported values and, when the fallback rule is used, the corresponding local target value. By equation 198,

$$|a_t^\star[k]| \le G_\infty.$$

Consequently,

$$\tau_{r,t}^2 \le \|a_t^\star\|^2 \tag{229}$$

$$= \sum_{k \in \mathcal{K}_t} |a_t^\star[k]|^2 \tag{230}$$

$$\le D_{\max} G_\infty^2. \tag{231}$$

Taking conditional expectations proves equation 204. □

Average approximation error. Define

$$\overline{B}_T = \frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}B_t,$$

$$\overline{\delta}_T^{\text{match}} = \frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\delta_t^{\text{match}},$$

$$\overline{\mathcal{T}}_T = \frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\mathcal{T}_t,$$

$$\overline{\mathcal{W}}_T = \frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\mathcal{W}_t,$$

and

$$\overline{\tau}_{r,T}^2 = \frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\tau_{r,t}^2.$$

Lemma C.7 immediately gives

$$\overline{B}_T \le \epsilon_B, \tag{232}$$

$$\overline{\delta}_T^{\text{match}} \le \epsilon_\delta, \tag{233}$$

$$\overline{\mathcal{T}}_T \le \epsilon_T, \tag{234}$$

$$\overline{\mathcal{W}}_T \le \epsilon_W, \tag{235}$$

$$\overline{\tau}_{r,T}^2 \le \epsilon_\tau. \tag{236}$$

Theorem C.6 therefore implies

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\mathcal{E}_t \le 4G^2\epsilon_\delta + 16\epsilon_T + 16G_\infty^2\epsilon_W + 2\epsilon_\tau. \tag{237}$$

Corollary C.8 (Convergence to a stationary neighborhood)**.** *Suppose that the assumptions of Theorems C.1 and C.6, together with equation 195–equation 199, hold. Choose*

$$\eta = \frac{c}{\sqrt{T}}, \tag{238}$$

where $c > 0$, and suppose T is sufficiently large that

$$\eta \le \frac{1}{8L}.$$

Then

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\|\nabla F(y_t)\|^2 \le \frac{4[F(y_0)-F_\star]}{c\sqrt{T}} + \frac{4Lc\sigma^2}{\sqrt{T}} + 5\epsilon_B + 18G^2\epsilon_\delta + 72\epsilon_T + 72G_\infty^2\epsilon_W + 9\epsilon_\tau. \tag{239}$$

Equivalently,

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\|\nabla F(y_t)\|^2 \le O(T^{-1/2}) + \mathcal{R}, \tag{240}$$

where

$$\mathcal{R} = 5\epsilon_B + 18G^2\epsilon_\delta + 72\epsilon_T + 72G_\infty^2\epsilon_W + 9\epsilon_\tau < \infty. \tag{241}$$

Thus V-HFLoRA converges to a finite first-order stationary neighborhood, with the residual radius controlled by population collaborative bias, functional-matching failure, compatible-transport perturbation, mask–reliability aggregation error, and rank-constrained recovery.

Proof. From

$$\eta \le \frac{1}{8L},$$

we have

$$\Gamma = \frac{1}{2} - 2L\eta \ge \frac{1}{4}.$$

Hence

$$\frac{1}{\Gamma} \le 4,$$

and

$$1 + 2L\eta \le \frac{5}{4}, \qquad 1 + L\eta \le \frac{9}{8}.$$

Therefore

$$\frac{1 + 2L\eta}{\Gamma} \le 5, \qquad \frac{1 + L\eta}{\Gamma} \le \frac{9}{2}.$$

Substituting Lemma C.7 and equation 237 into Theorem C.1 gives

$$\begin{aligned}\frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}\|\nabla F(y_t)\|^2 \le \frac{4[F(y_0) - F_\star]}{\eta T} + 5\epsilon_B \\ + \frac{9}{2}\left[4G^2\epsilon_\delta + 16\epsilon_T + 16G_\infty^2\epsilon_W + 2\epsilon_\tau\right] \\ + 4L\eta\sigma^2. \end{aligned} \tag{242}$$

Using

$$\eta = \frac{c}{\sqrt{T}}$$

yields equation 239. □

Vanishing-residual special case. Corollary C.8 does not require the component errors to vanish with the number of communication rounds. For fixed calibration data, fixed retained ranks, and persistent client heterogeneity, the residual $\mathcal{R}$ may remain nonzero.

