LOTUS-UCC: UNITARY SEMANTIC COCYCLES FOR SINGLE-MODEL SEQUENTIAL RECOMMENDATION

Anonymous authors
Paper under double-blind review

ABSTRACT

LLM-enhanced recommendation has improved semantic representation, but largely leaves semantic evolution unspecified: knowing what an item means does not explain how an interaction moves intent or when that motion changes a ranking decision. We introduce LOTUS-UCC (Language-Oriented Transport with Unitary Semantic Cocycles), a single-model semantics-to-decision dynamical system in which language defines the coordinate space of user intent. Causal spherical forecasting and persistent particle transport maintain multiple latent semantic states; a unitary cocycle carries the resulting representation without distorting its geometry; and realification produces one dense full-catalog score. Our analysis follows perturbations through the complete chain from semantic history and persistent state to particles, semantic endpoints, catalog scores, and Top-K decisions. It also quantifies the curvature error of Euclidean updates and proves that CNDB upper-controls cutoff failure risk. Under chronological leave-one-out and exhaustive ranking, LOTUS-UCC establishes new best results on 19 of 20 metrics across Beauty, Sports, Toys, Yelp, and Online Retail, with a 25.08% average relative gain over the strongest published result in each setting. Structural ablations provide evidence for the geometric and decision-aware mechanisms, and the proposed dynamics generalize consistently across four language representation spaces.

1 INTRODUCTION

Static semantic recommenders answer what an item means, while behavioral sequence models estimate what tends to follow (Hidasi et al., 2016; Kang & McAuley, 2018; Sun et al., 2019; Zhou et al., 2022). Neither explicitly represents the missing object between them: how an observed interaction moves user intent through a language-defined semantic space. Treating language as time-invariant side information cannot express this motion, and modeling transitions without semantic direction cannot explain it. Moreover, motion is useful only when it has controlled consequences for candidate membership and discounted order under exhaustive ranking. Sequential recommendation is therefore not merely semantic matching; it is semantic transport closed at a decision boundary.

Given a semantic coordinate system induced by a language model, sequential decision making requires four conditions. **C1: Causal evolution.** Every future state depends only on the observed prefix. **C2: Manifold consistency.** Updates respect the geometry of the language-induced state space. **C3: Controlled composition.** Repeated event-conditioned transitions admit explicit perturbation propagation bounds over time. **C4: Decision closure.** State variation admits explicit consequences at score and Top-K boundaries. Together, C1–C4 turn language from an auxiliary feature source into a coordinate system for a controlled semantics-to-decision process.

LOTUS-UCC realizes these conditions in one model. Language defines a spherical base space; causal transport and persistent particles maintain multiple latent semantic states; and an event-conditioned unitary cocycle evolves a complex recommendation fiber along that semantic path. A real direct-sum query scores the complete catalog, while CNDB targets operational cutoffs. Our analysis follows the complete propagation from semantic history and persistent particle states through latent dynamics to catalog scores and discrete Top-K decisions. Across five benchmarks, LOTUS-UCC achieves new best results on 19 of 20 HR/NDCG metrics, and structural, cross-LLM, stability, and efficiency studies support the complete chain.

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.

Our contributions are:

- We formulate sequential recommendation as stable semantic evolution satisfying causal, manifold, compositional, and decision-closure conditions.
- We develop LOTUS-UCC, combining a language-defined manifold, persistent multi-hypothesis particle transport, a unitary cocycle, and one dense catalog scorer.
- We establish end-to-end stability from semantic history and persistent state to score variation and Top-K invariance, together with curvature and cutoff-risk results connecting spherical transport and CNDB to their structural ablations and ranking metrics.
- Across five benchmarks, LOTUS-UCC achieves SOTA performance on 19 of 20 metrics, supported by component, structural, cross-LLM, stability, and efficiency studies.

2 Related Work

Sequential recommendation. Recurrent, attention, bidirectional, and frequency-domain models capture different views of transition structure (Hidasi et al., 2016; Kang & McAuley, 2018; Sun et al., 2019; Zhou et al., 2022). Universal pre-training transfers sequence representations across domains (Hou et al., 2022); text-driven models align language and collaborative spaces (Li et al., 2023); and large transducers scale recommendation sequence modeling (Zhai et al., 2024). MPT pre-trains on synthetic Markov chains and adapts the resulting transition prior to recommendation (Xu et al., 2026). WPGRec injects graph structure into aligned wavelet-packet bands (Liu et al., 2026), while IOCLRec learns coarse- and fine-grained intent through contrastive objectives (Wang et al., 2025). LOTUS-UCC takes a dynamical view: language defines the geometry in which intent evolves, and sequential recommendation becomes stable semantic transport through this state space.

Language representations for recommendation. Language features can improve transfer and item understanding, but static fusion does not specify how semantics should evolve across interactions. We use Qwen3-Embedding-0.6B (Yang et al., 2025) as the main semantic encoder and evaluate the same mechanism with BGE-M3 (Chen et al., 2024), E5 (Wang et al., 2022), and TinyLlama (Zhang et al., 2024).

Geometry and rank learning. Riemannian exponential maps provide intrinsic updates on spheres (Absil et al., 2008); Householder transformations are exact orthogonal maps (Golub & Van Loan, 2013); and unitary dynamics retain norms under repeated composition (Arjovsky et al., 2016). Pairwise ranking objectives such as BPR (Rendle et al., 2009) optimize preferences, but sampled evaluation can differ substantially from full ranking (Krichene & Rendle, 2020). We combine exact full-catalog likelihood with smooth cutoff barriers and use geometry as a structural constraint on model capacity.

3 Method

3.1 Problem Setup and Notation

Let $\mathcal{I}$ be the item catalog and $x_{1:t} = (i_1, \dots, i_t)$ a chronological prefix whose next item is $y = i_{t+1}$. A base behavioral encoder produces $\mathbf{h}_t^{(0)} = f_\theta(i_{1:t}) \in \mathbb{R}^{d_0}$. Particle transport and readout then construct the final state $\mathbf{h}_t \in \mathbb{R}^{d_h}$ used by the semantic-conditioned transformation and full-catalog scorer. Each item has a trainable behavioral state $\mathbf{b}_i \in \mathbb{R}^{d_h}$. We use i, y for item indices, t, j for time indices, k, ℓ for particle indices, and c for a ranking cutoff. The semantic dimension, complex rank, and particle dimension are d_s, r, d_p; P, N_h, S denote the numbers of particles, hard negatives, and Sinkhorn iterations, respectively.

For real or complex $\mathbf{x}$, define the numerically guarded unitizer

$$\mathsf{N}_{\epsilon_{\mathrm{norm}}}(\mathbf{x}) = \frac{\mathbf{x}}{\max(\|\mathbf{x}\|_2, \epsilon_{\mathrm{norm}})}, \qquad \epsilon_{\mathrm{norm}} = 10^{-6}. \tag{1}$$

It equals exact unit normalization for every nondegenerate state used in the analysis. LOTUS-UCC ultimately constructs one query $\mathbf{q}_t \in \mathbb{R}^{d_h+2r}$, one catalog key $\mathbf{k}_i \in \mathbb{R}^{d_h+2r}$ per item, and one full-catalog score vector $\mathbf{o}_t$; these objects are defined in Equation 26.

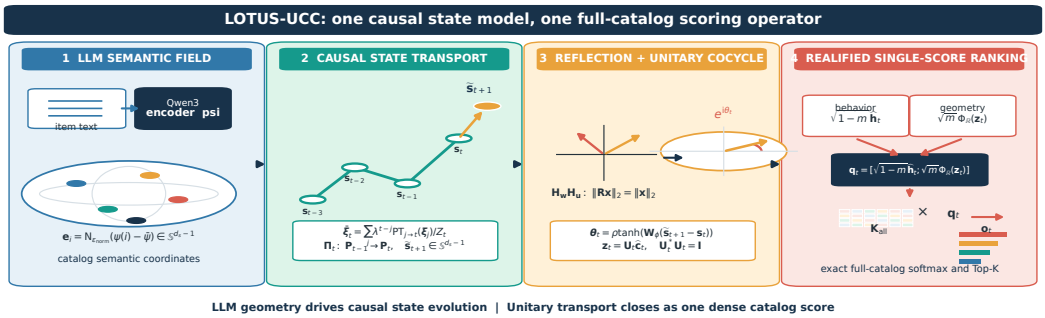

Figure 1: **LOTUS-UCC overview.** Semantic forecasting, norm-preserving evolution, and full-catalog ranking.

