



City Research Online

City St George's, University of London

Citation: Mi, Y., Hao, Z., Zhao, G., Wu, Y., Liu, C., Shen, J. & Qian, X. (2026). CanonFedRec: A Canonical Geometric Framework for Personalized Federated Recommendation. Paper presented at the KDD '26: The 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 9-13 Aug 2026, Jeju Island, Republic of Korea. doi: 10.1145/3770855.3817845

This is the published version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/38468/>

Link to published version: <https://doi.org/10.1145/3770855.3817845>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

CanonFedRec: A Canonical Geometric Framework for Personalized Federated Recommendation

Yunqi Mi
School of Software
Engineering
Xi'an Jiaotong University
Xi'an, Shaanxi, China
miyunqi@stu.xjtu.edu.cn

Zeyu Hao
Institute of Artificial
Intelligence and Robotics
Xi'an Jiaotong University
Xi'an, Shaanxi, China
hzy950509@stu.xjtu.edu.cn

Guoshuai Zhao*
School of Software
Engineering
Xi'an Jiaotong University
Xi'an, Shaanxi, China
guoshuai.zhao@xjtu.edu.cn

Yingjie Wu
School of Software
Engineering
Xi'an Jiaotong University
Xi'an, Shaanxi, China
nostalgiauw@stu.xjtu.edu.cn

Chengxu Liu
School of Software
Engineering
Xi'an Jiaotong University
Xi'an, Shaanxi, China
chengxu.liu@xjtu.edu.cn

Jialie Shen
City St George's
University of London
London, United Kingdom
jerry.shen@citystgeorges.ac.uk

Xueming Qian
School of Information and
Communication Engineering
Xi'an Jiaotong University
Xi'an, Shaanxi, China
qianxm@mail.xjtu.edu.cn

Abstract

With the increasing demand for privacy protection, the federated recommendation (FedRec) has become a critical research topic. Prevailing frameworks continuously decouple item representation to bypass the heterogeneity of client data, sacrificing the universality of aggregated item embeddings, losing the collaborative benefits of federated learning. This trend not only introduces prohibitive memory costs on resource-constrained edge devices, but also exacerbates the isolation of user interests, leading to a significant loss of recommendation diversity. In this paper, we attribute the failure of aggregated item embeddings to the server-side **Global Geometric Misalignment** and client-side **Local Filtering Bubbles** after empirical and theoretical analysis. Then, we propose CanonFedRec, a novel framework that resolves the above challenges through a canonical geometric perspective. First, we shift the optimization landscape from Euclidean space to a unit hypersphere to explicitly decouple local magnitude noise unrelated to popularity from semantic relationships. To rectify the global geometric inconsistency, we propose a Riemannian Projection-based Aggregation mechanism that performs global updates in a geometrically compatible tangent space. Furthermore, to mitigate the local filtering bubbles of clients, we propose a Divergence-Aware Elastic Alignment mechanism. This treats item-wise variance as a proxy for client-side cognitive divergence, and dynamically adapting optimization objectives via elastic decision boundaries. Extensive experiments on four datasets demonstrate that CanonFedRec not only significantly enhances the representational power of aggregated item embeddings, but also achieves superior performance while reducing memory costs by up to 40× compared to the best FedRec approach.

*Corresponding Author.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.
KDD 2026, Jeju Island, Republic of Korea.
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2259-2/2026/08
<https://doi.org/10.1145/3770855.3817845>

CCS Concepts

• Information systems → Recommender systems.

Keywords

Recommender Systems; Federated Learning; Manifold Learning; Representation Learning

ACM Reference Format:

Yunqi Mi, Zeyu Hao, Guoshuai Zhao, Yingjie Wu, Chengxu Liu, Jialie Shen, and Xueming Qian. 2026. CanonFedRec: A Canonical Geometric Framework for Personalized Federated Recommendation. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2 (KDD 2026)*, August 9–13, 2026, Jeju Island, Republic of Korea. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3770855.3817845>

Resource Availability:

The source code of this paper has been made publicly available at <https://doi.org/10.5281/zenodo.20438127>.

1 Introduction

With the explosion of information, recommender systems have permeated every corner of daily life [31, 36, 41]. With increasing demand for data privacy, federated recommendation (FedRec) has become a critical research topic [22, 26, 46]. In this paradigm, multiple clients jointly train a unified system with a server, while sensitive interaction data is siloed locally on clients. Each client optimizes three parameter types: user embeddings, item embeddings, and recommender parameters. After optimization, each client sends all parameters, except user embeddings, back to the server for aggregation and redistribution.

Compared to centralized recommendations, clients in FedRec hold siloed data subsets, thereby exhibiting heterogeneity. This leads to a decline in the personalization of the aggregation model, most pronounced in item embeddings, thereby reducing recommendation accuracy. Since a client updates only the interacted item embeddings in a sparse data subset, naive aggregation of these sparse updates often leads to performance drops. Hence, prevailing frameworks typically decouple parameters, using local bias

terms, client-specific parameters to rectify aggregated item embeddings and restore personalization [20, 47, 48]. As shown in Figure 1, prevailing frameworks increasingly rely less on unified item embeddings and, in some cases, no longer require them [8], which actually loses the collaborative benefits of federated learning [4, 16]. This not only introduces significant memory and maintenance costs on resource-constrained edge devices, but also exacerbates the silos of user interests, leading to a significant decrease in diversity. To address the above challenges, we must answer a fundamental question: **What causes the failure of aggregated item embeddings?**

Hence, we visualize the item embeddings on clients and the server during the training process in Figure 2, and identify three phenomena that characterize this failure¹: 1) *Local Unbounded Norm Inflation*. The magnitude of a client’s interacted item embeddings in the federated setting increases rapidly and unboundedly, no longer correlated with popularity [24]. 2) *Local Representation Homogeneity*. Within a client, the item-wise similarity and homogeneity of interacted items sharply increase, indicating that representations rapidly collapse into a narrow hypercone driven by local preferences. 3) *Global Item-Wise Energy Decay*². Paradoxically, while local norms explode, the norm of aggregated item embeddings decays, indicating that updates from different clients cancel each other out during aggregation. These observations highlight the dilemma within the FedRec: **Local Filtering Bubbles** [19, 42] and **Global Geometric Misalignment**. On the client side, extreme data sparsity leads to overconfidence in a locally biased sub-dataset. To maximize the effect of projection onto local user embeddings, the local optimizer actively pushes interactive item embeddings toward specific directions determined by local preferences. This allows item embeddings encode client-specific noise rather than intrinsic semantics. On the server side, this local overconfidence manifests as geometric misalignment. Unlike embeddings in Computer Vision or Natural Language Processing tasks, which capture intrinsic semantic features, embeddings in recommendation represent latent topological structures [11, 31, 43]. For instance, “beer” and “diapers” are geometrically close not because of semantic attributes, but due to their co-occurrence topological structure [23]. However, due to heterogeneity, different clients optimize these topological structures on incompatible geometric manifolds. Consequently, naive Euclidean averaging (*i.e.*, FedAVG [27]) fails to reconcile these divergent manifolds. The aggregation process erases learned topological information, directly causing the failure of aggregated embeddings.

Based on the above analysis, we propose CanonFedRec, a canonical geometric framework that unifies clients’ diverse, fragmented perspectives into a coherent global manifold. First, to address the local unbounded norm inflation, we shift the optimization landscape from the flat Euclidean space to a unit hypersphere[34], explicitly eliminating the interference of magnitude noise unrelated to popularity. This ensures the model focuses on learning high-quality topological structures. To solve the global geometric misalignment, we propose a novel Riemannian Projection-based Aggregation mechanism. We project local updates into a shared geometry-compatible tangent space, enabling servers to implement FedAVG simply and

¹We choose MF as the backbone recommender, training it under centralized and federated settings on the ML-100k dataset. See detailed empirical verification in Appendix A.

²We use aggregation energy ratio to measure the consistency of client updates, denoting as the ratio of the item-wise global aggregated norm to the sum of local update norms.

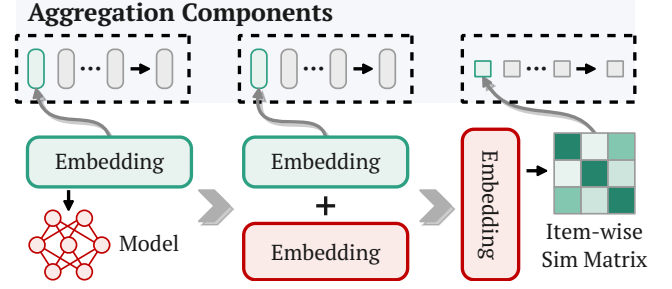


Figure 1: The decoupling trap in existing FedRec. Current approaches rely on non-aggregated strategies or parameters (Red) to rectify personalization, implicitly abandoning the collaborative benefits of aggregated embeddings (Green).

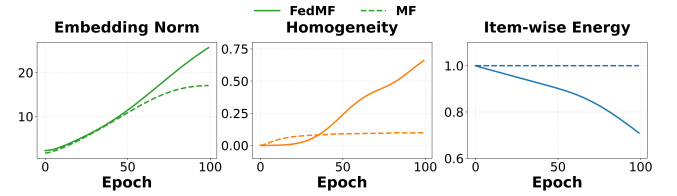


Figure 2: Trends of unbounded norm inflation and representation homogeneity on clients, and item-wise energy decay on the server during training on ML-100k dataset.

efficiently while strictly preserving the topological structure of aggregated item embeddings. To mitigate local filtering bubbles, we introduce a Divergence-Aware Elastic Alignment mechanism. We treat item-wise variance as a proxy for clients’ cognitive divergence and employ elastic decision boundaries in local optimization. This effectively prevents the client from over-optimizing items with divergence, achieving a dynamic trade-off between local personalization and global consistency. Extensive experiments on four real-world datasets demonstrate that CanonFedRec not only maintains the collaborative benefits to achieve superior performance, but also exhibits significant efficiency gains, reducing memory costs by up to 40× compared to the best FedRec approach.