Under stronger approximation conditions, however, suppose that a sequence of problem instances or algorithmic configurations satisfies

$$\epsilon_B(T) \to 0, \qquad \epsilon_\delta(T) \to 0, \qquad \epsilon_T(T) \to 0,$$
$$\epsilon_W(T) \to 0, \qquad \epsilon_\tau(T) \to 0.$$

Then

$$\mathcal{R}_T \to 0,$$

and equation 240 reduces to

$$\frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}\|\nabla F(y_t)\|^2 \to 0. \tag{243}$$

Thus exact first-order stationarity is recovered as a stronger special case, whereas the general result guarantees convergence to a finite stationary neighborhood without requiring all approximation errors to vanish.

D Computational Complexity

We report the additional server-side cost of V-HFLoRA, since local LoRA training is unchanged from standard federated adaptation. The main distinction is that representation matching is performed over candidate source–target layer pairs, whereas transport and aggregation are performed only for the pairs retained after matching. The latter operations are further restricted to the low-dimensional retained coordinates.

Let K_t be the number of active clients at round t, L the approximate number of adapted layers per client, n_{cal} the calibration-set size, k the reduced response dimension used for functional matching, q a representative retained coordinate dimension, r the LoRA rank, and $\bar{m}$ the average number of selected pairs per target layer.

Per-model preparation. The singular bases of all frozen backbone weights are computed once before federated training and therefore do not incur a round-wise SVD cost. At each round, the server projects the uploaded low-rank update

$$\Delta W = BA$$

onto the retained frozen-backbone coordinates. For one layer, this requires

$$O\big((d_{\mathrm{out}} + d_{\mathrm{in}})qr + q^2 r\big)\,, \tag{244}$$

without explicitly forming the dense $d_{\mathrm{out}} \times d_{\mathrm{in}}$ update.

Each active model is also evaluated once on the common calibration set. The resulting intermediate responses are cached and reused for all subsequent layer comparisons, so the calibration-forward cost is

$$\mathcal{C}_t^{\mathrm{cal}} = \sum_{i \in \mathcal{C}^t} \mathcal{C}_{\mathrm{cal},i,t}. \tag{245}$$

Pairwise functional matching. Functional matching is the only stage that is evaluated over candidate source–target layer pairs. After response reduction, regularised CCA, target-oriented ridge alignment, and CKA evaluation require

$$O(n_{\mathrm{cal}} k^2 + k^3)$$

for one candidate pair. With approximately L admissible layers per model, the round-level matching cost is therefore

$$O\big(K_t^2 L^2 \left(n_{\mathrm{cal}} k^2 + k^3\right)\big)\,. \tag{246}$$

The monotone layer-alignment step adds $O(K_t^2 L^2)$ operations and is typically lower order than the CCA–ridge–CKA computations. If matching is restricted to a local layer window, the corresponding candidate-layer factor L is replaced by the window size.

Post-matching target updates. Once the pair set has been selected, the remaining operations scale with $\bar{m}$, rather than with all candidate layer pairs.

For same-dimensional layers, the frozen source–target basis overlaps can be precomputed once. Applying the bilateral transport

$$C_{j \to i} = P^U_{j \to i} C_j P^{V\top}_{j \to i}$$

costs

$$O(q^3) \tag{247}$$

when the source and target retained dimensions are both of order q.

For cross-dimensional layers, V-HFLoRA first restricts both layers to their mutual retained dimension and registers the two coordinate systems using the calibration responses. For one matched pair, forming the retained cross-covariance matrices costs

$$O(n_{\mathrm{cal}} q^2),$$

while the corresponding orthogonal-Procrustes decompositions and bilateral coordinate transport require

$$O(q^3).$$

Hence the cross-dimensional transport cost is

$$O\big(n_{\mathrm{cal}} q^2 + q^3\big)\,. \tag{248}$$

These operations remain entirely in the retained coordinates; no dense cross-model ambient transport matrix is constructed.

Mask–reliability aggregation then scans only the supported coordinates of the selected pairs. For one target layer, its cost is

$$O(\bar{m} q^2). \tag{249}$$

Unsupported coordinates are ignored and sign-conflicting pairs are removed before reliability-weight normalisation, so no additional dense padding or alignment operation is required.

Finally, let

$$\bar{r} = \min(r, q).$$

A rank-$\bar{r}$ truncated decomposition of the aggregated target coordinate, followed by lifting the recovered factors through the retained target bases, costs

$$O\big(q^2 \bar{r} + (d_{\mathrm{out}} + d_{\mathrm{in}}) q \bar{r}\big)\,. \tag{250}$$

Table 6: Training and V-HFLoRA hyperparameter settings. Unless otherwise stated, the same configuration is used across all benchmarks.