3.2 Semantic Coordinate Field and Causal Spherical Transport

The language encoder ψ defines a fixed coordinate for every catalog item and a distinct state for every observed sequence position:

$$\bar{\psi} = |\mathcal{I}|^{-1} \sum_{i \in \mathcal{I}} \psi(i), \qquad \mathbf{e}_i = \mathsf{N}_{\epsilon_{\text{norm}}}\big(\psi(i) - \bar{\psi}\big) \in \mathbb{S}^{d_s-1}, \qquad \mathbf{s}_j := \mathbf{e}_{i_j}. \tag{2}$$

Thus $\mathbf{e}_i$ is an item-indexed catalog coordinate, whereas $\mathbf{s}_j$ is the semantic state at time j. Let $\mathrm{Log}_{\mathbf{x}}(\mathbf{y})$ and $\mathsf{PT}_{\mathbf{x}\to\mathbf{y}}$ denote the spherical logarithm and shortest-geodesic parallel transport. Each observed transition is represented in its endpoint tangent space and transported to the common space $T_{\mathbf{s}_t}\mathbb{S}^{d_s-1}$:

$$\boldsymbol{\xi}_j = -\,\mathrm{Log}_{\mathbf{s}_j}(\mathbf{s}_{j-1}), \qquad \widetilde{\boldsymbol{\xi}}_{j\to t} = \mathsf{PT}_{\mathbf{s}_j\to\mathbf{s}_t}(\boldsymbol{\xi}_j), \qquad \bar{\boldsymbol{\xi}}_t = \frac{\sum_{j=t-W_v+1}^{t} \lambda^{t-j} \widetilde{\boldsymbol{\xi}}_{j\to t}}{\sum_{j=t-W_v+1}^{t} \lambda^{t-j}}. \tag{3}$$

Here $\boldsymbol{\xi}_j \in T_{\mathbf{s}_j}\mathbb{S}^{d_s-1}$ and $\lambda \in (0,1]$ is the temporal decay. For non-antipodal states,

$$\mathsf{PT}_{\mathbf{x}\to\mathbf{y}}(\mathbf{v}) = \mathbf{v} - \frac{\langle \mathbf{v}, \mathbf{y}\rangle}{1 + \langle \mathbf{x}, \mathbf{y}\rangle}(\mathbf{x} + \mathbf{y}).$$

Define $\mathrm{Clip}_r(\mathbf{v}) = \min\{1, r/\|\mathbf{v}\|_2\}\mathbf{v}$, the causal local mean $\boldsymbol{\mu}_t = \mathsf{N}_{\epsilon_{\text{norm}}}(\sum_{j=t-W_m+1}^{t} \lambda^{t-j}\mathbf{s}_j)$, and

$$\mathbf{s}_t^{\text{geo}} = \mathrm{Exp}_{\mathbf{s}_t}\big(\mathrm{Clip}_{r_{\text{geo}}}(\alpha \bar{\boldsymbol{\xi}}_t)\big), \qquad \widetilde{\mathbf{s}}_{t+1}^{\text{geo}} = \mathsf{N}_{\epsilon_{\text{norm}}}\big(\beta \mathbf{s}_t^{\text{geo}} + (1-\beta)\boldsymbol{\mu}_t\big). \tag{4}$$

The extrapolation scale satisfies $\alpha > 0$, the mixing coefficient $\beta \in [0,1]$, and $r_{\text{geo}} > 0$ bounds the geodesic step. Every quantity depends only on $i_{1:t}$.

3.3 Persistent Particle Transport and Readout

The particle stack maintains multiple persistent latent semantic states, allowing the model to retain competing intent hypotheses while balanced transport reassigns historical mass to event-conditioned anchors. Let $\mathbf{h}_t^{(0)} = f_\theta(i_{1:t}) \in \mathbb{R}^{d_0}$ be the causal state before readout. We maintain P unit particles $\mathbf{p}_{t,k} \in \mathbb{S}^{d_p-1}$, initialized by $\mathbf{p}_{0,k} = \mathsf{N}_{\epsilon_{\text{norm}}}(\widetilde{\mathbf{p}}_{0,k})$, and define $\mathbf{d}_t = \mathbf{s}_t - \mathbf{s}_{t-1} \in \mathbb{R}^{d_s}$, with $\mathbf{d}_1 = \mathbf{0}$. The event-conditioned anchor for slot ℓ is

$$\mathbf{r}_{t,\ell} = \mathsf{N}_{\epsilon_{\text{norm}}}\Big(\big[\mathbf{A}_h \mathbf{h}_t^{(0)} + \mathbf{A}_s \mathbf{s}_t + \mathbf{A}_d \mathbf{d}_t\big]_\ell + \boldsymbol{\delta}_\ell^{\text{slot}}\Big), \tag{5}$$

where $\mathbf{A}_h \in \mathbb{R}^{Pd_p \times d_0}$, $\mathbf{A}_s, \mathbf{A}_d \in \mathbb{R}^{Pd_p \times d_s}$, and $[\mathbf{z}]_\ell = \mathbf{z}_{(\ell-1)d_p+1:\ell d_p}$ for $\mathbf{z} \in \mathbb{R}^{Pd_p}$.

The learned diagonal metric is

$$\widetilde{\mathbf{g}}_t = f_{\text{met}}([\mathbf{s}_t; \mathbf{d}_t]), \qquad g_{t,j} = \exp\big(0.75 \tanh(\widetilde{g}_{t,j})\big) > 0, \qquad \mathbf{G}_t = \mathrm{diag}(g_{t,1}, \ldots, g_{t,d_p}). \tag{6}$$

Writing $\|\mathbf{x}\|_{\mathbf{G}_t}^2 = \mathbf{x}^\top \mathbf{G}_t \mathbf{x}$ and $\lambda_{\text{in}} = \mathrm{softplus}(\eta_{\text{in}}) \geq 0$, the transport cost is

$$D_{t,k\ell} = d_p^{-1} \|\mathbf{p}_{t-1,k} - \mathbf{r}_{t,\ell}\|_{\mathbf{G}_t}^2 + \lambda_{\text{in}} \mathbf{1}\{k \neq \ell\}. \tag{7}$$

Let $\boldsymbol{\mu}_P = P^{-1}\mathbf{1}$ and $\mathcal{E}_{\mathrm{OT}}(\boldsymbol{\Pi}) = -\sum_{k,\ell} \Pi_{k\ell}(\log \Pi_{k\ell} - 1)$. The balanced plan is

$$\boldsymbol{\Pi}_t = \arg\min_{\boldsymbol{\Pi}\geq 0} \langle \boldsymbol{\Pi}, \mathbf{D}_t \rangle - \epsilon_{\mathrm{OT}} \mathcal{E}_{\mathrm{OT}}(\boldsymbol{\Pi}), \quad \boldsymbol{\Pi}\mathbf{1} = \boldsymbol{\mu}_P, \quad \boldsymbol{\Pi}^\top \mathbf{1} = \boldsymbol{\mu}_P. \tag{8}$$

The shared scalar $\eta_{\mathrm{OT}} \in \mathbb{R}$ sets $\epsilon_{\mathrm{OT}} = 0.05 + 0.95\sigma(\eta_{\mathrm{OT}}) \in (0.05, 1)$, controlling matching entropy; S log-domain Sinkhorn updates approximate the plan.

The plan directly advances the particle state. First form the transported source and a spherical proposal,

$$\bar{\mathbf{p}}_{t,\ell} = \mathsf{N}_{\epsilon_{\mathrm{norm}}}\left(P \sum_{k=1}^{P} \Pi_{t,k\ell} \mathbf{p}_{t-1,k} \right), \tag{9}$$

$$\eta_{t,\ell} = 0.05 + 0.45\sigma\Big([f_{\mathrm{step}}(\mathbf{h}_t^{(0)}, \mathbf{s}_t, \mathbf{d}_t)]_\ell \Big), \tag{10}$$

$$\mathbf{p}_{t,\ell}^{\mathrm{prop}} = \mathrm{Exp}_{\bar{\mathbf{p}}_{t,\ell}}\left(\eta_{t,\ell} \, \mathrm{Log}_{\bar{\mathbf{p}}_{t,\ell}}(\mathbf{r}_{t,\ell}) \right). \tag{11}$$

Here $\eta_{t,\ell} \in (0.05, 0.50)$ controls geodesic motion toward the anchor. The persistence gate $\zeta_{t,\ell} = 0.70 + 0.295\sigma([f_{\mathrm{surv}}(\mathbf{h}_t^{(0)}, \mathbf{s}_t, \mathbf{d}_t)]_\ell)$ controls retention of the preceding slot:

$$\mathbf{p}_{t,\ell} = \mathsf{N}_{\epsilon_{\mathrm{norm}}}\Big(\zeta_{t,\ell} \mathbf{p}_{t-1,\ell} + (1 - \zeta_{t,\ell}) \mathbf{p}_{t,\ell}^{\mathrm{prop}} \Big). \tag{12}$$

Set $\mathbf{P}_t = [\mathbf{p}_{t,1}^\top; \ldots; \mathbf{p}_{t,P}^\top] \in \mathbb{R}^{P \times d_p}$ and $\mathbf{r}_t^P = \sqrt{d_p}\,\mathrm{vec}(\mathbf{P}_t) \in \mathbb{R}^{P d_p}$. With $\mathbf{A}_{\mathrm{ro}} \in \mathbb{R}^{d_0 \times P d_p}$, $\mathbf{A}_{\mathrm{sem}} \in \mathbb{R}^{d_s \times P d_p}$, and the shared learned scalar $\kappa_{\mathrm{ro}} = 0.05 + 0.35\sigma(\eta_{\mathrm{ro}}) \in (0.05, 0.40)$, define

$$\widehat{\mathbf{h}}_t^{(0)} = \mathsf{N}_{\epsilon_{\mathrm{norm}}}(\mathbf{h}_t^{(0)}), \qquad \widehat{\mathbf{h}}_t^{P} = \mathsf{N}_{\epsilon_{\mathrm{norm}}}(\mathbf{A}_{\mathrm{ro}} \mathbf{r}_t^P), \tag{13}$$

$$\mathbf{h}_t^{\mathrm{tr}} = \mathrm{Exp}_{\widehat{\mathbf{h}}_t^{(0)}}\Big(\kappa_{\mathrm{ro}} \, \mathrm{Log}_{\widehat{\mathbf{h}}_t^{(0)}}(\widehat{\mathbf{h}}_t^P) \Big), \qquad \widetilde{\mathbf{s}}_{t+1} = \mathsf{N}_{\epsilon_{\mathrm{norm}}}(\mathbf{A}_{\mathrm{sem}} \mathbf{r}_t^P), \tag{14}$$

$$\mathbf{h}_t = \big[\mathbf{h}_t^{\mathrm{tr}}; \sqrt{\lambda_{\mathrm{geo}}}\, \widetilde{\mathbf{s}}_{t+1}^{\mathrm{geo}} \big] \in \mathbb{R}^{d_h}, \qquad d_h = d_0 + d_s, \tag{15}$$

where $\lambda_{\mathrm{geo}} = 0.04$. Equation 4 produces the causal pure geometric forecast from the observed semantic trajectory, while Equations 5–12 advance $\mathbf{P}_{t-1}$ to $\mathbf{P}_t$. Its readout $\mathbf{r}_t^P$ yields both $\mathbf{h}_t^{\mathrm{tr}}$ and the final post-particle semantic endpoint $\widetilde{\mathbf{s}}_{t+1}$. The final state $\mathbf{h}_t$ explicitly retains $\widetilde{\mathbf{s}}_{t+1}^{\mathrm{geo}}$, keeping intrinsic semantic motion visible to the catalog query. Consequently, both geometric motion and persistent particle dynamics enter $\boldsymbol{\chi}_t = [\mathbf{h}_t; \mathbf{s}_t; \widetilde{\mathbf{s}}_{t+1}]$, which controls the Householder frame and UCC. Padding retains $\mathbf{P}_{t-1}$.