The main contributions of this work can be summarized as:

- We theoretically analyze the failure of aggregated embeddings stems from local filtering bubbles and global geometric misalignment. And propose CanonFedRec, a canonical geometric framework to solve the above challenges.
- We propose a Riemannian Projection-based Aggregation mechanism. By mapping local updates to a geometrically compatible tangent space, we preserve the global topological structure while maintaining the simplicity of FedAVG.
- We introduce a Divergence-Aware Elastic Alignment mechanism. By flexibly adapting local optimization objectives via elastic decision boundaries, we achieve a dynamic trade-off between local personalization and global consistency.
- Extensive experiments on four datasets demonstrate that CanonFedRec achieves superior performance. It reduces memory costs by up to 40× compared to the best approach, demonstrating advantages in effectiveness and efficiency.

2 Notation and Definition

Consider a typical FedRec system with a user set $\mathcal{U}$ ($|\mathcal{U}| = M$) and an item set $\mathcal{V}$ ($|\mathcal{V}| = N$). Users are distributed across C clients ($C \leq M$), where each client $c \in C$ holds a private sub-dataset $\mathcal{D}_c = \{(u, v) | u \in \mathcal{U}_c, R_{uv} = 1\}$. $\mathcal{R} \in \mathbb{R}^{M \times N}$ denotes the co-occurrence matrix. The goal of FedRec is to collaboratively learn parameters of a recommender across clients without sharing private sub-datasets, including user embeddings $\mathbf{U} \in \mathbb{R}^{M \times d}$, item embeddings $\mathbf{V} \in \mathbb{R}^{N \times d}$ and recommendation parameters Θ . d is the latent dimension. In each communication round t , the server distributes the current global state $\Phi^{(t)} = \{\Theta^{(t)}, \mathbf{V}^{(t)}\}$ to a set of clients C_t . Each client $c \in C_t$ initializes their local model with $\Phi^{(t)}$ and performs local optimization on private dataset $\mathcal{D}_c$ with local user embeddings $\mathbf{U}_c \in \mathbb{R}^{M_c \times d}$. After local optimization, each client c sends the updated local state $\Phi_c^{(t+1)} = \{\Theta_c^{(t+1)}, \mathbf{V}_c^{(t+1)}\}$ back to the server for aggregation via FedAvg, formulated as:

$$\Theta^{(t+1)} \leftarrow \sum_{c \in C_t} \frac{|\mathcal{D}_c|}{\sum_c |\mathcal{D}_c|} \Theta_c^{(t+1)}, \mathbf{V}^{(t+1)} \leftarrow \sum_{c \in C_t} \frac{|\mathcal{D}_c|}{\sum_c |\mathcal{D}_c|} \mathbf{V}_c^{(t+1)}. \quad (1)$$

This means that the aggregated parameters are the weighted sum of the client-uploaded parameters in Euclidean space.

3 Theoretical Analyses of Dilemma in FedRec

3.1 Local Filtering Bubbles

We formalize this challenge as two properties, i.e., unbounded norm inflation and local representation homogeneity. Consider the most typical ranking loss (i.e., BPRLoss [29]) on client c :

$$\mathcal{L}_{\text{BPR}} = \frac{1}{\mathcal{R}} \sum_{(u, v) \in \mathcal{D}_c} -\log[\text{sigmoid}(s(u, v) - s(u, v^-))], \quad (2)$$

where $s(u, v) = f(\mathbf{u})^T f(\mathbf{v})$ means the inner product, $f(\cdot)$ is the encoder to map embeddings into a low-dimensional space, $\mathbf{u} \in \mathbf{U}_c$, $\mathbf{v} \in \mathbf{V}_c$. v^- denotes the randomly sampled negative item of the user.

PROPOSITION 3.1 (LOCAL UNBOUNDED NORM INFLATION). *In FedRec, the general contrastive objective $\mathcal{L}_{\text{BPR}}$ can be driven arbitrarily close to zero by scaling the embeddings without improving semantic representation in the absence of a hard norm constraint.*

PROPOSITION 3.2 (LOCAL REPRESENTATION HOMOGENEITY). *Due to the local supervision sparsity in FedRec, the local optimization tends to concentrate interacted item embeddings around a low-dimensional local user subspace.*

In summary³, a critical reason for the failure of aggregated item embeddings is the overconfidence of clients caused by extreme data sparsity. The local training process actively increases the embedding magnitude, making it highly noisy and unrelated to popularity. Furthermore, it concentrates interacted item embeddings into a low-dimensional subspace around user embeddings, which essentially constitutes different local filtering bubbles for clients.

3.2 Global Geometric Misalignment

Although local optimization causes embedding norm inflation, we observe item-wise energy decay on the server. We argue that this is because the naive Euclidean aggregation in FedAvg fails to preserve

³The proofs of the two propositions are in Appendix B.

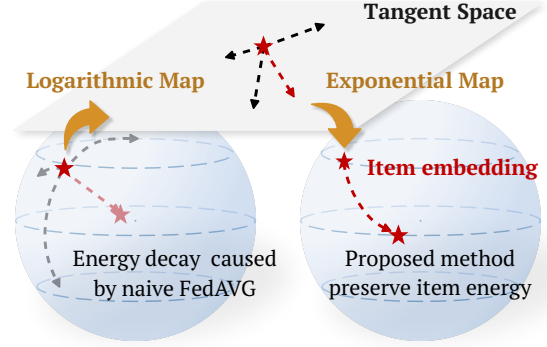


Figure 3: Aggregation via Riemannian Projection.

geometric structures in the presence of conflicts. For any item v , client c learns a local embedding $\mathbf{v}_c$. We decompose it into angular and radial pieces as $\mathbf{v}_c = r_c \mathbf{p}_c$, where $r_c = \|\mathbf{v}_c\| \geq 0$, $\|\mathbf{p}_c\| = 1$. In the server, FedAvg aggregates updates $\mathbf{v}$ from all clients as $\hat{\mathbf{v}} = \sum_c w_c \mathbf{v}_c$, where $\sum_c w_c = 1$.

ASSUMPTION 1 (DIRECTION HETEROGENEITY FACTOR). *There exists a $\rho \in [0, 1)$ for each item v , that $\mathbf{p}_c^T \mathbf{p}_{c'} \leq \rho$ for all $c \neq c'$.*

PROPOSITION 3.3 (FEDAVG SHRINKS AGGREGATED ITEM ENERGY). *The energy of the aggregated item embedding $\hat{\mathbf{v}}$ is upper bounded by the weighted energy of local embeddings, affected by the direction heterogeneity factor ρ .*

Since ρ serves as a geometrical proxy for data heterogeneity, a smaller ρ indicates greater semantic divergence among clients of the item. This proposition⁴ reveals a fundamental geometric misalignment that naive Euclidean aggregation fails to account for the relational semantics of embeddings in recommendation. It cancels out the updates of clients, rather than aggregating them, leading to observed energy decay.

4 CanonFedRec: A Canonical Geometric Framework for Personalized FedRec

Based on our observations and analyses, we propose CanonFedRec, a simple and practical framework to address challenges from a canonical geometric perspective. Following prior work [34], we adopt a Riemannian manifold $\mathcal{M} := \mathbb{S}^{d-1} = \{\mathbf{v} \in \mathbb{R}^d, \|\mathbf{v}\| = 1\}$, explicitly mitigating local noise not related to popularity from the semantic relationships. On the server, we propose a Riemannian Projection-based Aggregation mechanism that aggregates local updates in a geometrically compatible tangent space, explicitly preserving the simplicity of FedAVG. On the client, we introduce a Divergence-Aware Elastic Alignment strategy that dynamically adapts local optimization objectives via elastic decision boundaries.

4.1 Aggregation via Riemannian Projection

Although FedAVG is simple and effective, it struggles to handle item embeddings that represent semantic relationships in recommendations, leading to an inherent geometric misalignment. Therefore,

⁴The proof of the proposition is in Appendix C.

we propose a Riemannian Projection-based Aggregation mechanism that explicitly retains simplicity by implementing FedAVG in a geometrically compatible global tangent space.

Specifically, we treat the global item state $\mathbf{V}^{(t-1)}$ from the previous round $t-1$ as a semantic anchor, mapping local updates into its tangent space. However, due to heterogeneity, client embeddings $\mathbf{V}_c^{(t)}$ always drift into directionally inconsistent subspaces relative to the $\mathbf{V}^{(t-1)}$. To rectify this, we employ Orthogonal Procrustes Transformation [30] to align local updates with global anchors before aggregation, formulated as:

$$\tilde{\mathbf{V}}_c^{(t)} = \mathbf{V}_c^{(t)} \mathbf{Q}_c, \quad (3)$$

$$\mathbf{Q}_c = \arg \min_{\mathbf{Q}} \|\mathbf{V}_{\text{anch}}^{(t-1)} - \mathbf{V}_{c,\text{anch}}^{(t)} \mathbf{Q}\|_F, \quad (4)$$

where $\mathbf{Q}_c$ is the solved rotation matrix. $\mathbf{V}_{\text{anch}}^{(t-1)}$ and $\mathbf{V}_{c,\text{anch}}^{(t)}$ denote the subset of item embeddings used for alignment. Among these, the interacted items naturally serve as semantic anchor points. To avoid the ill-posed alignment issues caused by sparse interactions, we expand the anchor set by randomly sampling a subset of uninteracted items as negative anchor points. $\tilde{\mathbf{V}}_c^{(t)}$ denotes the aligned item embeddings of client c . This process ensures that all client updates are in a unified global coordinate system while preserving their inherent topological structure.