Component	Hyperparameter / setting	Default value
Federated training	Number of clients K	10
	Dirichlet parameter α_{dir}	0.5
	Communication rounds	100
	Local epochs per round	1
	Learning rate	5×10^{-3}
	Batch size	16
	Optimizer	AdamW
	Weight decay	10^{-4}
Backbone	Pretrained backbone	Frozen
	Task classifier	Trainable
LoRA	Rank r	16
	Scaling α_{LoRA}	8
Update-conditioned geometry	Retained update-energy threshold ρ	0.90
Calibration	Calibration-set size n_{cal}	128
	Calibration batch size	32
	Functional-response batches	2
	Maximum response samples	512
Response reduction	PCA variance-retention threshold τ_{pca}	0.99
	Maximum PCA dimension $d_{\mathrm{pca}}^{\max}$	128
CCA matching	CCA regularization ϵ_{reg}	10^{-4}
	Minimum canonical-correlation threshold τ_{cca}	$10^{-4\dagger}$
Target-oriented alignment	Ridge penalty λ_m	10^{-4}
Monotone matching	Pair-order penalty λ_{pair}	129
	Maximum pair count k_{pair}	5

Overall scaling. Combining the above terms, the representative round-level server cost is

$$\begin{aligned}\mathcal{C}_t^{\mathrm{server}} = \mathcal{C}_t^{\mathrm{cal}} &+ O\big(K_t L\left[(d_{\mathrm{out}} + d_{\mathrm{in}})qr + q^2 r\right]\big) \\ &+ O\big(K_t^2 L^2\left[n_{\mathrm{cal}} k^2 + k^3\right]\big) \\ &+ O(K_t L \bar{m}\, \mathcal{C}_{\mathrm{tr}}(q)) \\ &+ O\big(K_t L\left[\bar{m} q^2 + q^2 \bar{r} + (d_{\mathrm{out}} + d_{\mathrm{in}})q\bar{r}\right]\big)\,, \end{aligned} \tag{251}$$

where

$$\mathcal{C}_{\mathrm{tr}}(q) = \begin{cases} O(q^3), & \text{same-dimensional transport,} \\ O(n_{\mathrm{cal}} q^2 + q^3), & \text{cross-dimensional registered transport.} \end{cases} \tag{252}$$

The dominant additional cost is therefore generally the pairwise functional matching in equation 246, which scales with the number of candidate model–layer pairs. In contrast, transport, mask–weight aggregation, and recovery are evaluated only for the selected pairs and remain confined to the retained coordinates. Thus pair pruning and update-conditioned retention limit the cost of the heterogeneous-model operations after matching.

E Hyperparameter Setting

F Baseline Implementations

All baselines follow the same heterogeneous-client protocol as V-HFLoRA. They use the same federated data partitions, client backbone assignments, pretrained initialization, optimizer, number of communication rounds, local epochs, LoRA rank, LoRA scaling, batch size, and evaluation protocol.

For every client, the pretrained vision backbone is frozen, while the LoRA adapters and task classifier are optimized locally. Server-side information sharing follows the aggregation rule of each baseline.

Let θ_i^t denote the locally trainable parameters of client i at round t, and let n_i denote the number of its local training examples. For clarity, the main comparison groups the baselines according to the type of cross-client information transfer they perform: no communication, same-backbone parameter aggregation, cross-backbone aligned aggregation, and prediction-space distillation.

Local Only. Local Only removes all server-side communication and aggregation. Each client independently fine-tunes its LoRA adapters and classifier using only its local data:

$$\theta_i^{t+1} = \mathrm{LocalTrain}\left(\theta_i^t; \mathcal{D}_i\right). \tag{253}$$

No parameters, logits, or adapter updates are exchanged across clients. This baseline measures the performance of purely personalized local fine-tuning under the same backbone-heterogeneous client setting.