Proposition 1 (balanced transport). Since $\epsilon_{\mathrm{OT}} > 0$ and $\mathbf{D}_t$ is finite, the transport kernel is strictly positive; the converged plan has its prescribed marginals and unit mass.

3.4 Semantic-Conditioned Householder Frames

Set $\boldsymbol{\chi}_t = [\mathbf{h}_t; \mathbf{s}_t; \widetilde{\mathbf{s}}_{t+1}] \in \mathbb{R}^{d_h + 2d_s}$. A source frame network and an item frame network produce raw vectors

$$(\mathbf{u}_t^{\mathrm{raw}}, \mathbf{w}_t^{\mathrm{raw}}) = f_{\mathrm{frame}}^{\mathrm{src}}(\boldsymbol{\chi}_t), \qquad (\mathbf{u}_i^{\mathrm{raw}}, \mathbf{w}_i^{\mathrm{raw}}) = f_{\mathrm{frame}}^{\mathrm{item}}(\mathbf{e}_i), \tag{16}$$

where every raw vector lies in $\mathbb{R}^{2r}$. Gram–Schmidt orthogonalization is explicit:

$$\mathbf{u}_t = \mathsf{N}_{\epsilon_{\mathrm{norm}}}(\mathbf{u}_t^{\mathrm{raw}}), \qquad \mathbf{w}_t = \mathsf{N}_{\epsilon_{\mathrm{norm}}}\big(\mathbf{w}_t^{\mathrm{raw}} - \mathbf{u}_t (\mathbf{u}_t^\top \mathbf{w}_t^{\mathrm{raw}}) \big), \tag{17}$$

$$\mathbf{u}_i = \mathsf{N}_{\epsilon_{\mathrm{norm}}}(\mathbf{u}_i^{\mathrm{raw}}), \qquad \mathbf{w}_i = \mathsf{N}_{\epsilon_{\mathrm{norm}}}\big(\mathbf{w}_i^{\mathrm{raw}} - \mathbf{u}_i (\mathbf{u}_i^\top \mathbf{w}_i^{\mathrm{raw}}) \big). \tag{18}$$

We assume each first raw vector and each Gram–Schmidt residual in Equations 16–18 has norm greater than ϵ_{norm}; therefore all four frame vectors are unit and the Householder factors below are well defined. For a unit vector $\mathbf{u}$, let $\mathbf{H}_{\mathbf{u}} = \mathbf{I}_{2r} - 2\mathbf{u}\mathbf{u}^\top$. The source and item transformations are

$$\mathbf{R}_t = \mathbf{H}_{\mathbf{w}_t} \mathbf{H}_{\mathbf{u}_t}, \qquad \mathbf{R}_i = \mathbf{H}_{\mathbf{w}_i} \mathbf{H}_{\mathbf{u}_i} \in \mathbb{R}^{2r \times 2r}. \tag{19}$$

Let $\Phi_{\mathbb{C}}([\mathbf{x}; \mathbf{y}]) = \mathbf{x} + \mathrm{i}\mathbf{y}$ map $\mathbb{R}^{2r}$ to $\mathbb{C}^r$. The source and target adapters contain explicit bounded residuals,

$$\mathbf{g}_{\mathrm{src}}(\boldsymbol{\chi}_t) = \mathbf{A}_{\mathrm{src}} \boldsymbol{\chi}_t + r_{\mathrm{src}} \tanh(f_{\mathrm{src}}(\boldsymbol{\chi}_t)), \tag{20}$$

$$\mathbf{g}_{\mathrm{tgt}}(\mathbf{e}_i) = \mathbf{A}_{\mathrm{tgt}} \mathbf{e}_i + \boldsymbol{\delta}_i^{\mathrm{tgt}}, \qquad \boldsymbol{\delta}_i^{\mathrm{tgt}} = r_{\mathrm{tgt}} \tanh(\mathbf{r}_i^{\mathrm{tgt}}), \qquad r_{\mathrm{src}} = r_{\mathrm{tgt}} = 0.30, \tag{21}$$

so each residual coordinate is bounded by its stated amplitude. The complex source and item fibers and their normalized forms are

$$\mathbf{c}_t = \Phi_{\mathbb{C}}\big(\mathbf{R}_t \mathbf{g}_{\mathrm{src}}(\boldsymbol{\chi}_t)\big), \quad \mathbf{c}_i = \Phi_{\mathbb{C}}\big(\mathbf{R}_i \mathbf{g}_{\mathrm{tgt}}(\mathbf{e}_i)\big), \quad \widehat{\mathbf{c}}_t = \mathsf{N}_{\epsilon_{\mathrm{norm}}}(\mathbf{c}_t), \quad \widehat{\mathbf{c}}_i = \mathsf{N}_{\epsilon_{\mathrm{norm}}}(\mathbf{c}_i). \tag{22}$$

3.5 Unitary Semantic Cocycle

We separate motion on the semantic base manifold from representation evolution in a complex fiber. Spherical and particle transport update intent, while the cocycle transports the recommendation state along that path. The ambient endpoint displacement $\boldsymbol{\Delta}_t = \widetilde{\mathbf{s}}_{t+1} - \mathbf{s}_t$ controls a bounded phase:

$$\boldsymbol{\theta}_t = \rho \tanh(\mathbf{W}_\phi \boldsymbol{\Delta}_t), \qquad \mathbf{U}_t = \mathrm{diag}\big(e^{\mathrm{i}\theta_{t,1}}, \dots, e^{\mathrm{i}\theta_{t,r}}\big), \qquad 0 < \rho \le \frac{\pi}{2}. \tag{23}$$

The evolved endpoint is $\mathbf{z}_t = \mathbf{U}_t \widehat{\mathbf{c}}_t \in \mathbb{C}^r$. For $a < b$, define the path operator

$$\boldsymbol{\mathcal{U}}_{b \leftarrow a} = \mathbf{U}_b \mathbf{U}_{b-1} \cdots \mathbf{U}_{a+1}. \tag{24}$$

It satisfies $\boldsymbol{\mathcal{U}}_{c \leftarrow a} = \boldsymbol{\mathcal{U}}_{c \leftarrow b} \boldsymbol{\mathcal{U}}_{b \leftarrow a}$ and preserves complex norms and inner products.

3.6 Realification and Full-Catalog Scoring

Define the realification operator $\Phi_{\mathbb{R}}(\mathbf{z}) = [\mathrm{Re}\,\mathbf{z}; \mathrm{Im}\,\mathbf{z}] \in \mathbb{R}^{2r}$. It obeys $\Phi_{\mathbb{R}}(\mathbf{z})^\top \Phi_{\mathbb{R}}(\mathbf{c}) = \mathrm{Re}(\mathbf{z}^* \mathbf{c})$. Let $\widehat{\mathbf{h}}_t = \mathsf{N}_{\epsilon_{\mathrm{norm}}}(\mathbf{h}_t)$, $\widehat{\mathbf{b}}_i = \mathsf{N}_{\epsilon_{\mathrm{norm}}}(\mathbf{b}_i)$, and $m = m_{\max}\sigma(\eta_m) \in (0, m_{\max})$, where $0 < m_{\max} = 0.55 < 1$. The single-model query and catalog key are

$$\mathbf{q}_t = \begin{bmatrix} \sqrt{1-m}\,\widehat{\mathbf{h}}_t \\ \sqrt{m}\,\Phi_{\mathbb{R}}(\mathbf{z}_t) \end{bmatrix}, \qquad \mathbf{k}_i = \begin{bmatrix} \sqrt{1-m}\,\widehat{\mathbf{b}}_i \\ \sqrt{m}\,\Phi_{\mathbb{R}}(\widehat{\mathbf{c}}_i) \end{bmatrix}. \tag{25}$$

Stacking $\mathbf{k}_i^\top$ gives $\mathbf{K}_{\mathrm{all}} \in \mathbb{R}^{|\mathcal{I}| \times (d_h + 2r)}$, and

$$\mathbf{o}_t = \mathbf{K}_{\mathrm{all}} \mathbf{q}_t, \qquad o_{t,i} = \mathbf{k}_i^\top \mathbf{q}_t, \qquad \pi_t = \mathrm{argsort}_{i \notin x_{1:t}}\, o_{t,i}. \tag{26}$$

Because the behavioral and complex blocks are unit-normalized and their squared weights sum to one, $\|\mathbf{k}_i\|_2 = 1$ for every catalog item.