As shown in Figure 3, we utilize the Riemannian Logarithmic Map mechanism to project aligned updates onto the tangent space $T_{\mathbf{V}^{(t-1)}} \mathcal{M}$ at the global state $\mathbf{V}^{(t-1)}$, formulated as:

$$\mathbf{T}_c = \text{Log}_{\mathbf{V}^{(t-1)}}(\tilde{\mathbf{V}}_c^{(t)}). \quad (5)$$

Specifically, for each item v , the tangent vector is computed as:

$$\mathbf{T}_{c,v} = \frac{\theta}{\sin \theta} \left(\tilde{\mathbf{v}}_c^{(t)} - \mathbf{v}^{(t-1)} \cos \theta \right), \quad (6)$$

$\theta = \arccos(x^T y)$ represents the item-wise geodesic distance between $x = \mathbf{v}^{(t-1)} \in \mathbf{V}^{(t-1)}$ and $y = \tilde{\mathbf{v}}_c^{(t)} \in \tilde{\mathbf{V}}_c^{(t)}$. In the flat tangent space, FedAVG is locally valid. As $\mathbf{V}^{(t-1)}$ is the starting point in round t , the aggregated embeddings in the tangent space represent the geodesic displacement of the global state, which is denoted as:

$$\tilde{\mathbf{V}}^{\text{agg}} = \sum_{c \in \mathcal{C}_t} \frac{|\mathcal{D}_c|}{\sum_c |\mathcal{D}_c|} \mathbf{T}_c. \quad (7)$$

However, due to heterogeneity and sparsity, clients' updates do not always lie on the global manifold, leading to energy decay when projected onto the global tangent space. Hence, we employ a scaling parameter γ for compensation, which serves as the global learning rate. We then map $\gamma \tilde{\mathbf{V}}^{\text{agg}}$ back to the hypersphere via the item-wise Riemannian Exponential Map to update the global state:

$$\mathbf{V}^{(t)} = \text{Exp}_{\mathbf{V}^{(t-1)}}(\gamma \tilde{\mathbf{V}}^{\text{agg}}). \quad (8)$$

For each item v , the updated embedding is calculated as:

$$\mathbf{v}^{(t)} = \mathbf{v}^{(t-1)} \cos \|\gamma \tilde{\mathbf{v}}^{\text{agg}}\| + \frac{\tilde{\mathbf{v}}^{\text{agg}}}{\|\tilde{\mathbf{v}}^{\text{agg}}\|} \sin(\|\gamma \tilde{\mathbf{v}}^{\text{agg}}\|). \quad (9)$$

Through this approach, CanonFedRec strictly ensures geometric alignment of aggregation. While preserving the simplicity, it avoids the structural collapse caused by the naive Euclidean aggregation.

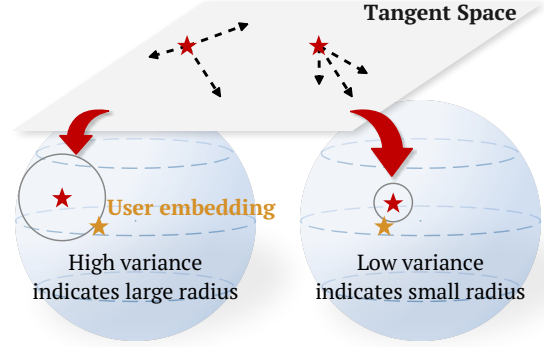


Figure 4: Divergence-Aware Elastic Alignment.

4.2 Divergence-Aware Elastic Alignment

When performing optimization on a hypersphere, the client typically enforces strict alignment of positive user-item pairs [34, 39]. However, clients could hold divergent yet valid perspectives on a specific item (e.g., a movie with polarizing reviews). Blindly enforcing strict alignment on such high-variance items forces the local recommender to compress complex semantics into a single point, erasing the inherent diversity. To address this, we propose a Divergence-Aware Elastic Alignment mechanism derived from a probabilistic perspective.

Following the contrastive paradigm on hyperspheres [34], we aim to align positive user-item pairs while maintaining representation uniformity of all embeddings on the client c :

$$\begin{aligned} \mathcal{L}_{\text{align}}(c) &= \mathbb{E}_{(u,v) \in \mathcal{D}_c} \|f(u) - f(v)\|^2 \\ \mathcal{L}_{\text{uniform}}(c) &= \log \mathbb{E}_{u,u' \in \mathcal{D}_c} e^{-2\|f(u) - f(u')\|/2} \\ &\quad + \log \mathbb{E}_{v,v' \in \mathcal{D}_c} e^{-2\|f(v) - f(v')\|/2}. \end{aligned} \quad (10)$$

To account for semantic uncertainty and divergence, we model the conditional probability of a user embedding $f(u)$ given the item embedding $f(v)$ not as a fixed point, but as a distribution. Theoretically, the von Mises-Fisher (vMF) distribution is the best choice to model the distribution on a hypersphere [1]. However, optimizing vMF directly involves complex Bessel functions, which are expensive in practice [5]. Since the tangent space $T_{\mathbf{V}^{(t-1)}} \mathcal{M}$ is a local Euclidean approximation of the manifold, we approximate the vMF distribution with a heteroscedastic Gaussian distribution in:

$$p(f(u)|f(v), \sigma_v^2) \propto \exp\left(-\frac{\|\phi_v(u)\|^2}{2\beta\sigma_v^2}\right), \quad (11)$$

where $\phi_v(u) = f(u) - f(v)$ is a local approximate tangent vector. σ_v^2 represents the global semantic divergence of item v , β is a temperature hyperparameter controlling the sensitivity to uncertainty.

As shown in Figure 4, the server quantifies the divergence σ_v^2 by analyzing the dispersion of client updates in the tangent space:

$$\sigma_v^2 = \mathbb{E}_c[\|\gamma \mathbf{T}_{c,v}\|^2] - \|\mathbb{E}_c[\gamma \mathbf{T}_{c,v}]\|^2, \quad (12)$$

where $\mathbf{T}_{c,v}$ is the updated tangent vector obtained via Eq. 6, γ is the global learning rate defined in Eq. 8. $\mathbb{E}_c[\cdot]$ denotes the weighted average of participating clients. A larger σ_v^2 indicates that there is a significant semantic divergence and uncertainty regarding the item v . This allows the server to efficiently quantify the uncertain radius

Table 1: Statistics of the datasets in this paper.

Datasets	#Users	#Items	#Interactions	Density
ML-100k	943	1,682	100,000	6.30%
LastFM	992	10,000	517,817	5.22%
BX (Book-Crossing)	18,964	19,998	482,153	0.13%
Video (Video Games)	24,303	10,672	231,780	0.09%

$R_v = 2\beta\sigma_v^2 + \epsilon$ for each item, representing the confidence region from a global perspective. $\epsilon > 0$ to prevent numerical instability. Then, the server distributes the uncertainty set $\{R_v | v \in \mathcal{V}\}$ to the client c along with the updated global state.

Formally, maximizing the log-likelihood in Eq. 11 is equivalent to minimizing the uncertainty-weighted Euclidean distance. We thus formulate the alignment loss into an Elastic Alignment Loss:

$$\mathcal{L}_{\text{align}}^{\text{elastic}}(c) = \mathbb{E}_{(u,v) \in \mathcal{D}_c} \left[\frac{\|f(u) - f(v)\|^2}{R_v} \right], \quad (13)$$

The final learning objective for client c is denoted as:

$$\mathcal{L}_{\text{rec}}(c) = \mathcal{L}_{\text{align}}^{\text{elastic}}(c) + \lambda \mathcal{L}_{\text{uniform}}(c). \quad (14)$$

Compared to $\mathcal{L}_{\text{align}}(c)$, $\mathcal{L}_{\text{align}}^{\text{elastic}}(c)$ adaptively allows more relaxed boundaries for items with divergence, effectively preventing representation homogeneity and local overconfidence.

5 Experiments

In this section, we conduct extensive experiments to answer the following research questions:

- **RQ 1:** Does CanonFedRec outperform other FedRec methods in terms of accuracy, efficiency and diversity?
- **RQ 2:** How do specific components in the CanonFedRec contribute to the performance?
- **RQ 3:** How robust is CanonFedRec under heterogeneity and noise injection?
- **RQ 4:** What is the impact of key hyperparameters?
- **RQ 5:** How does CanonFedRec perform when integrated with other inner product based encoders?

5.1 Experimental Settings

5.1.1 Datasets. We conduct experiments on four widely used real-world datasets: ML-100k [9], LastFM [17], Book-Crossing (BX) [51], and Amazon Video Games (Video) [10]. All the datasets are publicly available and widely adopted in previous studies. Following existing practice [8], we split the data into training and test sets in an 8:2 ratio, with 10% of the training set for evaluation. In experiments excluding heterogeneity, we divide the dataset among 100 clients for federated learning. Statistics of the datasets are shown in Table 1.

5.1.2 Compared Methods. We compare the performance of CanonFedRec with various superior personalized FedRec methods:

- **FedMF** [2]: Following the typical FedRec setting, clients jointly train a matrix factorization-based recommender [15].
- **FedLightGCN**: A federated version of LightGCN [11], which integrates user-item interaction data while avoiding the introduction of additional parameters.

- **FedNCF** [28]: A federated version of NeuCF [12], where clients jointly learn a shared recommender and item embeddings, while user embeddings remain locally.
- **PFedRec** [47]: It proposes a dual-personalization strategy. By asynchronously optimizing the client’s recommender and item embeddings, it achieves personalization without introducing additional parameters.
- **GPFedRec** [48]: A graph-guided personalized FedRec framework, which constructs a user-relation graph on the server and generates the global and user-specific perspective of item embeddings for each client to achieve personalization.
- **FedRAP** [20]: It decouples the global and personalized representations of item embeddings, rectifying personalization in item embeddings through an additive mechanism.
- **MetaMF** [21]: It utilizes a meta-network to generate a score function module and private item embeddings, achieving efficient client-side personalization.
- **FedCIA** [8]: It proposes a novel collaborative information aggregation method, aggregating item similarity matrices instead of raw embeddings, which allows clients to align without forcing unified item embeddings.