Grouped-Backbone FedAvg. Grouped-Backbone FedAvg is a fixed-group variant inspired by FedGroup Duan et al. (2021). Rather than dynamically grouping clients according to update similarity, we group them by backbone identity, since clients using the same backbone have directly compatible layer structures and LoRA parameter shapes. Let b_i denote the backbone type of client i, and define

$$\mathcal{G}_b = \{i : b_i = b\}. \tag{254}$$

After local training, the server performs standard data-size-weighted FedAvg within each backbone group:

$$\bar{\theta}_b^{t+1} = \sum_{i \in \mathcal{G}_b} \frac{n_i}{\sum_{j \in \mathcal{G}_b} n_j} \theta_i^{t+1}. \tag{255}$$

The aggregated parameters $\bar{\theta}_b^{t+1}$ are returned only to clients in $\mathcal{G}_b$. Thus, this baseline enables conventional federated sharing among architecturally compatible clients but does not transfer LoRA information across different backbone variants.

Aligned FedAvg. Aligned FedAvg is a cross-backbone aggregation baseline designed to test whether layer alignment and geometric transport alone are sufficient for heterogeneous LoRA aggregation. It uses the same functional layer matching and compatible transport procedures as V-HFLoRA to rewrite matched source updates in the retained coordinate system of each target layer.

For target client i and target layer ℓ, let $\mathcal{M}_i^{(\ell)}$ denote the matched source set and $C_{p\to i,t}^{(\ell)}$ the transported coordinate from matched source $p = (j, \pi)$. Aligned FedAvg directly averages all transported source coordinates using data-size weights:

$$C_{i,t}^{\mathrm{AF},(\ell)} = \sum_{p=(j,\pi)\in\mathcal{M}_i^{(\ell)}} \frac{n_j}{\sum_{p'=(j',\pi')\in\mathcal{M}_i^{(\ell)}} n_{j'}} C_{p\to i,t}^{(\ell)}. \tag{256}$$

The resulting coordinate is then recovered to the prescribed target LoRA rank using the same low-rank recovery procedure as V-HFLoRA.

Unlike V-HFLoRA, Aligned FedAvg does not apply coordinate masking or coordinate-wise numerical compatibility weighting after transport. All matched and transported source coordinates therefore contribute directly to the weighted average. This baseline isolates whether functional matching and compatible transport alone are sufficient, or whether the proposed conflict-aware mask–weight aggregation is also necessary.

FedDF. FedDF Lin et al. (2020) avoids direct parameter aggregation across heterogeneous backbones by transferring knowledge in prediction space. After local training, the server evaluates all participating client models on the same server-side unlabelled calibration set $\mathcal{D}_{\mathrm{cal}}$ and constructs an ensemble teacher. For $x \in \mathcal{D}_{\mathrm{cal}}$, the ensemble prediction is

$$p_{\mathrm{ens}}(y \mid x) = \sum_{i=1}^{K} \frac{n_i}{\sum_{j=1}^{K} n_j} p_i(y \mid x), \tag{257}$$

where $p_i(y \mid x)$ is the predictive distribution of client i.

Each client-compatible student is then updated by minimizing

$$\mathcal{L}_{\mathrm{KD}} = \frac{1}{|\mathcal{D}_{\mathrm{cal}}|} \sum_{x \in \mathcal{D}_{\mathrm{cal}}} \mathrm{KL}\left(p_{\mathrm{ens}}(\cdot \mid x) \,\|\, p_{\mathrm{student}}(\cdot \mid x)\right). \tag{258}$$

Because FedDF transfers predictions rather than parameter tensors, it remains applicable when participating clients use structurally different backbones. For fairness, FedDF uses the same server-side unlabelled calibration-data budget as V-HFLoRA.

F.1 Same-Backbone LoRA-FL Baselines

Table 10 compares V-HFLoRA with representative same-backbone LoRA-FL baselines. The reported results are obtained by re-implementing the corresponding methods under our unified OpenCLIP ViT-B/16 federated vision fine-tuning protocol rather than copying results from the original papers. Specifically, all methods use the same client partition, pretrained backbone, LoRA insertion modules, LoRA rank, LoRA scaling, communication rounds, local epochs, optimizer, learning rate, batch size, and evaluation protocol as V-HFLoRA.

For method-specific hyperparameters, we follow the values recommended in the corresponding original papers whenever available. When a method was originally proposed outside the CV/VFM setting, such as for language-model or general foundation-model fine-tuning, we port its aggregation rule to our OpenCLIP LoRA fine-tuning pipeline while keeping all common training and evaluation factors fixed. This protocol-normalised comparison is intended to compare aggregation mechanisms under a common federated vision fine-tuning environment rather than to reproduce the original task-specific experimental setting of each baseline.