3.7 Full-Catalog Learning with CNDB

The catalog likelihood uses all items:

$$\mathcal{L}_{\mathrm{cat}} = -\frac{o_{t,y}}{\tau_{\mathrm{rec}}} + \log \sum_{i \in \mathcal{I}} \exp\left(\frac{o_{t,i}}{\tau_{\mathrm{rec}}}\right), \qquad \tau_{\mathrm{rec}} > 0. \tag{27}$$

Let $\mathcal{H}_t$ contain the indices of the $N_h = 128$ highest-scoring non-target items. With rank-comparison temperature $\tau_r > 0$, define

$$\widetilde{r}_t = 1 + \sum_{j \in \mathcal{H}_t} \sigma\left(\frac{o_{t,j} - o_{t,y}}{\tau_r}\right). \tag{28}$$

For cutoff $c \in \mathcal{K} = \{1, 5, 10, 20\}$, the normalized smooth barrier is

$$B_c(r) = \frac{\mathrm{softplus}((r-c)/(2\sqrt{c}))}{\mathrm{softplus}((N_h + 1 - c)/(2\sqrt{c}))}, \qquad \mathrm{softplus}(x) = \log(1 + e^x). \tag{29}$$

CNDB is

$$\mathcal{L}_{\mathrm{cndb}} = \frac{\log(1 + \widetilde{r}_t)}{\log(N_h + 2)} + \lambda_b \sum_{c \in \mathcal{K}} w_c B_c(\widetilde{r}_t), \quad \lambda_b = 0.75, \quad (w_1, w_5, w_{10}, w_{20}) = (0.15, 0.25, 0.40, 0.20). \tag{30}$$

The semantic loss directly supervises the final post-particle endpoint against the target coordinate $\mathbf{e}_y$:

$$\mathcal{L}_{\mathrm{sem}} = 1 - \langle \widetilde{\mathbf{s}}_{t+1}, \mathbf{e}_y \rangle. \tag{31}$$

Section 3.8 combines these terms with the particle transport regularizers.

3.8 Complete Training Objective

Let $\mathcal{B}$ be the valid, non-padding training checkpoints in the current optimization batch and $N = |\mathcal{B}|$. Using the particle stack defined above,

$$\mathcal{L}_{\text{transport}} = N^{-1} \sum_{t \in \mathcal{B}} \sum_{k,\ell} \Pi_{t,k\ell} D_{t,k\ell}, \tag{32}$$

$$\mathcal{L}_{\text{diversity}} = N^{-1} \sum_{t \in \mathcal{B}} \|\mathbf{P}_t \mathbf{P}_t^\top - \mathbf{I}_P\|_F^2 / P^2, \tag{33}$$

$$\mathcal{L}_{\text{metric}} = N^{-1} \sum_{t \in \mathcal{B}} \left(d_p^{-1} \sum_{j=1}^{d_p} \log g_{t,j} \right)^2. \tag{34}$$

The full objective is

$$\mathcal{L} = \mathcal{L}_{\text{cat}} + \lambda_r \mathcal{L}_{\text{cndb}} + \lambda_s \mathcal{L}_{\text{sem}} + \lambda_t \mathcal{L}_{\text{transport}} + \lambda_d \mathcal{L}_{\text{diversity}} + \lambda_g \mathcal{L}_{\text{metric}}, \tag{35}$$

where all coefficients are nonnegative and respectively control cutoff-aware rank learning, semantic forecasting, transport action, particle separation, and metric-volume regularization.

3.9 Geometry, Dynamics, and Decision

Layer I: geometric validity. Orthogonal Householder frames (Proposition 2) and unitary path operators (Theorem 1) preserve norms and pairwise distances and provide the geometric base case for the dynamical and decision-level analysis.

Proposition 3 (intrinsic–extrinsic curvature gap). For $\mathbf{s} \in \mathbb{S}^{d_s-1}$, $\mathbf{v} \in T_{\mathbf{s}}\mathbb{S}^{d_s-1}$, and $a = \|\mathbf{v}\|_2 \le r_{\text{geo}} < \pi$,

$$d_{\mathbb{S}}(\text{Exp}_{\mathbf{s}}(\mathbf{v}), \mathsf{N}_{\epsilon_{\text{norm}}}(\mathbf{s} + \mathbf{v})) = a - \arctan a \le a^3/3. \tag{36}$$

Thus Euclidean projection incurs third-order curvature error; under isometric continuation, path mismatch is at most $\frac{1}{3}\sum_t \|\mathbf{v}_t\|_2^3$.

Layer II: dynamical stability. Theorems 2–3 establish that the framed, normalized, and unitarily evolved endpoint $\mathcal{T}(\boldsymbol{\chi})$ is $L_{\mathcal{T}}$-Lipschitz under the component and nondegeneracy conditions specified in Appendix A, which also gives the explicit constant.

Lemma 1 (entropic-plan sensitivity). For fixed positive marginals and common $\epsilon_{\text{OT}} \ge \epsilon_{\min} > 0$, exact plans satisfy

$$\|\mathbf{\Pi}(\mathbf{D}) - \mathbf{\Pi}(\mathbf{D}')\|_F \le \epsilon_{\min}^{-1} \|\mathbf{D} - \mathbf{D}'\|_F. \tag{37}$$

Let $W = \max(W_v, W_m)$, $\mathbf{S}_t = (\mathbf{s}_{t-W+1}, \ldots, \mathbf{s}_t)$, and $\delta_t^{\text{in}} = \|\mathbf{h}_t^{(0)} - \mathbf{h}_t^{(0)\prime}\|_2 + \sum_{j=t-W+1}^{t} \|\mathbf{s}_j - \mathbf{s}_j'\|_2$. The complete causal input is $(\mathbf{P}_{t-1}, \mathbf{h}_t^{(0)}, \mathbf{S}_t)$, since the geometric forecast uses both velocity and local-mean histories.

Theorem 4 (end-to-end semantics-to-decision sensitivity). Compare two executions of the same learned model and catalog. Assume component maps are Lipschitz on a bounded causal domain, $\epsilon_{\text{OT}} \ge \epsilon_{\min} > 0$, normalized inputs have norm at least $\nu_N > \epsilon_{\text{norm}}$, and spherical logarithms remain $\delta_{\text{cut}} > 0$ from the antipodal cut locus. Then model-dependent constants $L_P, L_X, L_R, L_o < \infty$, determined by the component Lipschitz moduli, entropy floor, normalization margin, and cut-locus separation, satisfy, for $E_t = \|\mathbf{P}_t - \mathbf{P}_t'\|_F$,

$$E_t \le L_P E_{t-1} + L_X \delta_t^{\text{in}}, \tag{38}$$

$$\|\mathbf{h}_t - \mathbf{h}_t'\|_2 + \|\widetilde{\mathbf{s}}_{t+1} - \widetilde{\mathbf{s}}_{t+1}'\|_2 \le L_R(E_{t-1} + \delta_t^{\text{in}}), \tag{39}$$

$$\|\mathbf{o}_t - \mathbf{o}_t'\|_\infty \le L_o(E_{t-1} + \delta_t^{\text{in}}). \tag{40}$$

Here $L_P = C_P(1 + L_D/\epsilon_{\min})$, exposing the role of positive transport entropy, and recursion gives

$$E_t \le L_P^t E_0 + L_X \sum_{r=1}^{t} L_P^{t-r} \delta_r^{\text{in}}. \tag{41}$$

Corollary 1 (uniform compositional stability). If $L_P < 1$ and $\delta_t^{\text{in}} \le \delta_{\max}$ for all t, then particle-state and catalog-score discrepancies are uniformly bounded over time; Appendix A gives the explicit geometric-series bounds.

Layer III: decision closure. Since $\|\mathbf{k}_i\|_2 = 1$, the catalog is uniformly controlled by the last bound above. If $\gamma_K = o_{(K)} - o_{(K+1)} > 2L_o(E_{t-1} + \delta_t^{\text{in}})$, the Top-$K$ set is unchanged.

Proposition 4 (CNDB controls cutoff risk). Let $r_t^\star = 1 + \sum_{i \neq y} \mathbf{1}\{o_{t,i} > o_{t,y}\}$ be the exact rank. For every $c \le N_h$,

$$\mathbf{1}\{r_t^\star > c\} \le \frac{B_c(\widetilde{r}_t)}{B_c(1 + c/2)}, \qquad 1 - \text{HR}\,@c \le \frac{\mathbb{E}[B_c(\widetilde{r}_t)]}{B_c(1 + c/2)}. \tag{42}$$

Thus CNDB directly upper-controls cutoff failures, while its normalized log-rank term penalizes displacement within the retained set. Proofs are in Appendix A.

3.10 Causal and Numerical Realization

All event states, anchors, metrics, gates, frames, and phases at time t are functions of $i_{1:t}$; the target coordinate $\mathbf{e}_y$ occurs only in the training loss. The explicit entropy floor supports $\epsilon_{\min} > 0$, while r_{geo} and bounded gates limit local motion. Coincident semantic points use identity transport, and a deterministic least-aligned axis handles antipodal cases. The guarded unitizer in Equation 1 supports the normalization margin used by Theorems 3–4 and prevents division by zero in finite precision.

4 Experiments

Evaluation questions. The experiments mirror the theoretical chain. RQ1 measures full-catalog ranking quality (Table 1). RQ2 evaluates manifold consistency, controlled composition, and decision closure through five component removals, Euclidean transport, a non-unitary transition, and removal of CNDB (Table 3). RQ3 evaluates whether the semantic dynamics persist across four language representation spaces (Table 4). RQ4 studies local sensitivity to particle, Sinkhorn, semantic, and transport choices, while RQ5 measures computational efficiency and catalog exposure. Validation selects configurations and every table reports held-out test performance.

4.1 Setup

Data and protocol. We evaluate five processed benchmark instances: Amazon-2014 five-core Beauty, Sports and Outdoors, and Toys and Games; Yelp 2018 with 10-core filtering; and the Online Retail instance used by MPT (Xu et al., 2026). Appendix Table 6 gives their exact statistics. For each user, the final interaction is test, the penultimate interaction is validation, and all earlier interactions are training. Sequences are truncated to 50. We rank the target against the complete item catalog after masking previously seen items and report HR and NDCG at 10 and 20. Hyperparameters and checkpoints are selected only by validation NDCG@10.

Implementation. We use Qwen3-Embedding-0.6B as the semantic encoder. Unless stated otherwise, $W_v = 3$, $\lambda = 0.8$, $W_m = 5$, $\alpha = 0.75$, $\beta = 0.70$, complex rank is 64, and the phase radius is $\pi/2$. Main results are means over seeds 42, 43, and 44; Appendix C reports the complete mean and standard deviation for every dataset–metric pair. Every component ablation uses the same split, training budget, and full-catalog evaluator.