5.1.3 Evaluation Metrics. We evaluate performance based on both the accuracy and diversity of Top-K recommendations. For accuracy, we employ F1 [13], Mean Reciprocal Rank [33] (MRR) and Normalized Discounted Cumulative Gain [14] (NDCG) as evaluation metrics following existing practice. For diversity, we employ Intra-List Diversity [51] (ILD) and Coverage [6] as metrics. The former measures diversity by the average semantic gap between item pairs in the recommendation list, preventing duplicate content. The latter evaluates whether all items in the system have been recommended, reflecting the ability to uncover long-tail items. The higher the ILD and Coverage, the greater the diversity in the recommendation list, reducing the risk of filter bubbles. We repeat each experiment 5 times with different random seeds, and report the mean results, except in Table 2. We set $K = 10$ in this paper.

5.1.4 Implementation Details. The implementation of CanonFedRec utilizes the PyTorch library on an NVIDIA 4090 GPU. Adam is used as the default optimizer and the maximum number of epochs is set to 300. Early stop is adopted if F1@10 on the validation set continues to drop for 10 epochs. Following existing practice [8], we set the embedding size to 512. We use the implementation of LightGCN in the Recbole library [50]. We maintain the same parameter range and backbone recommender (*i.e.*, MF) for CanonFedRec and other methods for a fair comparison. As for baseline-specific hyperparameters, we tune them in the ranges suggested by the original paper. For the key hyperparameters of CanonFedRec, we search γ in $\{1, 1.5, 2, 2.5, 3\}$ and search β in the range $[0.1, 1]$.

5.2 RQ 1: Overall Performance

5.2.1 Accuracy Comparison. Table 2 reports the performance comparison of CanonFedRec with SOTA FedRec methods across four benchmark datasets, revealing the following observations: 1) CanonFedRec consistently outperforms all baselines across most metrics. Especially in ranking quality metrics (MRR@10 and NDCG@10), our approach achieves SOTA performance on all datasets. This

Table 2: The overall comparison for all baseline methods in four datasets. The best results are in bold face, and the best baselines are underlined. M@10 and N@10 denote MRR@10 and NDCG@10. The superscripts ** indicate $p \leq 0.01$ for the paired t-test of CanonFedRec v.s. the best baseline.

Metrics		FedRec w/ personalized strategy					FedRec w/o personalized strategy			
		PFedRec	GFedRec	FedRAP	MetaMF	FedCIA	FedMF	FedNCF	FedLightGCN	CanonFedRec
ML-100k	F1@10	14.51±0.43	14.01±0.19	16.19±0.25	17.76±0.62	23.29±0.18	20.71±0.05	13.08±0.33	16.93±0.43	23.06±0.09
	M@10	47.64±0.17	46.77±0.14	47.76±0.14	43.19±0.38	<u>60.14±0.07</u>	55.73±0.10	43.43±0.16	55.82±0.37	61.79±0.08**
	N@10	55.80±0.11	55.51±0.10	56.13±0.07	52.53±0.37	<u>67.58±0.15</u>	64.57±0.07	52.31±0.23	61.98±0.22	68.76±0.14**
LastFM	F1@10	4.41±0.17	4.76±0.15	4.96±0.13	5.51±0.26	<u>5.81±0.12</u>	4.02±0.06	4.39±0.18	4.68±0.54	8.55±0.09**
	M@10	33.20±1.32	41.57±0.31	43.31±0.22	34.20±0.29	45.76±0.13	34.10±0.07	40.11±0.05	<u>46.01±0.93</u>	58.17±0.36**
	N@10	46.23±0.08	49.95±0.19	50.01±0.17	46.92±0.83	53.83±0.10	43.39±0.04	48.45±0.16	<u>54.06±0.72</u>	65.33±0.41**
BX	F1@10	<u>1.30±0.12</u>	0.97±0.43	0.82±0.19	1.06±0.28	1.21±0.17	0.99±0.21	1.26±0.29	0.91±0.08	1.88±0.10**
	M@10	<u>2.84±0.19</u>	2.56±0.61	2.68±0.33	2.41±0.35	2.54±0.33	2.14±0.60	2.59±0.15	2.13±0.06	3.46±0.21**
	N@10	4.01±0.28	3.99±0.15	3.84±0.47	<u>4.14±0.26</u>	3.47±0.28	3.07±0.91	4.10±0.11	3.06±0.14	4.92±0.13**
Video	F1@10	0.91±0.28	0.74±0.20	0.86±0.14	0.54±0.19	<u>0.99±0.26</u>	0.68±0.11	0.76±0.07	0.84±0.18	2.21±0.25**
	M@10	1.25±0.36	1.10±0.24	1.21±0.15	1.25±0.29	<u>1.68±0.34</u>	1.04±0.06	1.08±0.09	1.35±0.20	3.82±0.27**
	N@10	1.99±0.37	1.72±0.51	1.83±0.18	1.68±0.23	<u>2.27±0.35</u>	1.76±0.19	1.69±0.13	1.77±0.32	5.83±0.30**

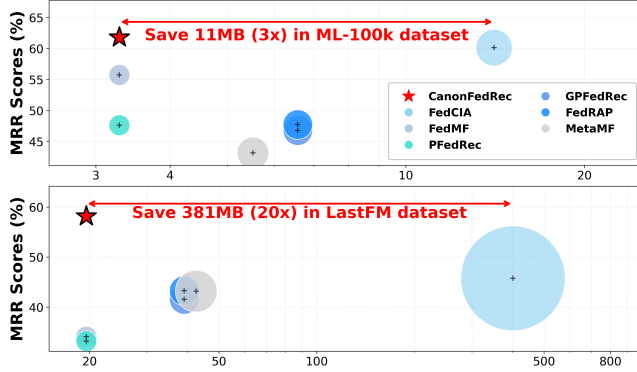


Figure 5: Performance vs. Parameter Efficiency in Log scale. The bubble size represents the communication overhead.

demonstrates that the two strategies we proposed effectively enhance the representation quality of item embeddings, achieving more accurate ranking. 2) CanonFedRec demonstrates more significant improvements on sparser LastFM and Video datasets. This indicates that as heterogeneity increases, existing methods are limited by the collapse of item embeddings, making it challenging to learn clear decision boundaries. The proposed aggregation mechanism effectively preserves the energy of aggregated item embeddings, demonstrating superior robustness. 3) On the relatively dense ML-100k dataset, although FedCIA achieves marginally higher F1@10 scores, CanonFedRec still dominates in ranking quality. This indicates that while the SOTA method can effectively retrieve target items in dense scenarios, our elastic alignment strategy enables learning fine-grained item relationships, which is particularly crucial for practical recommendations.

5.2.2 Efficiency Comparison. As shown in Figure 5, CanonFedRec dominates the Pareto frontier for storage and performance, achieving superior accuracy while requiring significantly fewer parameters compared to the baselines. Specifically, our method reduces client-side storage requirements by 3× on the ML-100k dataset, and up to 20× on the larger LastFM dataset (and 40× on the BX

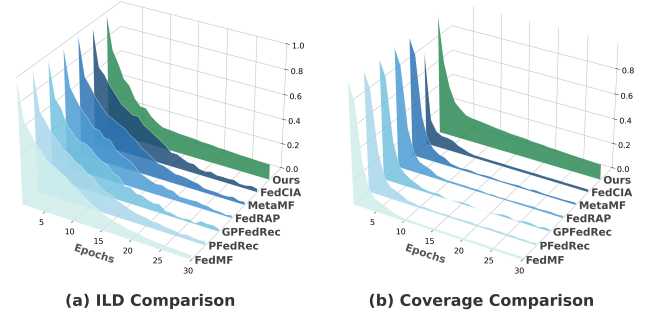


Figure 6: Performance comparison of recommendation diversity metrics (ILD and Coverage) on the ML-100k dataset.

dataset) compared to the strongest baseline FedCIA. Furthermore, the smallest bubbles indicate that CanonFedRec maintains minimal communication bandwidth, making it highly suitable for resource-constrained edge devices. The scalability bottleneck of FedCIA stems from its reliance on dense item-wise similarity matrices with $O(N^2)$ space complexity, where N denotes the scale of items. Although low-rank approximation techniques (e.g., SVD) can be employed to reduce communication costs, they fundamentally fail to alleviate the heavy storage burden. In contrast, CanonFedRec fundamentally enhances the representational capability of aggregated item embeddings by mitigating both client-side local filtering bubbles and server-side global geometric misalignment. As a result, it achieves a linear $O(Nd)$ space complexity ($d \ll N$). This makes CanonFedRec not only accurate but also a scalable solution for large-scale FedRec systems.

5.2.3 Diversity Comparison. To evaluate the ability to break filtering bubbles, we track the training dynamics of ILD and Coverage metrics on the ML-100k dataset. As shown in Figure 6, existing methods exhibit a significant collapse in these metrics during training, indicating the degeneration of item embeddings. CanonFedRec consistently maintains high ILD values throughout training, indicating

Table 3: Ablation studies (%) on key components.

HMH	RPA	DEA	ML-100k			LastFM		
			F1@10	M@10	N@10	F1@10	M@10	N@10
			20.71	55.73	64.57	4.02	34.10	43.39
	✓		22.39	59.33	66.92	7.01	51.33	59.58
✓			22.17	57.78	66.29	3.99	34.08	43.37
✓	✓		22.86	61.46	68.51	7.96	54.61	62.53
✓		✓	21.20	55.52	64.08	5.50	44.15	52.43
✓	✓	✓	23.06	61.79	68.76	8.55	58.17	65.33

that our proposed Riemannian Projection-based Aggregation mechanism successfully preserves the intrinsic topological structure of item embeddings. This ensures that aggregated embeddings retain their representational power, resulting in diverse recommendation lists. Similarly, the continuous high Coverage values demonstrate the superiority of our Divergence-Aware Elastic Alignment strategy. It effectively maintains the fine-grained relationships of item embeddings, preserving the visibility of the long-tail distribution.