Concretely, FLoRA, FlexLoRA, HLoRA-P, HetLoRA, pFedLoRA, LFAIR, and HeLoRA are implemented as same-backbone LoRA-FL baselines using OpenCLIP ViT-B/16 for all clients. Because all clients share the same backbone architecture, these methods do not require cross-backbone functional layer matching or dimensional transport.

For V-HFLoRA, the homogeneous-backbone setting likewise removes the need for cross-dimensional transport, while retaining the update-conditioned backbone representation and the proposed target-space mask–weight aggregation. This comparison therefore isolates whether the conflict-aware aggregation mechanism remains beneficial when structural backbone incompatibility is removed.

G DINOv2 Backbones

Our main experiments use heterogeneous OpenCLIP backbones. To examine whether V-HFLoRA is tied to OpenCLIP-specific representations, we further conduct an applicability check using DINOv2 backbones Oquab et al. (2023). Specifically, clients are assigned heterogeneous DINOv2 small/base backbones on CIFAR-10, while keeping the same federated fine-tuning protocol.

Table 7 shows that V-HFLoRA remains effective after replacing OpenCLIP with DINOv2. It achieves 98.31% average personalised accuracy and slightly outperforms all compared baselines. Because CIFAR-10 is already close to saturation under DINOv2 features, the margin is small; nevertheless, this result indicates that the proposed target-centred aggregation pipeline is not tied to OpenCLIP-specific representations and can be applied to alternative pretrained vision backbones without changing the core algorithm.

Table 7: DINOv2 applicability on CIFAR-10. Accuracy (%).

Method	Acc.
Local Only	98.20
Grouped FedAvg	98.12
FedDF	97.59
V-HFLoRA	**98.31**

H Additional Cross-Backbone Stress Tests

We further evaluate V-HFLoRA under stronger backbone heterogeneity beyond the main OpenCLIP ViT-B/32 and ViT-B/16 setting. These experiments are intended as supplementary stress tests rather than full baseline comparisons. The cross-family ViT setting uses OpenCLIP ViT-B/32 and timm ViT-S/16 clients, while the cross-size OpenCLIP setting uses OpenCLIP ViT-B/32 and ViT-L/14 clients.

Table 8: Per-backbone accuracy breakdown for the OpenCLIP ViT-B/32 + ViT-L/14 cross-size stress test on CIFAR-10. Overall denotes the average personalized accuracy across all clients.

Method	ViT-B/32 clients	ViT-L/14 clients	Overall
Local Only	92.5	96.5	94.5
Grouped FedAvg	95.8	97.0	96.4
V-HFLoRA	**96.8**	**97.4**	**97.1**

Table 9: Effect of the functional-matching recomputation interval on CIFAR-10 with heterogeneous OpenCLIP backbones.

Matching interval	Avg. client acc.
Every round	**96.32%**
Every 2 rounds	95.83%
Every 3 rounds	95.17%

The latter introduces a stricter hidden-dimension mismatch and therefore requires source-to-target transport before aggregation.

I Ablation on Functional-Matching Frequency

V-HFLoRA recomputes response-based functional matching at every communication round by default. To assess whether this cost can be amortised, we reuse the same layer correspondence for several consecutive rounds before recomputing it.

Table 9 reports the results on CIFAR-10 with heterogeneous OpenCLIP backbones. Recomputing the correspondence every round achieves the highest average client accuracy of 96.32%. Increasing the matching interval to two and three rounds reduces the accuracy to 95.83% and 95.17%, respectively. The gradual degradation suggests that functional correspondence remains reasonably stable over short intervals, but becomes less reliable as the client models continue to evolve.

These results expose a direct accuracy–efficiency trade-off. Per-round matching gives the strongest performance, while less frequent recomputation reduces server-side matching overhead at the cost of a modest loss in accuracy.

J Additional Experimental Analysis

J.1 Same-backbone Study.

Although V-HFLoRA is designed primarily for heterogeneous-backbone federated adaptation, we additionally evaluate it in a homogeneous setting where all clients use OpenCLIP ViT-B/16. This setting removes cross-backbone architectural and dimensional incompatibility and therefore serves as a compatibility test of the proposed aggregation mechanism. The gain is small on CIFAR-10, where several same-backbone aggregation methods already attain similar accuracy, but becomes larger on CIFAR-100 and particularly on TinyImageNet.

Table 10: Same-backbone LoRA-FL comparison under OpenCLIP ViT-B/16. Gain denotes the absolute improvement of V-HFLoRA over the strongest same-backbone LoRA-FL baseline.