4.2 Main Results

Table 1 compares all methods under the same dataset and full-catalog evaluation setting. LOTUS-UCC achieves SOTA performance on 11 of 12 metric–dataset pairs, including all eight metrics on Sports and Toys. Its average improvement over the strongest baseline is 6.32%, reaching 12.23% on Sports NDCG@20 and 10.70% on Toys NDCG@10. Beauty HR@10, HR@20, and NDCG@20 also surpass the previous best results. Appendix Figure 2 visualizes gains across both metric families. The gains appear at both cutoffs and in both relevance- and position-sensitive metrics, indicating that the

Table 1: Full-catalog results on Beauty, Sports, and Toys. Best/second-best values are red/blue; LOTUS-UCC is shaded green.

	Beauty				Sports				Toys			
Model	H@10	N@10	H@20	N@20	H@10	N@10	H@20	N@20	H@10	N@10	H@20	N@20
GRU4Rec	0.0524	0.0278	0.0788	0.0344	0.0259	0.0135	0.0407	0.0172	0.0490	0.0282	0.0704	0.0335
SASRec	0.0753	0.0417	0.1060	0.0495	0.0405	0.0212	0.0607	0.0263	0.0846	0.0493	0.1115	0.0561
FMLPRec	0.0775	0.0454	0.1072	0.0529	0.0413	0.0238	0.0593	0.0283	0.0855	0.0521	0.1124	0.0589
HSTU	0.0902	0.0535	0.1246	0.0621	0.0550	0.0317	0.0788	0.0377	0.0981	0.0610	0.1288	0.0687
SASRec+	0.0982	0.0591	0.1348	0.0683	0.0576	0.0337	0.0824	0.0399	0.1040	0.0645	0.1393	0.0733
UniSRec	0.0803	0.0382	0.1233	0.0490	0.0446	0.0221	0.0699	0.0285	0.0829	0.0380	0.1235	0.0483
RecFormer	0.0676	0.0333	0.1017	0.0419	0.0345	0.0167	0.0555	0.0220	0.0842	0.0405	0.1221	0.0501
E4SRec	0.0773	0.0469	0.1046	0.0538	0.0387	0.0218	0.0563	0.0262	0.0798	0.0492	0.1061	0.0558
Qwen2.5	0.0826	0.0464	0.1200	0.0558	0.0473	0.0256	0.0712	0.0316	0.0784	0.0433	0.1114	0.0515
MPT-Adaptor	0.1118	0.0673	0.1531	0.0777	0.0651	0.0373	0.0924	**0.0442**	0.1202	0.0705	0.1617	0.0809
MPT+LoRA	0.1100	0.0658	0.1514	0.0763	0.0644	0.0365	0.0927	0.0436	0.1192	0.0701	0.1611	0.0806
WPGRec	**0.1197**	**0.0742**	**0.1583**	**0.0813**	**0.0704**	**0.0401**	**0.0991**	–	**0.1222**	**0.0747**	**0.1620**	**0.0848**
LOTUS-UCC	**0.1222**	**0.0741**	**0.1657**	**0.0851**	**0.0733**	**0.0423**	**0.1024**	**0.0496**	**0.1341**	**0.0827**	**0.1762**	**0.0933**
Improve	+2.12%	-0.13%	+4.70%	+4.65%	+4.16%	+5.51%	+3.29%	+12.23%	+9.70%	+10.70%	+8.78%	+10.05%

Table 2: Full-catalog results on Yelp and Online Retail under the MPT setting (Xu et al., 2026). Best/second-best values are red/blue.

	Yelp				Online Retail			
Model	H@10	N@10	H@20	N@20	H@10	N@10	H@20	N@20
GRU4Rec	0.0246	0.0116	0.0448	0.0166	0.1175	0.0561	0.1931	0.0751
SASRec	0.0298	0.0123	0.0539	0.0184	0.1245	0.0524	0.2014	0.0718
FMLPRec	0.0334	0.0138	0.0606	0.0207	0.1333	0.0548	0.2237	0.0775
HSTU	0.0453	0.0215	0.0773	0.0295	0.1452	0.0656	0.2304	0.0870
SASRec+	0.0466	0.0209	0.0796	0.0292	0.1589	0.0719	**0.2575**	0.0968
UniSRec	0.0343	0.0170	0.0567	0.0226	0.1424	0.0662	0.2275	0.0877
RecFormer	0.0247	0.0121	0.0429	0.0166	0.1407	0.0683	0.2194	0.0881
E4SRec	0.0291	0.0139	0.0513	0.0195	0.1551	**0.0749**	0.2528	**0.0995**
Qwen2.5	0.0411	0.0204	0.0682	0.0272	0.1503	0.0731	0.2356	0.0945
MPT-Adaptor	0.0461	0.0224	0.0783	0.0305	**0.1608**	0.0741	0.2572	0.0982
MPT+LoRA	**0.0495**	**0.0241**	**0.0832**	**0.0325**	0.1568	0.0746	0.2502	0.0980
LOTUS-UCC	**0.0686**	**0.0361**	**0.1068**	**0.0457**	**0.2553**	**0.1508**	**0.3448**	**0.1733**
Improve	+38.61%	+49.96%	+28.34%	+40.66%	+58.77%	+101.32%	+33.89%	+74.19%

learned transition improves candidate inclusion and within-list ordering together. Their consistency across three Amazon domains further shows that the effect is not tied to one catalog density.

Extension to Yelp and Online Retail. Table 2 extends the same chronological leave-one-out and full-catalog protocol to the remaining two MPT benchmarks. We compare with protocol-matched published results from Xu et al. (2026). LOTUS-UCC leads all eight metrics: its relative gains range from 28.34% to 49.96% on Yelp and from 33.89% to 101.32% on Online Retail. Together, Tables 1 and 2 establish new best results on 19 of 20 dataset–metric pairs across five benchmarks. The larger improvements on Yelp and Online Retail extend this pattern beyond product-only catalogs to business and transaction sequences with different sparsity and metadata structure. The shared direction across every metric shows that the same dynamics remain effective as catalog organization changes.

4.3 WHAT DOES EACH COMPONENT CONTRIBUTE?

Table 3 combines five component removals with three structural variants; all eight matched alternatives yield lower point estimates across the 12 metrics. Mean decreases for Euclidean Transport, Non-unitary UCC, and w/o CNDB are $(1.68, 2.08, 1.24) \times 10^{-3}$, respectively. These observations are consistent with the curvature, fiber, and cutoff distinctions in Equations 36 and 42. UCC has the

Table 3: Component and structural ablations. Best/second-best values are red/blue.

	Beauty				Sports				Toys			
Model	H@10	N@10	H@20	N@20	H@10	N@10	H@20	N@20	H@10	N@10	H@20	N@20
w/o UCC	0.1222	0.0727	0.1651	0.0837	0.0716	0.0412	0.1011	0.0486	0.1328	0.0788	0.1772	0.0900
w/o semantic forecast	0.1245	0.0748	0.1672	0.0856	0.0730	0.0427	0.1022	0.0500	0.1355	0.0834	0.1764	0.0937
w/o transport	**0.1249**	**0.0751**	**0.1681**	**0.0860**	**0.0748**	**0.0435**	0.1026	**0.0505**	0.1367	0.0837	0.1772	0.0939
w/o diversity	0.1245	**0.0751**	0.1680	**0.0860**	0.0741	0.0433	**0.1027**	**0.0505**	**0.1369**	**0.0842**	0.1776	**0.0945**
w/o metric	0.1228	0.0740	0.1653	0.0847	0.0733	0.0427	0.1024	0.0500	0.1364	0.0839	0.1772	0.0941
Euclidean Transport	0.1241	0.0740	0.1672	0.0852	0.0737	0.0425	0.1021	0.0498	0.1351	0.0818	0.1773	0.0926
Non-unitary UCC	0.1238	0.0737	0.1668	0.0849	0.0733	0.0422	0.1018	0.0495	0.1346	0.0813	0.1768	0.0920
w/o CNDB	0.1246	0.0741	0.1678	0.0850	0.0744	0.0427	**0.1027**	0.0499	0.1362	0.0821	**0.1784**	0.0928
Full LOTUS-UCC	**0.1253**	**0.0756**	**0.1685**	**0.0865**	**0.0752**	**0.0438**	**0.1032**	**0.0510**	**0.1375**	**0.0846**	**0.1792**	**0.0952**

Table 4: Cross-LLM results. Each row is a separately trained single model; best/second-best values are red/blue.

	Beauty				Sports				Toys			
Semantic field	H@10	N@10	H@20	N@20	H@10	N@10	H@20	N@20	H@10	N@10	H@20	N@20
Qwen3	0.12485	**0.07549**	0.16760	0.08625	0.07419	0.04301	0.10281	0.05020	**0.13703**	**0.08410**	0.17917	**0.09467**
BGE-M3	**0.12597**	**0.07591**	0.16791	**0.08649**	0.07307	0.04270	**0.10293**	0.05020	0.13543	**0.08324**	**0.17932**	**0.09433**
E5-large-v2	**0.12467**	0.07523	**0.16867**	**0.08635**	0.07402	**0.04376**	0.10270	**0.05097**	0.13471	0.08249	0.17803	0.09342
TinyLlama	0.12311	0.07423	0.16626	0.08514	**0.07453**	**0.04310**	**0.10377**	**0.05045**	**0.13548**	0.08312	0.17587	0.09332

largest component effect at 3.38×10^{-3} and paired user-level bootstrap support; Appendix Figure 3 summarizes all 60 point-estimate differences. More specifically, removing semantic forecasting weakens anticipatory state updates, while removing transport, diversity, or metric regularization produces smaller but consistent losses. The three structural alternatives also remain below the full model in every column, aligning empirical ranking behavior with the distinct geometric roles assigned by the theory rather than with a single dominant regularizer.