5.3 RQ 2: Ablation Study

We conduct a comprehensive ablation study in Table 3, along with a visualization in Figure 7 to validate the contributions of key components: Hypersphere Manifold Hypothesis (HMH), Divergence-Aware Elastic Alignment (DEA), and Riemannian Projection-based Aggregation (RPA). Comparing the variants in Table 3 reveals that excluding RPA results in suboptimal performance. Figure 7 provides the most intuitive evidence that the embedding distribution in the methods without the tangent space aggregation mechanism exhibits severe collapse. This indicates that there is a geometric misalignment between the naive Euclidean aggregation and the item embedding space, leading to energy decay. Even with HMH to regularize the latent space, we cannot address this issue when the data is sparse. In contrast, RPA effectively preserves the global semantic structure by aggregating in geometrically compatible tangent spaces. We observe a critical dependency between DEA and RPA. Introducing DEA without RPA results in performance degradation, even worse than the variant using HMH alone in the ML-100k dataset. This is because DEA encourages clients to explore diverse semantic directions, while naive Euclidean aggregation treats these as noise. This causes disruptive interference, completely erasing the fine-grained relationships for ranking. However, when integrating DEA and RPA, the performance reaches SOTA levels, demonstrating that RPA can effectively accommodate the diverse user preferences generated by DEA. While the effective representations produced by RPA can provide DEA with a reliable source of divergence. Therefore, the fine-grained preference clusters shown in Figure 7 occur.

5.4 RQ 3: Robustness of CanonFedRec

5.4.1 Impact of Data Heterogeneity. We split the ML-100k dataset into $C = \{100, \dots, 943\}$ clients to validate the effects of different levels of heterogeneity. At $C = 100$, each client holds sufficient users and interactions, whereas at $C = 943$, each client represents a single user, thereby maximizing data heterogeneity. As shown in Figure 8, FedCIA and CanonFedRec perform comparably under mild heterogeneity settings ($C = 100$). However, as heterogeneity increases, the gap between the two widens significantly. Especially in the

Table 4: Results of privacy-enhanced CanonFedRec under various Laplacian noise strength on LastFM dataset.

Metrics	Noise Strength					
	0	0.1	0.2	0.3	0.4	0.5
F1@10	8.55	8.25	8.20	8.10	8.06	7.74
MRR@10	58.17	57.54	54.68	53.88	52.50	52.43
NDCG@10	65.33	64.84	64.50	63.31	63.06	62.55

extreme setting ($C = 943$), CanonFedRec significantly outperforms FedCIA. This highlights the limitations of naive Euclidean aggregation: even with personalized rectification mechanisms, it cannot alleviate the geometric misalignment caused by highly divergent local embeddings. In contrast, CanonFedRec demonstrates greater stability, indicating that it effectively resolves global geometric misalignment and local filtering bubble issues, thereby ensuring the representational capability of aggregated item embeddings.

5.4.2 Impact of Noise Injection. We investigate the performance of privacy-enhanced CanonFedRec using local differential privacy techniques on the LastFM dataset. We use Laplace noise and set the strength from 0.1 to 0.5 with an interval of 0.1. Given that a naive application of the standard Laplace mechanism would violate the geometric constraints, we use the Riemannian Projection mechanism to retract noisy embeddings onto the unit hypersphere. This means that all the noise is an angular perturbation. As shown in Table 4, CanonFedRec demonstrates high tolerance to perturbations. Under moderate noise strength of 0.2, CanonFedRec retains over 98% of its NDCG@10 performance. Even at higher noise levels, the performance degradation is still acceptable. This demonstrates that the proposed geometric framework inherently constrains the perturbation magnitude, exhibiting considerable robustness.

5.5 RQ 4: Hyperparameter Sensitivity Study

We validate on the ML-100k dataset to demonstrate the sensitivity of two key hyperparameters in CanonFedRec: γ and β .

5.5.1 Impact of γ . γ governs the magnitude of updates in the global space, aiming to rectify inherent energy decay caused by geometric misalignment between client-specific sub-manifolds and the global manifold. As shown in Figure 9, it exhibits a distinct trend of first increasing and then decreasing. When γ is small, the performance is suboptimal, indicating that the energy of aggregated item embeddings drops. Performance peaks at $\gamma = 2$ on ML-100k dataset, indicating that aggregated item embeddings are in their proper energy levels. As γ increases further, performance begins to degrade, indicating that γ over-amplifies the update steps, causing aggregated embeddings to exceed their target positions and disrupting the semantic relationships established during training.

5.5.2 Impact of β . β modulates the radius of the elastic decision boundary, determining the client’s tolerance for divergence. CanonFedRec exhibits significant robustness for moderate β . Within the range $\beta \in [0.1, 0.7]$, the elastic alignment mechanism successfully preserves fine-grained item relationships, mitigating client-side filtering bubbles. When $\beta \geq 0.8$, we observe a sharp performance drop. This is because a huge β leads to overly loose alignment boundaries, causing client updates to deviate from global consensus, thus losing

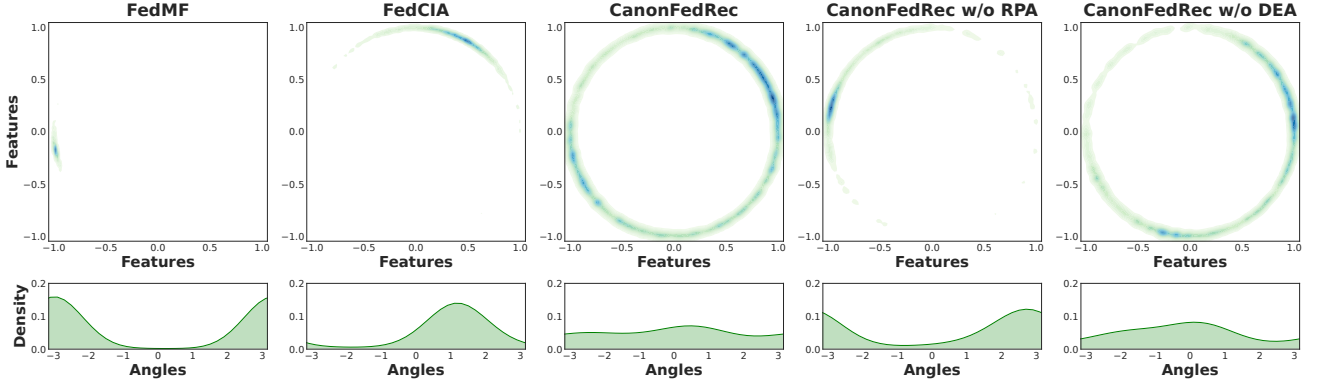


Figure 7: Distribution of interacted embeddings with Gaussian kernel density estimation [45] of the 20th client in LastFM dataset. Top: The darker the color, the more items concentrations. Bottom: The flatter the curve, the more uniform the distribution.

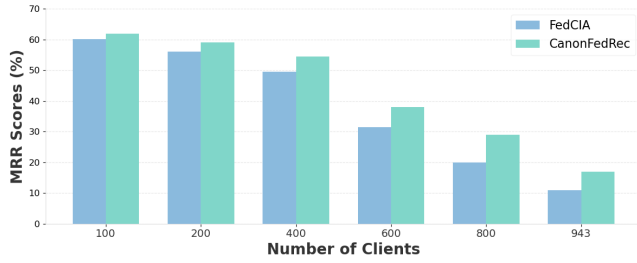


Figure 8: Performance improvement on clients of different heterogeneity levels on ML-100k dataset. The more clients, the fewer users within it, and the stronger the heterogeneity.

the collaborative benefits. This phenomenon validates that while elastic mechanisms are needed to break filtering bubbles, they must remain constrained by the global structure.

5.6 RQ 5: Integration with Other Encoders

We integrate CanonFedRec with representative inner product based encoders (*i.e.*, MF and LightGCN with different numbers of layers) to verify its architectural agnosticism on LastFM and BX datasets. FedLightGCN-X means the FedLightGCN with X GNN layers. As shown in Table 5, CanonFedRec consistently improves performance across all backbones, validating it as a practical geometric framework. We observe that the relative improvement decreases as the encoder complexity grows. This is attributed to structural redundancy: FedMF relies entirely on our design to reconstruct geometric relationships, whereas FedLightGCN inherently captures local structural information, partially overlapping with ours. Crucially, the significant gains achieved on the strongest baseline confirm that local graph is insufficient to replace the global geometric structure. Even when local structures are effectively learned, CanonFedRec still mitigates geometric misalignment and filtering bubbles.

6 Related Work

6.1 Personalized Federated Recommendation

FedRec aims to train recommenders collaboratively while keeping user data locally, thus presenting fundamental heterogeneity challenges [7, 22, 26, 32, 46, 49]. Standard aggregation methods (*i.e.*,

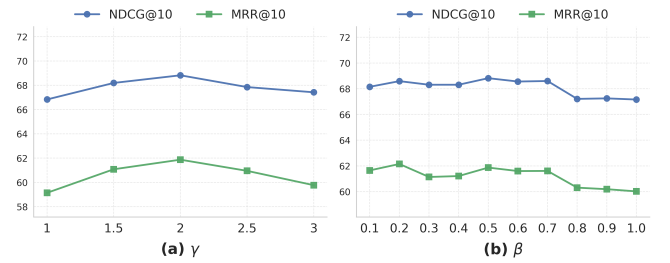


Figure 9: Impact of key hyperparameters on ML-100k dataset.

Table 5: Performance comparison of different encoders when integrated with CanonFedRec framework.