	Same-backbone setting: OpenCLIP ViT-B/16								
	Same-backbone LoRA-FL baselines							**Ours**	Improvement
Dataset	FLoRA Wang et al. (2024)	FlexLoRA Bai et al. (2024)	HLoRA-P Fan et al. (2025)	HetLoRA Cho et al. (2024)	pFedLoRA Yi et al. (2023)	LFAIR Bian et al. (2025)	HeLoRA Fan et al. (2025)	**V-HFLoRA**	Gain
CIFAR-10	90.68 ± 1.95	92.43 ± 1.88	91.08 ± 1.81	91.28 ± 1.24	95.25 ± 1.37	94.61 ± 1.13	94.32 ± 1.86	$\mathbf{95.38 \pm 0.50}$	+0.13
CIFAR-100	73.54 ± 2.74	74.10 ± 1.04	74.45 ± 1.41	74.18 ± 2.83	72.24 ± 1.31	71.68 ± 1.56	71.82 ± 1.42	$\mathbf{75.28 \pm 0.84}$	+0.83
TinyImageNet	51.13 ± 1.07	61.77 ± 1.66	60.92 ± 1.26	62.84 ± 1.33	62.41 ± 1.52	62.57 ± 1.39	65.38 ± 1.91	$\mathbf{68.98 \pm 0.71}$	+3.60

In the same-backbone setting, V-HFLoRA no longer needs cross-dimensional transport, but the client updates can still differ because of heterogeneous local data. The method therefore retains its update-conditioned backbone representation and mask–weight aggregation: incompatible source coordinates are filtered before the remaining updates are weighted and combined. The improvements on CIFAR-100 and TinyImageNet suggest that this conflict-aware aggregation mechanism remains useful even when structural backbone mismatch is absent. Thus, the method does not rely solely

on cross-backbone dimensional incompatibility for its gains, while its primary use case remains heterogeneous-backbone federated adaptation.

J.2 Hyperparameter Sensitivity

We evaluate the sensitivity of V-HFLoRA to three key hyperparameters: the retained-energy threshold ρ, the size of the server-side unlabelled calibration set n_{cal}, and the maximum number of source pairs k_{pairs} retained for each target layer. Unless otherwise specified, all experiments use CIFAR-100 under the main heterogeneous-backbone setting, and only the indicated hyperparameter is varied while all remaining settings are kept fixed.

Retained-energy threshold. The threshold ρ determines the smallest frozen-backbone subspace that retains a prescribed fraction of the projected LoRA-update energy. A larger ρ therefore preserves more backbone directions but also increases the dimensionality of subsequent transport and aggregation. Table 11 reports both accuracy and the average retained ratio across clients, layers, and communication rounds.

Table 11: **Sensitivity to the retained-energy threshold ρ on CIFAR-100.** Retained ratio denotes the average fraction of frozen-backbone directions kept by the update-conditioned selection rule.

ρ	0.80	**0.90**	0.95	0.99
Retained ratio	31%	**44%**	57%	79%
Accuracy (%)	75.31	**76.43**	76.24	75.79

Performance remains stable around $\rho = 0.9$–0.95. Increasing ρ beyond this range substantially enlarges the retained coordinate space but provides no additional accuracy gain. We therefore use $\rho = 0.9$ as the default setting, which provides a favourable balance between preserving update-relevant backbone geometry and controlling the dimensionality of cross-client transport.

Calibration-set size and number of source pairs. The server-side unlabelled calibration set is used only for functional layer matching. We vary n_{cal} to determine how many calibration examples are required to obtain stable source–target correspondences. We additionally vary k_{donor}, the maximum number of matched source layers retained for each target layer before compatible transport and aggregation.

Table 12: **Sensitivity to calibration-set size and the number of source pairs on CIFAR-100.**

Hyperparameter	Setting	Acc. (%)
Calibration size n_{cal}	32	74.71
	64	75.64
	128	**76.43**
	256	76.51
Maximum source pairs k_{pair}	1	74.86
	2	75.57
	4	**76.43**
	5	76.27

Increasing the calibration-set size improves performance from 32 to 128 examples, after which the improvement largely saturates. This suggests that functional matching does not require a large server-side unlabelled dataset. Similarly, using multiple source pairs improves over relying on a single matched source, but expanding the pairs set beyond the default value provides little additional benefit. This is consistent with the role of pairs selection: additional sources can provide useful cross-client information, but increasingly weakly matched sources need not improve the target update.