4.4 CROSS-LLM GENERALIZATION

Table 4 reports all 12 metrics for four separately trained semantic encoders. Column wins distribute as 3/4/3/2 across Qwen3, BGE-M3, E5-large-v2, and TinyLlama, while the mean best-to-worst span is 0.00178. The distributed leaders demonstrate consistent transfer of LOTUS-UCC dynamics across distinct semantic fields. No encoder dominates every dataset or cutoff; instead, each supplies a different coordinate realization while the final scores remain tightly clustered. This stability indicates that LOTUS-UCC relies on transferable semantic directions and their induced dynamics rather than on idiosyncratic dimensions of one frozen language model.

5 CONCLUSION

LOTUS-UCC reframes language-enhanced sequential recommendation as a semantics-to-decision dynamical problem. Language supplies the coordinate space, persistent and spherical transport evolve intent, norm-preserving fiber dynamics control representation propagation, and full-catalog scoring closes the process at discrete recommendation boundaries. LOTUS-UCC achieves SOTA performance on 19 of 20 metrics across five benchmarks and averages a 25.08% relative gain over the strongest baselines. All eight matched ablations yield lower point estimates across the 12 controlled metrics; UCC has the largest component effect and paired user-level bootstrap support, while the Euclidean, non-unitary, and CNDB variants support the corresponding structural mechanisms. Four language encoders exhibit consistent transfer across semantic spaces, and full-catalog throughput exceeds 1,100 users per second. The resulting view advances language-enhanced recommendation from static semantic matching toward controlled semantic evolution for sequential decision making. By linking semantic geometry, compositional stability, and cutoff-aware ranking in one trainable system, LOTUS-UCC provides a unified foundation for language-conditioned sequential recommendation.

Reproducibility Statement

Section 4 and Appendix B specify the data, split, candidate set, masking, metrics, main-result seeds, and model selection. The main text summarizes the complete cross-LLM evaluation and three-seed stability; Appendix C gives the complete run statistics, and Appendices D–F document statistical tests, sensitivity, efficiency, and exposure. Appendix A contains the proofs, and the final appendix records implementation constants.

AI Use Statement

AI assistants were used for language editing, figure-code drafting, and formatting. The authors designed the method, conducted the experiments, and verified the mathematical statements and results.

References

P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008. ISBN 978-0-691-13298-3.

Martin Arjovsky, Amar Shah, and Yoshua Bengio. Unitary evolution recurrent neural networks. In *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 1120–1128. PMLR, 2016. URL https://proceedings.mlr.press/v48/arjovsky16.html.

Jianlv Chen, Shitao Xiao, Peitian Zhang, Kun Luo, Defu Lian, and Zheng Liu. M3-Embedding: Multi-linguality, multi-functionality, multi-granularity text embeddings through self-knowledge distillation. In *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 2318–2335. Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.findings-acl.137. URL https://aclanthology.org/2024.findings-acl.137/.

Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, 4 edition, 2013.

Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. Session-based recommendations with recurrent neural networks. In *International Conference on Learning Representations*, 2016.

Yupeng Hou, Shanlei Mu, Wayne Xin Zhao, Yaliang Li, Bolin Ding, and Ji-Rong Wen. Towards universal sequence representation learning for recommender systems. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 585–593, 2022. doi: 10.1145/3534678.3539381.

Wang-Cheng Kang and Julian McAuley. Self-attentive sequential recommendation. In *IEEE International Conference on Data Mining*, pp. 197–206, 2018. doi: 10.1109/ICDM.2018.00035.

Walid Krichene and Steffen Rendle. On sampled metrics for item recommendation. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1748–1757, 2020. doi: 10.1145/3394486.3403226.

Jiacheng Li, Ming Wang, Jin Li, Jinmiao Fu, Xin Shen, Jingbo Shang, and Julian McAuley. Text is all you need: Learning language representations for sequential recommendation. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 1258–1267, 2023. doi: 10.1145/3580305.3599519.

Peilin Liu, Zhiquan Ji, and Gang Yan. WPGRec: Wavelet packet guided graph enhanced sequential recommendation. In *Proceedings of the 49th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 3964–3969, 2026. doi: 10.1145/3805712.3809907. URL https://doi.org/10.1145/3805712.3809907.

Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence*, pp. 452–461. AUAI Press, 2009.

Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pp. 1441–1450, 2019. doi: 10.1145/3357384.3357895.

Liang Wang, Nan Yang, Xiaolong Huang, Binxing Jiao, Linjun Yang, Daxin Jiang, Rangan Majumder, and Furu Wei. Text embeddings by weakly-supervised contrastive pre-training. *arXiv preprint arXiv:2212.03533*, 2022.

Wuhong Wang, Jianhui Ma, Yuren Zhang, Kai Zhang, Junzhe Jiang, Yihui Yang, Yacong Zhou, and Zheng Zhang. Intent oriented contrastive learning for sequential recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 12748–12756, 2025. doi: 10.1609/aaai.v39i12.33390.

Cong Xu, Guoliang Li, Jun Wang, and Wei Zhang. Markovian pre-trained transformer for next-item recommendation. *arXiv preprint arXiv:2601.08275*, 2026. URL https://arxiv.org/abs/2601.08275.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, Chujie Zheng, Dayiheng Liu, Fan Zhou, Fei Huang, Feng Hu, Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jing Zhou, Jingren Zhou, Junyang Lin, Kai Dang, Keqin Bao, Kexin Yang, Le Yu, Lianghao Deng, Mei Li, Mingfeng Xue, Mingze Li, Pei Zhang, Peng Wang, Qin Zhu, Rui Men, Ruize Gao, Shixuan Liu, Shuang Luo, Tianhao Li, Tianyi Tang, Wenbiao Yin, Xingzhang Ren, Xinyu Wang, Xinyu Zhang, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yinger Zhang, Yu Wan, Yuqiong Liu, Zekun Wang, Zeyu Cui, Zhenru Zhang, Zhipeng Zhou, and Zihan Qiu. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*, 2025.

Jiaqi Zhai, Lucy Liao, Xing Liu, Yueming Wang, Rui Li, Xuan Cao, Leon Gao, Zhaojie Gong, Fangda Gu, Jiayuan He, Yinghai Lu, and Yu Shi. Actions speak louder than words: Trillion-parameter sequential transducers for generative recommendations. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 58484–58509. PMLR, 2024. URL https://proceedings.mlr.press/v235/zhai24a.html.

Peiyuan Zhang, Guangtao Zeng, Tianduo Wang, and Wei Lu. TinyLlama: An open-source small language model. *arXiv preprint arXiv:2401.02385*, 2024.

Kun Zhou, Hui Yu, Wayne Xin Zhao, and Ji-Rong Wen. Filter-enhanced MLP is all you need for sequential recommendation. In *Proceedings of the ACM Web Conference 2022*, pp. 2388–2399, 2022. doi: 10.1145/3485447.3512111.

A PROOFS

A.1 PROOF OF PROPOSITION 1: BALANCED TRANSPORT

Let $\mathbf{K}_t^{\mathrm{OT}} = \exp(-\mathbf{D}_t/\epsilon_{\mathrm{OT}})$ elementwise. Positivity and uniform positive marginals give the unique entropic scaling $\mathbf{\Pi}_t = \mathrm{diag}(\boldsymbol{\kappa}^{\mathrm{row}})\mathbf{K}_t^{\mathrm{OT}}\,\mathrm{diag}(\boldsymbol{\kappa}^{\mathrm{col}})$. A row normalization enforces $\mathbf{\Pi}_t\mathbf{1} = \boldsymbol{\mu}_P$, and a column normalization enforces $\mathbf{\Pi}_t^\top\mathbf{1} = \boldsymbol{\mu}_P$. Alternation converges to the jointly feasible scaling, so $\mathbf{1}^\top\mathbf{\Pi}_t\mathbf{1} = \mathbf{1}^\top\boldsymbol{\mu}_P = 1$. □

A.2 PROOF OF PROPOSITION 2: ISOMETRIC HOUSEHOLDER FRAMES

For a unit vector $\mathbf{u}$,

$$\mathbf{H}_\mathbf{u}^\top\mathbf{H}_\mathbf{u} = (\mathbf{I} - 2\mathbf{u}\mathbf{u}^\top)^2 = \mathbf{I}.$$

Thus each Householder factor is orthogonal, and $\mathbf{R}^\top\mathbf{R} = \mathbf{H}_\mathbf{u}^\top\mathbf{H}_\mathbf{w}^\top\mathbf{H}_\mathbf{w}\mathbf{H}_\mathbf{u} = \mathbf{I}$. Consequently $\|\mathbf{R}\mathbf{x}\|_2 = \|\mathbf{x}\|_2$ and $\langle\mathbf{R}\mathbf{x}, \mathbf{R}\mathbf{y}\rangle = \langle\mathbf{x}, \mathbf{y}\rangle$. □

A.3 PROOF OF THEOREM 1: NON-EXPANSIVE COCYCLE DYNAMICS

Every diagonal entry of $\mathbf{U}_t$ has modulus one, hence $\mathbf{U}_t^*\mathbf{U}_t = \mathbf{I}$. Products of unitary matrices are unitary, and associativity gives

$$\boldsymbol{\mathcal{U}}_{c\leftarrow b}\boldsymbol{\mathcal{U}}_{b\leftarrow a} = (\mathbf{U}_c\cdots\mathbf{U}_{b+1})(\mathbf{U}_b\cdots\mathbf{U}_{a+1}) = \boldsymbol{\mathcal{U}}_{c\leftarrow a}.$$

The norm and difference identities follow from $\|\boldsymbol{\mathcal{U}}\mathbf{z}\|_2^2 = \mathbf{z}^*\boldsymbol{\mathcal{U}}^*\boldsymbol{\mathcal{U}}\mathbf{z} = \|\mathbf{z}\|_2^2$. □