Method	LastFM			BX		
	F1@10	M@10	N@10	F1@10	M@10	N@10
FedMF	4.02	34.10	43.39	0.99	2.14	3.07
+CanonFedRec	8.68	58.01	65.57	1.85	3.35	4.87
FedLightGCN-1	4.77	45.72	53.61	0.84	2.01	2.99
+CanonFedRec	6.67	51.68	57.84	1.31	2.97	4.43
FedLightGCN-2	4.68	46.01	54.06	0.91	2.13	3.06
+CanonFedRec	6.16	50.18	57.09	1.15	2.78	4.24

FedAVG) typically produce an average representation of diverse client preferences, leading to personalization degradation [20]. To mitigate this, extensive research focuses on decoupling global and personal representations of item embeddings [20, 21, 25, 47, 48]. For instance, GPFedRec [48] proposes a graph-guided personalization framework, which constructs a user-relation graph on the server and generates a global and user-specific perspective of item embeddings for each client. Although these decoupling methods alleviate immediate conflicts arising from heterogeneity, they carry an inherent risk of exacerbating information silos, thereby undermining the collaborative benefits of federated learning. FedCIA [8] points out that the drop in personalization is due to misalignment in aggregated content and proposes aggregating item-wise similarity matrices rather than raw embeddings. This allows clients to share their preference manifolds without enforcing rigid alignment of embeddings. Despite achieving SOTA performance, it trades off precision for storage in resource-constrained clients, reducing its practical usability. Hence, we empirically and theoretically analyze

the fundamental geometric misalignment inherent in Euclidean-space aggregation and bridge this gap using Riemannian geometry.

6.2 Manifold Learning in Recommendation

Traditional recommender systems usually embed users and items in a flat Euclidean space [37, 40]. However, this assumption results in geometric mismatches when modeling scale-free or hierarchical real-world interactions, thereby degrading representation. To mitigate this, pioneering research explores Hyperbolic manifolds [3, 38], leveraging their exponential expansion capacity to capture hierarchies and tail distributions, significantly improving cold start and long-tail recommendation capabilities. However, hyperbolic embedding requires expensive Riemannian optimization and is mathematically incompatible with the standard inner-product models of recommenders, posing significant engineering challenges in practice. Therefore, recent research adopts hyperspheres as a compromise solution [34, 35, 44]. For example, DirectAU [34] shows that optimizing alignment and uniformity on the hypersphere can more effectively capture semantic structure than a complex hyperbolic manifold while maintaining compatibility with inner-product models. However, enforcing alignment in FedRec poses challenges due to data sparsity and heterogeneity. This inherently undermines diverse user interests, creating filtering bubbles and diminishing the system’s ability to represent long-tail content, which motivates us to design an elastic alignment strategy on hyperspherical manifolds.

7 Conclusion

In this work, we empirically and theoretically analyze the failure of aggregated item embeddings, which stems from local filtering bubbles and global geometric misalignment. We propose CanonFedRec, a canonical geometric framework to address these challenges. By using a Riemannian Projection-based Aggregation mechanism on the server and implementing Divergence-Aware Elastic Alignment strategy on clients, CanonFedRec effectively avoids the collapse of preference diversity while maintaining global topological structure. Extensive experimental results demonstrate that compared to SOTA approaches, CanonFedRec significantly improves recommendation quality while reducing memory cost by up to 40×, thereby breaking the trilemma of accuracy-efficiency-diversity in FedRec.

Acknowledgments

This work is supported by National Natural Science Foundation of China (No. 62372364) and Technical Innovation Guidance Plan of Shaanxi Province, China (No. 2024QCY-KXJ-199).

References

- [1] Arindam Banerjee, Inderjit S Dhillon, Joydeep Ghosh, Suvrit Sra, and Greg Ridgeway. 2005. Clustering on the Unit Hypersphere using von Mises-Fisher Distributions. *JMLR* 6, 9 (2005).
- [2] Di Chai, Leye Wang, Kai Chen, and Qiang Yang. 2020. Secure federated matrix factorization. *IEEE Intelligent Systems* 36, 5 (2020), 11–20.
- [3] Yong Chen, Li Li, Wei Peng, and Songzhi Su. 2025. VoRec: Enhancing Recommendation with Voronoi Diagram in Hyperbolic Space. In *SIGIR*. 2081–2091.
- [4] Sen Cui, Jian Liang, Weishen Pan, Kun Chen, Changshui Zhang, and Fei Wang. 2022. Collaboration equilibrium in federated learning. In *SIGKDD*. 241–251.
- [5] Tim R Davidson, Luca Falorsi, Nicola De Cao, Thomas Kipf, and Jakub M Tomczak. 2018. Hyperspherical variational auto-encoders. *UAI* (2018).
- [6] Mouzhi Ge, Carla Delgado-Battenfeld, and Dietmar Jannach. 2010. Beyond accuracy: evaluating recommender systems by coverage and serendipity. In *RecSys*. 257–260.
- [7] Xiaoqiang Gui, Bowen Chen, Qiaoyu Tan, Jun Wang, Yongqing Zheng, Qingzhong Li, Lizhen Cui, and Guoxian Yu. 2026. Federated Recommendation via Stochastic Aggregation and Consistency Inference. *TOIS* (2026).
- [8] Mingzhe Han, Dongsheng Li, Jiafeng Xia, Jiahao Liu, Hansu Gu, Peng Zhang, Ning Gu, and Tun Lu. 2025. FedCIA: Federated collaborative information aggregation for privacy-preserving recommendation. In *SIGIR*. 1687–1696.
- [9] F Maxwell Harper and Joseph A Konstan. 2015. The movielens datasets: History and context. *TIIS* 5, 4 (2015), 1–19.
- [10] Ruining He and Julian McAuley. 2016. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *ACM WWW*. 507–517.
- [11] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, and Meng Wang. 2020. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *SIGIR*. 639–648.
- [12] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *ACM WWW*. 173–182.
- [13] Jonathan L Herlocker, Joseph A Konstan, Loren G Terveen, and John T Riedl. 2004. Evaluating collaborative filtering recommender systems. *TOIS* 22, 1 (2004), 5–53.
- [14] Kalervo Järvelin and Jaana Kekäläinen. 2002. Cumulated gain-based evaluation of IR techniques. *TOIS* 20, 4 (2002), 422–446.
- [15] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *Computer* 42, 8 (2009), 30–37.
- [16] Weirui Kuang, Bingchen Qian, Zitao Li, Daoyuan Chen, Dawei Gao, Xuchen Pan, Yuexiang Xie, Yaliang Li, Bolin Ding, and Jingren Zhou. 2024. Federatedscope-llm: A comprehensive package for fine-tuning large language models in federated learning. In *SIGKDD*. 5260–5271.
- [17] Mark Levy and Klaas Bosteele. 2010. Music recommendation and the long tail. In *WOMRAD*. 55–58.
- [18] Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. 2020. Federated optimization in heterogeneous networks. *MISys* 2 (2020), 429–450.
- [19] Zhenyang Li, Yancheng Dong, Chen Gao, Yizhou Zhao, Dong Li, Jianye Hao, Kai Zhang, Yong Li, and Zhi Wang. 2023. Breaking filter bubble: A reinforcement learning framework of controllable recommender system. In *ACM WWW*. 4041–4049.
- [20] Zhiwei Li, Guodong Long, and Tianyi Zhou. 2024. Federated recommendation with additive personalization. *ICLR* (2024).
- [21] Yujie Lin, Pengjie Ren, Zhumin Chen, Zhaochun Ren, Dongxiao Yu, Jun Ma, Maarten de Rijke, and Xiuzhen Cheng. 2020. Meta matrix factorization for federated rating predictions. In *SIGIR*. 981–990.
- [22] Zhiwei Liu, Liangwei Yang, Ziwei Fan, Hao Peng, and Philip S Yu. 2022. Federated social recommendation with graph neural network. *TIST* 13, 4 (2022), 1–24.
- [23] Zhongzhou Liu, Hao Zhang, Kuicai Dong, and Yuan Fang. 2024. Collaborative cross-modal fusion with large language model for recommendation. In *CIKM*. 1565–1574.
- [24] Donald Loveland, Mingxuan Ju, Tong Zhao, Neil Shah, and Danaï Koutra. 2025. On the Role of Weight Decay in Collaborative Filtering: A Popularity Perspective. In *SIGKDD*. 1975–1986.
- [25] Sichun Luo, Yuanzhang Xiao, Xinyi Zhang, Yang Liu, Wenbo Ding, and Linqi Song. 2024. Perfedrec++: Enhancing personalized federated recommendation with self-supervised pre-training. *TIST* 15, 5 (2024), 1–24.
- [26] Kiwan Maeng, Haiyu Lu, Luca Melis, John Nguyen, Mike Rabbat, and Carole-Jean Wu. 2022. Towards fair federated recommendation learning: Characterizing the inter-dependence of system and data heterogeneity. In *RecSys*. 156–167.
- [27] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*. PMLR, 1273–1282.
- [28] Vasileios Perifanis and Pavlos S Efraimidis. 2022. Federated neural collaborative filtering. *KBS* 242 (2022), 108441.
- [29] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. *AUAI Press* (2009), 452–461.
- [30] Peter H Schönemann. 1966. A generalized solution of the orthogonal procrustes problem. *Psychometrika* 31, 1 (1966), 1–10.
- [31] Rong Shan, Jianghao Lin, Chenxu Zhu, Bo Chen, Menghui Zhu, Kangning Zhang, Jieming Zhu, Ruiming Tang, Yong Yu, and Weinan Zhang. 2025. An automatic graph construction framework based on large language models for recommendation. In *SIGKDD*. 4806–4817.
- [32] Zehua Sun, Yonghui Xu, Yong Liu, Wei He, Lanju Kong, Fangzhao Wu, Yali Jiang, and Lizhen Cui. 2024. A survey on federated recommendation systems. *TNNLS* 36, 1 (2024), 6–20.
- [33] Ellen M Voorhees et al. 1999. The trec-8 question answering track report.. In *Trec*, Vol. 99. 77–82.
- [34] Chenyang Wang, Yuanqing Yu, Weizhi Ma, Min Zhang, Chong Chen, Yiqun Liu, and Shaoping Ma. 2022. Towards representation alignment and uniformity in collaborative filtering. In *SIGKDD*. 1816–1825.