A.4 PROOF OF THEOREM 2: SEMANTIC SENSITIVITY

For real x, y, $|e^{\mathrm{i}x} - e^{\mathrm{i}y}| \le |x - y|$. The spectral norm of a diagonal matrix is its largest diagonal magnitude. The coordinatewise 1-Lipschitz property of $\tanh$ therefore gives

$$\|\mathbf{U}(\boldsymbol{\Delta}) - \mathbf{U}(\boldsymbol{\Delta}')\|_2 \le \rho\|\mathbf{W}_\phi(\boldsymbol{\Delta} - \boldsymbol{\Delta}')\|_2 \le \rho\|\mathbf{W}_\phi\|_2\|\boldsymbol{\Delta} - \boldsymbol{\Delta}'\|_2.$$

□

A.5 PROOF OF THEOREM 3: COMPLETE ENDPOINT PERTURBATION

Assume the framed source has norm at least $\nu > \epsilon_{\mathrm{norm}}$, the source adapter has norm at most B_{src}, and the frame, source, and displacement maps have constants $L_{\mathrm{frame}}, L_{\mathrm{src}}, L_\Delta$, respectively. Write

$$\mathbf{c}(\boldsymbol{\chi}) = \Phi_{\mathbb{C}}(\mathbf{R}(\boldsymbol{\chi})\mathbf{g}_{\mathrm{src}}(\boldsymbol{\chi}))\,, \qquad \widehat{\mathbf{c}}(\boldsymbol{\chi}) = \frac{\mathbf{c}(\boldsymbol{\chi})}{\|\mathbf{c}(\boldsymbol{\chi})\|_2}.$$

For unit frame vectors,

$$\|\mathbf{H}_\mathbf{u} - \mathbf{H}_{\mathbf{u}'}\|_2 = 2\|\mathbf{u}\mathbf{u}^\top - \mathbf{u}'\mathbf{u}'^\top\|_2 \le 4\|\mathbf{u} - \mathbf{u}'\|_2.$$

Adding and subtracting $\mathbf{H}_\mathbf{w}\mathbf{H}_{\mathbf{u}'}$, then applying the L_{frame} assumption, yields

$$\|\mathbf{R}(\boldsymbol{\chi}) - \mathbf{R}(\boldsymbol{\chi}')\|_2 \le 4L_{\mathrm{frame}}\|\boldsymbol{\chi} - \boldsymbol{\chi}'\|_2.$$

Since real-to-complex identification is an isometry,

$$\|\mathbf{c}(\boldsymbol{\chi}) - \mathbf{c}(\boldsymbol{\chi}')\|_2 \le (L_{\mathrm{src}} + 4B_{\mathrm{src}}L_{\mathrm{frame}})\|\boldsymbol{\chi} - \boldsymbol{\chi}'\|_2.$$

Exact normalization is $2/\nu$-Lipschitz on $\|\mathbf{c}\|_2 \ge \nu$. Moreover, Theorem 2 and the displacement assumption give

$$\|\mathbf{U}(\boldsymbol{\Delta}(\boldsymbol{\chi})) - \mathbf{U}(\boldsymbol{\Delta}(\boldsymbol{\chi}'))\|_2 \le \rho\|\mathbf{W}_\phi\|_2 L_\Delta\|\boldsymbol{\chi} - \boldsymbol{\chi}'\|_2.$$

Add and subtract $\mathbf{U}(\boldsymbol{\Delta}(\boldsymbol{\chi}))\widehat{\mathbf{c}}(\boldsymbol{\chi}')$. Unitarity controls the normalized-state difference, while $\|\widehat{\mathbf{c}}\|_2 = 1$ controls the operator difference. Summing both terms gives the stated $L_\mathcal{T}$ bound. □

A.6 REALIFICATION IDENTITY

For $\mathbf{z} = \mathbf{z}_{\mathrm{R}} + \mathrm{i}\mathbf{z}_{\mathrm{I}}$ and $\mathbf{c} = \mathbf{c}_{\mathrm{R}} + \mathrm{i}\mathbf{c}_{\mathrm{I}}$,

$$\mathrm{Re}(\mathbf{z}^*\mathbf{c}) = \mathbf{z}_{\mathrm{R}}^\top\mathbf{c}_{\mathrm{R}} + \mathbf{z}_{\mathrm{I}}^\top\mathbf{c}_{\mathrm{I}} = \Phi_{\mathbb{R}}(\mathbf{z})^\top\Phi_{\mathbb{R}}(\mathbf{c}).$$

Thus the complex Hermitian score is exactly one real dot product. □

A.7 Proof of Proposition 3: intrinsic–extrinsic curvature gap

Let $a = \|\mathbf{v}\|_2$ and $\widehat{\mathbf{v}} = \mathbf{v}/a$ for $a > 0$. Since $\mathbf{v} \perp \mathbf{s}$,

$$\mathrm{Exp}_{\mathbf{s}}(\mathbf{v}) = \cos(a)\mathbf{s} + sin(a)\widehat{\mathbf{v}}, \qquad \frac{\mathbf{s}+\mathbf{v}}{\|\mathbf{s}+\mathbf{v}\|_2} = \cos(\arctan a)\mathbf{s} + sin(\arctan a)\widehat{\mathbf{v}}.$$

Both points lie on the same great circle, so their geodesic separation is $a - \arctan a$. Moreover,

$$a - \arctan a = \int_0^a \frac{x^2}{1+x^2}\,dx \le \int_0^a x^2\,dx = \frac{a^3}{3}.$$

The claim is trivial for $a = 0$. Under isometric continuation, telescoping the local deviations and applying the triangle inequality gives the stated path bound. □

A.8 Proof of Lemma 1: entropic-plan sensitivity

On the balanced transport polytope, write

$$F_{\mathbf{D}}(\mathbf{\Pi}) = \langle \mathbf{D}, \mathbf{\Pi}\rangle + \epsilon_{\mathrm{OT}} \sum_{k,\ell} \Pi_{k\ell}(\log \Pi_{k\ell} - 1).$$

Every feasible entry is at most one, hence the entropy Hessian has diagonal entries $\epsilon_{\mathrm{OT}}/\Pi_{k\ell} \ge \epsilon_{\min}$. Therefore $F_{\mathbf{D}}$ is $\epsilon_{\min}$-strongly convex in Frobenius norm. Applying the two variational inequalities for minimizers $\mathbf{\Pi}, \mathbf{\Pi}'$, adding them, and using Cauchy–Schwarz yields

$$\epsilon_{\min}\|\mathbf{\Pi} - \mathbf{\Pi}'\|_F^2 \le \langle \mathbf{D} - \mathbf{D}', \mathbf{\Pi}' - \mathbf{\Pi}\rangle \le \|\mathbf{D} - \mathbf{D}'\|_F \|\mathbf{\Pi} - \mathbf{\Pi}'\|_F.$$

Cancellation proves Equation 37. □

A.9 Proof of Theorem 4: end-to-end stability

The metric entries satisfy $e^{-0.75} \le g_{t,j} \le e^{0.75}$. Unit particles, bounded metrics, and Lipschitz anchors therefore make the cost map Lipschitz:

$$\|\mathbf{D}_t - \mathbf{D}'_t\|_F \le L_D(E_{t-1} + \delta_t^{\mathrm{in}})$$

for finite L_D. Lemma 1 then controls the transport plan by $L_D/\epsilon_{\min}$ times the same input discrepancy. Transported barycenters are affine before normalization. The lower norm bound makes normalization $2/\nu_N$-Lipschitz, while exclusion of the antipodal cut locus makes Exp and Log Lipschitz on the resulting compact domain. Composing these maps with the bounded proposal and persistence gates gives

$$E_t \le C_P(1 + L_D/\epsilon_{\min})E_{t-1} + L_X\delta_t^{\mathrm{in}}.$$

This establishes the first line with $L_P = C_P(1 + L_D/\epsilon_{\min})$; induction gives Equation 41.

The constants are determined by the concrete model components. In particular, $L_o = L_o(L_{\mathrm{frame}}, L_{\mathrm{src}}, L_\Delta, L_{\mathrm{anchor}}, L_{\mathrm{met}}, L_{\mathrm{step}}, L_{\mathrm{surv}}, L_{\mathrm{read}}, \epsilon_{\min}, \nu_N, \delta_{\mathrm{cut}})$, with analogous dependencies for L_P, L_X, L_R.

Vectorization and both readout heads are linear, and the geometric forecast is Lipschitz on the same nondegenerate semantic window. Their composition with normalization and geodesic interpolation yields the second line of Equation 40. Consequently $\boldsymbol{\chi}_t = [\mathbf{h}_t; \mathbf{s}_t; \widetilde{\mathbf{s}}_{t+1}]$ is Lipschitz in $(\mathbf{P}_{t-1}, \mathbf{h}_t^{(0)}, \mathbf{S}_t)$. Theorem 3 controls the complex endpoint, realification is an isometry, and the behavioral normalizer is Lipschitz under the same lower norm bound. Finally, $\|\mathbf{k}_i\|_2 = 1$ and Cauchy–Schwarz imply $\|\mathbf{o}_t - \mathbf{o}'_t\|_\infty \le \|\mathbf{q}_t - \mathbf{q}'_t\|_2$, proving the last line and the margin claim. □

A.10 Proof of Corollary 1: uniform compositional stability

Substituting $\delta_t^{\mathrm{in}} \le \delta_{\max}$ into Equation 41 and summing the geometric series gives

$$E_t \le L_P^t E_0 + \frac{L_X\delta_{\max}(1 - L_P^t)}{1 - L_P}.$$

Taking the supremum and applying the score line of Equation 40 gives

$$\sup_t E_t \le E_0 + \frac{L_X\delta_{\max}}{1 - L_P}, \qquad \sup_t \|\mathbf{o}_t - \mathbf{o}'_t\|_\infty \le L_o\left(E_0 + \delta_{\max} + \frac{L_X\delta_{\max}}{1 - L_P}\right).$$

□

A.11 Proof of Proposition 4: CNDB cutoff control

If $r_t^\star > c$, at least c non-target items outrank the target. Since $c \le N_h$, all of these occur among the selected hard negatives, and each contributes more than $1/2$ to Equation 28. Hence $\widetilde{r}_t \ge 1 + c/2$. The function B_c is positive and increasing, so on this event

$$1 \le B_c(\widetilde{r}_t)/B_c(1 + c/2).$$

Off the event the indicator is zero, proving the pointwise inequality. Taking expectations gives the HR bound. □

B Evaluation Protocol

The experiments use the five processed benchmark instances in Table 6: Amazon-2014 five-core Beauty, Sports, and Toys; Yelp 2018 with 10-core filtering; and Online Retail. Every dataset uses chronological leave-one-out with one validation and one test item per user, maximum sequence length 50, seen-item masking, and ranking over the complete item set. We report HR@10, NDCG@10, HR@20, and NDCG@20 under this shared full-catalog evaluation configuration.