- [35] Liang Wang, Xiang Tao, Qiang Liu, and Shu Wu. 2024. Rethinking graph masked autoencoders through alignment and uniformity. In *AAAI*, Vol. 38. 15528–15536.
- [36] Lu Wang, Di Zhang, Fangkai Yang, Pu Zhao, Jianfeng Liu, Yuefeng Zhan, Hao Sun, Qingwei Lin, Weiwei Deng, Dongmei Zhang, et al. 2025. Lettingo: Explore user profile generation for recommendation system. In *SIGKDD*. 2985–2995.
- [37] Maolin Wang, Yutian Xiao, Binhao Wang, Sheng Zhang, Shanshan Ye, Wanyu Wang, Hongzhi Yin, Ruocheng Guo, and Zenglin Xu. 2025. FindRec: Stein-Guided Entropic Flow for Multi-Modal Sequential Recommendation. In *SIGKDD*. 3008–3018.
- [38] Shicheng Wang, Shu Guo, Lihong Wang, Tingwen Liu, and Hongbo Xu. 2023. Hdnr: A hyperbolic-based debiased approach for personalized news recommendation. In *SIGIR*. 259–268.
- [39] Tongzhou Wang and Phillip Isola. 2020. Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In *ICML*. PMLR, 9929–9939.
- [40] Yi Wen, Yue Liu, Derong Xu, Huishi Luo, Pengyue Jia, Yiqing Wu, Siwei Wang, Ke Liang, Maolin Wang, Yiqi Wang, et al. 2025. Measure Domain’s Gap: A Similar Domain Selection Principle for Multi-Domain Recommendation. In *SIGKDD*. 3156–3167.
- [41] Binrui Wu, Shisong Tang, Fan Li, Bing Han, Chang Meng, Jingyu Xiao, and Jiechao Gao. 2025. Aligning and Balancing ID and Multimodal Representations for Recommendation. In *SIGKDD*. 5029–5038.
- [42] Yunjia Xi, Muyan Weng, Wen Chen, Chao Yi, Dian Chen, Gaoyang Guo, Mao Zhang, Jian Wu, Yuning Jiang, Qingwen Liu, et al. 2025. Bursting filter bubble: Enhancing serendipity recommendations with aligned large language models. In *SIGKDD*. 5059–5070.
- [43] Kang Yang, Ruiyun Yu, Bingyang Guo, and Jie Li. 2024. Interaction subgraph sequential topology-aware network for transferable recommendation. *TKDE* 36, 10 (2024), 5221–5233.
- [44] Liangwei Yang, Zhiwei Liu, Chen Wang, Mingdai Yang, Xiaolong Liu, Jing Ma, and Philip S Yu. 2023. Graph-based alignment and uniformity for recommendation. In *CIKM*. 4395–4399.
- [45] Junliang Yu, Hongzhi Yin, Xin Xia, Tong Chen, Lizhen Cui, and Quoc Viet Hung Nguyen. 2022. Are graph augmentations necessary? simple graph contrastive learning for recommendation. In *SIGIR*. 1294–1303.
- [46] Wei Yuan, Hongzhi Yin, Fangzhao Wu, Shijie Zhang, Tiek He, and Hao Wang. 2023. Federated unlearning for on-device recommendation. In *WSDM*. 393–401.
- [47] Chunxu Zhang, Guodong Long, Tianyi Zhou, Peng Yan, Zijian Zhang, Chengqi Zhang, and Bo Yang. 2023. Dual personalization on federated recommendation. *IJCAI* (2023).
- [48] Chunxu Zhang, Guodong Long, Tianyi Zhou, Zijian Zhang, Peng Yan, and Bo Yang. 2024. Gpfedrec: Graph-guided personalization for federated recommendation. In *SIGKDD*. 4131–4142.
- [49] Honglei Zhang, Haoxuan Li, Jundong Chen, Sen Cui, Kunda Yan, Abudukelimu Wuerkaixi, Xin Zhou, Zhiqi Shen, and Yidong Li. 2026. Beyond similarity: Personalized federated recommendation with composite aggregation. *TOIS* (2026), 1–28.
- [50] Wayne Xin Zhao, Shanlei Mu, Yupeng Hou, Zihan Lin, Yushuo Chen, Xingyu Pan, Kaiyuan Li, Yujie Lu, Hui Wang, Changxin Tian, et al. 2021. Recbole: Towards a unified, comprehensive and efficient framework for recommendation algorithms. In *CIKM*. 4653–4664.
- [51] Cai-Nicolas Ziegler, Sean M McNee, Joseph A Konstan, and Georg Lausen. 2005. Improving recommendation lists through topic diversification. In *ACM WWW*. 22–32.

A Empirical Verification

To empirically verify our concerns regarding overconfidence and geometric collapse, we choose MF as the backbone recommender, training it under centralized and federated settings on the ML-100k dataset. As shown in Figure 2, we observe three phenomena:

Local Unbounded Norm Inflation. As shown on the left, the norm of client’s interacted item embeddings in FedMF exhibits a rapid and unbounded increase. This phenomenon lies in the client attempts to minimize local training loss by increasing the embedding norm of positive user-item pairs, leading to severe overfitting.

Local Representation Homogeneity. Then, we study the item-wise representation homogeneity, defined as the average pairwise cosine similarity of all interacting items within a client. As shown on the mid, this metric rapidly increases in FedRec, showing that

the client is too confident to force all interactions to become geometrically close, leading to the loss of fine-grained preferences.

Global Item-Wise Energy Decay. Inspired by the wave interference phenomenon, we introduce a novel metric to explicitly measure the consistency of updates between clients, *i.e.*, aggregation energy ratio. It is denoted as the ratio of the item-wise global aggregated norm to the sum of local update norms, formulated as:

$$\text{EnergyRatio}(\mathbf{v}) = \frac{\|\sum_c w_c \mathbf{v}_c\|^2}{\sum_c w_c \|\mathbf{v}_c\|^2}, \quad w_c = \frac{|\mathcal{D}_c|}{\sum_{i \in C_t} |\mathcal{D}_i|}. \quad (15)$$

This metric exhibits significant decrease during training, revealing severe geometric directional conflicts between clients. In direct Euclidean averaging, clients’ uploaded item energies are largely canceled out, causing aggregated item embeddings ineffective.

These observations provide strong empirical evidence that standard FedAVG is geometrically incompatible with the item latent space, leading to the failure of aggregated embeddings.

B Proof of Local Filtering Bubbles

B.1 Proof of Local Unbounded Norm Inflation

PROPOSITION B.1 (LOCAL UNBOUNDED NORM INFLATION). *In FedRec, the general contrastive objective $\mathcal{L}_{\text{BPR}}$ can be driven arbitrarily close to zero by scaling the embeddings without improving semantic representation in the absence of a hard norm constraint.*

PROOF. For any triplet (u, v, v^-) , define the pairwise margin $\Delta s(u, v, v^-) = s(u, v) - s(u, v^-) = f(\mathbf{u})^T (f(\mathbf{v}) - f(\mathbf{v}'))$. There exists a margin $\gamma > 0$ for all sampled triplets, $\Delta s \geq \gamma > 0$.

Scaling embeddings by $f(\mathbf{u}) \leftarrow \alpha f(\mathbf{u})$ and $f(\mathbf{v}) \leftarrow \alpha f(\mathbf{v})$. The margin becomes:

$$\Delta s_\alpha(u, v, v^-) = \alpha^2 \Delta s(u, v, v^-) \geq \alpha^2 \gamma. \quad (16)$$

As $\mathcal{L}_{\text{BPR}}$ is a strictly decreasing function, and $\mathcal{L}_{\text{BPR}}(x) \rightarrow 0$ as $x \rightarrow \infty$. Therefore, for every sampled triplet,

$$\mathcal{L}_{\text{BPR}}(\Delta s_\alpha(u, v, v^-)) \leq \mathcal{L}_{\text{BPR}}(\alpha^2 \gamma), \quad (17)$$

where $\mathcal{L}_{\text{BPR}} \rightarrow 0$ when $\alpha \rightarrow \infty$. Soft norm regularization (*e.g.*, weight decay) can only limit the optimum to a finite but still very large norm when the regularization is weak, eliminating the short-cut typically requires a strong penalty that trades off personalization. Hence, in the absence of a hard norm constraint, general contrastive objective can be driven arbitrarily close to zero by scaling embeddings. $\square$

B.2 Proof of Local Representation Homogeneity

PROPOSITION B.2 (LOCAL REPRESENTATION HOMOGENEITY). *Due to the local supervision sparsity in FedRec, the local optimization tends to concentrate interacted item embeddings around a low-dimensional local user subspace.*

PROOF. Consider a client c with a set of local user embeddings $\mathbf{U}_c \in \mathbb{R}^{M_c \times d}$, we define the local latent space as:

$$S_c := \text{span}(\mathbf{U}_c), \quad r_c := \dim(S_c) \leq \min(M_c, d). \quad (18)$$

In a typical FedRec, each client usually contains only a small fraction of users and observes highly sparse interactions, while the embedding dimension d is moderately large. Hence, we typically find that $M_c \ll d$ and consequently $r_c \leq M_c \ll d$. Due to local

supervision sparsity, the update of a positive item v at local epoch t can be formulated as:

$$\mathbf{v}^{(t+1)} = \mathbf{v}^{(t)} + \eta g_v^{(t)}, \quad g_v^{(t)} = \sum_{k=1}^{|U_c|} a_{vk}^{(t)} \mathbf{u}_k + \xi_v^{(t)}, \quad (19)$$

where η denotes the local learning rate, and $g_v^{(t)}$ is the gradient of the item. $a_{vk}^{(t)} \in \mathbb{R}$ are alignment coefficients, and $\xi_v^{(t)}$ is a residual term from negative sampling, regularization, etc.

We decompose each item embedding into components parallel and orthogonal to S_c , formulated as:

$$\mathbf{v}^{(t)} = \mathbf{p}^{(t)} + \mathbf{q}^{(t)}, \quad \mathbf{p}^{(t)} \in S_c, \quad \mathbf{q}^{(t)} \in S_c^\perp. \quad (20)$$

Hence, the update of these components can be projected as:

$$\begin{aligned} \mathbf{p}^{(t+1)} &= \mathbf{p}^{(t)} + \eta \left(\sum_{k=1}^{|U_c|} a_{vk}^{(t)} \mathbf{u}_k + \Pi_{S_c} \xi_v^{(t)} \right), \\ \mathbf{q}^{(t+1)} &= \mathbf{q}^{(t)} + \eta \Pi_{S_c^\perp} \xi_v^{(t)}. \end{aligned} \quad (21)$$

Since the orthogonal component $\mathbf{q}^{(t)}$ is only driven by the projected residual, while the parallel component $\mathbf{p}^{(t)}$ accumulates the structured signal. The optimization process tends to increase the relative dominance of $\mathbf{p}^{(t)}$ over $\mathbf{q}^{(t)}$.

Hence, the angle between $\mathbf{v}^{(t)}$ and S_c satisfies:

$$\angle(\mathbf{v}^{(t)}, S_c) = \arcsin \left(\frac{\|\mathbf{q}^{(t)}\|}{\|\mathbf{v}^{(t)}\|} \right) \approx \arcsin \left(\frac{\|\mathbf{q}^{(t)}\|}{\|\mathbf{p}^{(t)}\|} \right), \quad (22)$$

which tends to be small as $\mathbf{p}^{(t)}$ dominates $\mathbf{q}^{(t)}$, especially under weight decay. Therefore, interacted item embeddings concentrate in a narrow hypercone around the local user subspace S_c , yielding local representation homogeneity. $\square$

C Proof of Global Energy Decay

PROPOSITION C.1 (FEDAVG SHRINKS AGGREGATED ITEM ENERGY). *The energy of the aggregated item embedding $\hat{\mathbf{v}}$ is upper bounded by the weighted energy of local embeddings, affected by the direction heterogeneity factor ρ .*

PROOF. The energy of the aggregated item can be denoted as:

$$\begin{aligned} \|\hat{\mathbf{v}}\|^2 &= \left\| \sum_c w_c \mathbf{v}_c \right\|^2 = \left\| \sum_c w_c r_c \mathbf{p}_c \right\|^2 \\ &= \sum_c w_c^2 r_c^2 + \sum_{c \neq c'} w_c w_{c'} r_c r_{c'} \mathbf{p}_c^T \mathbf{p}_{c'}. \end{aligned} \quad (23)$$

Since $\mathbf{p}_c^T \mathbf{p}_{c'} \leq \rho$ for all $c \neq c'$ by the definition of ρ , it can be denoted as:

$$\|\hat{\mathbf{v}}\|^2 \leq \sum_c w_c^2 r_c^2 + \rho \sum_{c \neq c'} w_c w_{c'} r_c r_{c'}. \quad (24)$$

Noting that:

$$\sum_{c \neq c'} w_c w_{c'} r_c r_{c'} = \left(\sum_c w_c r_c \right)^2 - \sum_c w_c^2 r_c^2, \quad (25)$$

The energy of item can be denoted as:

$$\begin{aligned} \|\hat{\mathbf{v}}\|^2 &\leq \sum_c w_c^2 r_c^2 + \rho \left(\left(\sum_c w_c r_c \right)^2 - \sum_c w_c^2 r_c^2 \right) \\ &= (1 - \rho) \sum_c w_c^2 r_c^2 + \rho \left(\sum_c w_c r_c \right)^2. \end{aligned} \quad (26)$$

D Direction Heterogeneity across Clients

Assumption 1 posits that a direction heterogeneity factor $\rho \in [0, 1)$ exists among clients' item embeddings. To empirically verify this, we measure the mean inter-client pairwise cosine similarity (%) of item embeddings during FedMF training across all four datasets and report the dynamics in Figure 10.

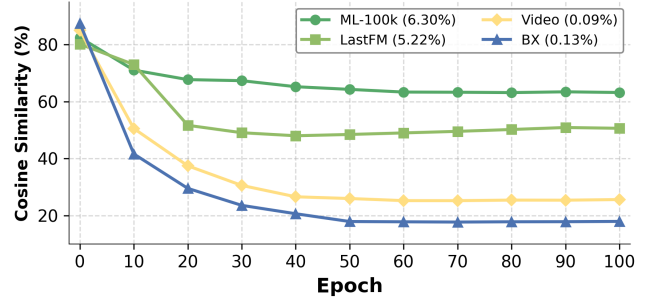


Figure 10: Mean inter-client pairwise cosine similarity (%) of item embeddings during FedMF training on four datasets.

All datasets show substantial divergence during training, validating our assumption. In addition, sparser datasets exhibit stronger divergence, which is consistent with our analysis in Proposition 3.3.

E Additional Analyses and Ablations

E.1 Anchor Sampling Strategy in Eq.(4)

In Eq.(4) interacted items serve as primary anchors. However, on sparse datasets, clients have very limited interactions (often fewer than dimension d), making the Procrustes SVD rank-deficient and the problem ill-posed: the rotation matrix $\mathbf{Q}_c$ becomes numerically unstable across rounds, causing training oscillation.

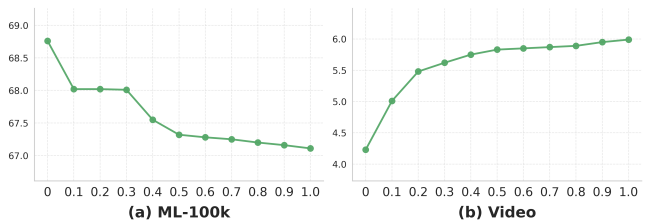


Figure 11: Ablation study on sampling ratio (%) on ML-100k and Video datasets, reporting N@10.

Hence, we need to expand the anchor set from the uninteracted items, making $\mathbf{Q}_c$ well-posed and stabilizing convergence. Since popularity-based methods depend on item frequency information, which is inherently unreliable in FedRec, where clients typically inject pseudo items to protect privacy, corrupting the observed popularity statistics on the server. In this paper, we adopt the uniform strategy of uniformly sampling a subset of uninteracted items as

negative anchors. As shown in Figure 11, the ratio is a density-dependent hyperparameter. On dense ML-100k dataset, interacted items alone suffice, and excessive sampling dilutes the alignment signal. On sparse Video dataset, additional anchors help stabilize Q_c , improving convergence. Specifically, Figure 11(b) shows that an appropriate negative sampling ratio in sparse datasets can quickly compensate the rank deficit of the Procrustes problem, then stabilizes once directional coverage saturates. Crucially, even at 0%, CanonFedRec on Video (4.23) already outperforms the best baseline FedCIA (2.27 in Table 2) by 86%, confirming core mechanisms are effective regardless of sampling.

E.2 Necessity of RPA Module

To validate the necessity of the RPA module, we compare it with simpler strategies on the same HMH+DEA base, including: 1) Cosine Constraint: cosine regularization to enforce directional consistency. 2) FedProx [18]: L2 proximal term to penalize deviation. Both use Euclidean averaging for aggregation.

Table 6: Ablation study on ML-100k and LastFM datasets.

Method	ML-100k			LastFM		
	F1@10	M@10	N@10	F1@10	M@10	N@10
Base (HMH+DEA)	22.86	61.46	68.51	7.96	54.61	62.53
+ Cosine Cons.	22.06	58.47	65.73	6.52	53.26	59.99
+ FedProx	22.49	61.55	67.52	7.33	55.58	62.17
+ RPA (Ours)	23.06	61.79	68.76	8.55	58.17	65.33

As shown in Table 6, neither resolves the misalignment in Proposition 3.3. The cosine constraint degrades on both datasets, as forcing directional alignment in Euclidean space ignores manifold structure, disrupting learned topological relationships. FedProx yields marginal benefits on the dense ML-100k dataset but degrades on the sparser LastFM dataset, indicating that L2 proximal regularization alone cannot reconcile the geometric misalignment. RPA consistently achieves the best performance on both datasets by performing alignment in the geometrically compatible tangent space, confirming that the misalignment identified in Proposition 3.3 requires a Riemannian rather than a Euclidean-space correction.

F Deployment Feasibility

To validate the deployment feasibility of CanonFedRec, we further compare it against the simplest (FedMF) and the strongest (FedCIA) baselines from both client-side and server-side on the LastFM dataset ($N=10,000$, $d=512$).

F.1 Client-Side Overhead

We report storage, communication costs per federated communication round, training time and FLOPs per local epoch in Table 7. It is clear that CanonFedRec introduces only an additional $O(N)$ term over FedMF for distributing the uncertainty radius $\{R_o\}$ from Eq. (12). Since $N \ll Nd$, the storage and communication costs are essentially identical to FedMF. In terms of local training time, CanonFedRec is comparable to FedMF and $21\times$ faster than FedCIA. In addition, the FLOPs of CanonFedRec are $155\times$ fewer than FedCIA, which must fine-tune local embeddings first. Our client-side FLOPs are well within the capacity of modern mobile SoCs. Combined with the accuracy and storage comparison in Table 2 and Figure 5, CanonFedRec achieves the best accuracy-efficiency trade-off among all compared methods.

Table 7: Client-side overhead comparison on LastFM dataset.

Method	FedMF	FedCIA	CanonFedRec
Storage	Nd	$N^2 + Nd$	$Nd + N$
Comm.	$2Nd$	$2N^2$	$2Nd + N$
Time (s)	0.10	2.54	0.12
FLOPs	11.05M	30.82T	198.96G

F.2 Server-Side Overhead

We further decompose the server-side FLOPs per communication round in Table 8.

Table 8: Server-side FLOPs per round on LastFM dataset.

Component	FLOPs
Variance computation in Eq. (12)	19.53M
Tangent-space FedAVG in Eq. (7)	488.28M
Others (Log/Exp Maps, Procrustes)	1.08T

The variance computation in Eq. (12), which enables the Divergence-Aware Elastic Alignment, accounts for less than 0.002% of the total server cost. It involves only element-wise operations that scale linearly with the item count N , introducing no numerically sensitive operations. The dominant cost comes from the Logarithmic/Exponential Maps and Procrustes alignment, which are inherent to the Riemannian Projection-based Aggregation. Notably, the entire server-side cost per round is far below FedCIA’s per-client cost, confirming CanonFedRec’s efficiency.