C Full Three-Seed Results

Table 5: Three-seed test results (mean ± standard deviation).

Dataset	HR@10	NDCG@10	HR@20	NDCG@20
Beauty	0.12224 ± 0.00227	0.07411 ± 0.00121	0.16574 ± 0.00190	0.08508 ± 0.00103
Sports	0.07333 ± 0.00079	0.04231 ± 0.00080	0.10236 ± 0.00123	0.04961 ± 0.00054
Toys	0.13406 ± 0.00261	0.08269 ± 0.00123	0.17623 ± 0.00275	0.09333 ± 0.00117
Yelp	0.06861 ± 0.00031	0.03614 ± 0.00022	0.10678 ± 0.00050	0.04572 ± 0.00025
Online Retail	0.25531 ± 0.00134	0.15079 ± 0.00151	0.34475 ± 0.00253	0.17332 ± 0.00079

C.1 Dataset Statistics

Table 6: Processed dataset statistics.

Dataset	Users	Items	Interactions
Beauty	22,363	12,101	198,502
Sports	35,598	18,357	296,337
Toys	19,412	11,924	167,597
Yelp	77,277	45,638	2,103,896
Online Retail	16,520	3,469	519,906

D Ablations and Statistical Evidence

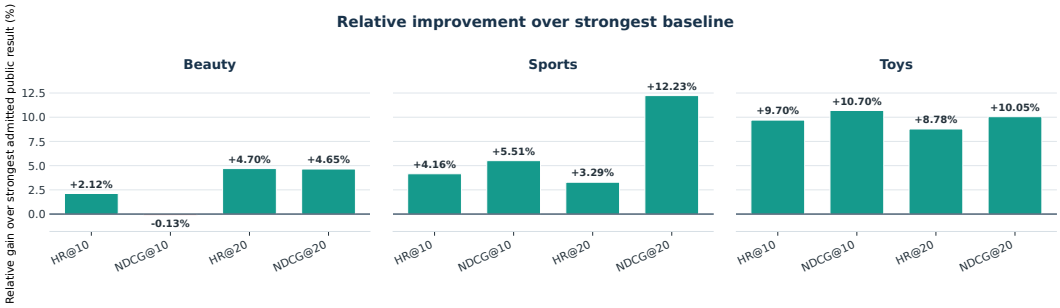

Figure 2: Relative improvement over the strongest baseline.

The study contains five component removals and three structural tests. Every matched alternative yields lower point estimates across the 12 metrics relative to Full LOTUS-UCC. Euclidean Transport is consistent with the intrinsic curvature distinction, Non-unitary UCC isolates norm-preserving phase dynamics, and w/o CNDB probes cutoff-aware rank optimization. Their mean decreases are 1.68×10^{-3}, 2.08×10^{-3}, and 1.24×10^{-3}, respectively. Among component removals, UCC has the largest mean decrease and paired statistical support across every dataset; forecasting, transport, diversity, and metric regularization provide distinct gains across ranking cutoffs.

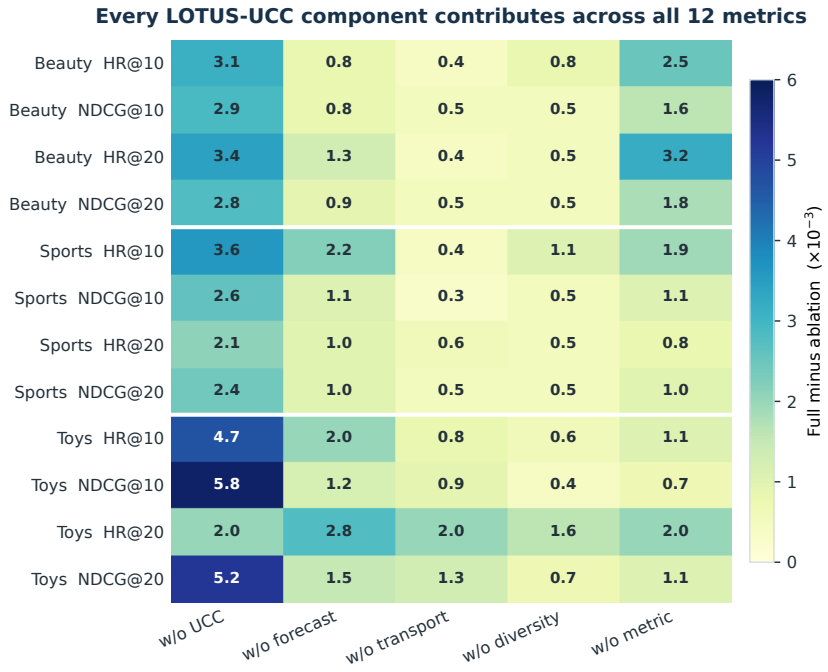

Figure 3: Metric decrease after component removal ($\times 10^3$).

Table 7: Paired bootstrap for Full minus w/o UCC (5,000 resamples).

Dataset	Metric	Mean diff.	95% CI	Two-sided p
Beauty	HR@10	0.00309	[0.00107, 0.00496]	0.003599
Beauty	NDCG@10	0.00276	[0.00173, 0.00381]	0.000400
Beauty	HR@20	0.00255	[0.00054, 0.00461]	0.018796
Beauty	NDCG@20	0.00260	[0.00163, 0.00358]	0.000400
Sports	HR@10	0.00256	[0.00135, 0.00376]	0.000400
Sports	NDCG@10	0.00184	[0.00120, 0.00246]	0.000400
Sports	HR@20	0.00174	[0.00037, 0.00312]	0.014397
Sports	NDCG@20	0.00165	[0.00104, 0.00226]	0.000400
Toys	HR@10	0.00428	[0.00237, 0.00623]	0.000400
Toys	NDCG@10	0.00528	[0.00415, 0.00644]	0.000400
Toys	HR@20	0.00201	[0.00005, 0.00392]	0.046791
Toys	NDCG@20	0.00464	[0.00359, 0.00571]	0.000400

All intervals exclude zero at the 5% level. The resampling unit is the user, preserving the paired dependence between both checkpoints on the same test cases.

E Hyperparameter Sensitivity

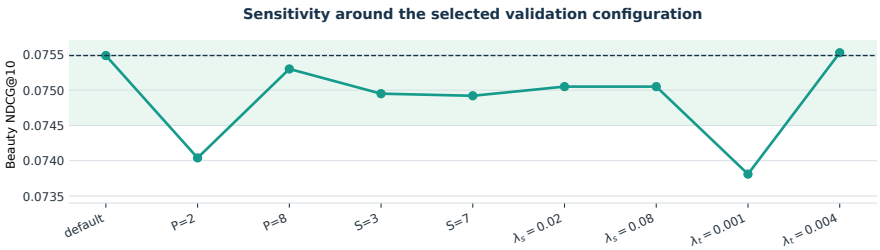

Figure 4: Beauty NDCG@10 sensitivity around the selected configuration.

The default configuration lies in a broad high-performing region. Increasing particles from 2 to 8 and varying Sinkhorn iterations from 3 to 7 produce smooth changes; semantic and transport weights exhibit the same local stability. All sensitivity choices were made with validation outcomes rather than test feedback.

F Efficiency and Exposure

Computational profile. The complete Beauty, Sports, and Toys models use 3.11M, 4.71M, and 3.06M trainable parameters and require 299.48, 398.10, and 297.65 MiB, respectively. Under full-catalog scoring, they process 1,477.24, 1,458.89, and 1,177.61 users per second. Thus, the geometric construction retains a compact trainable footprint and practical exhaustive-ranking throughput across all three catalog scales. Across datasets, the largest model remains below 5M trainable parameters while throughput stays above 1,100 users per second. This consistent operating profile shows that spherical transport, unitary phase action, and cutoff-aware learning add structured modeling capacity while retaining efficient exhaustive scoring. Appendix F reports the accompanying coverage, exposure-concentration, and tail-exposure measurements.

Table 8: Model footprint and full-catalog efficiency.

Dataset	Parameters	Users/s	Peak MiB
Beauty	3,110,275	1,477.24	299.48
Sports	4,711,811	1,458.89	398.10
Toys	3,064,963	1,177.61	297.65

Catalog exposure. For a Top-K list R_u^K, catalog size M, and exposure count $n_i = |\{u : i \in R_u^K\}|$, define

$$\text{Coverage@}K = |\cup_u R_u^K|/M, \qquad p_i = n_i/(K|\mathcal{U}|), \tag{43}$$

$$\text{Gini@}K = \frac{2\sum_{i=1}^{M} i p_{(i)}}{M\sum_i p_{(i)}} - \frac{M+1}{M}, \tag{44}$$

$$\text{TailShare@}K = \frac{\sum_u |R_u^K \cap \mathcal{I}_{\text{tail}}|}{K|\mathcal{U}|}, \qquad \text{PopLift@}K = \frac{\overline{\text{pop}}(R)}{\overline{\text{pop}}(x)}. \tag{45}$$

Here $p_{(i)}$ is exposure sorted increasingly and $\mathcal{I}_{\text{tail}}$ is the bottom 80% of items by training frequency.

Table 9: Exposure analysis. Arrows indicate the preferred direction.

Dataset	Variant	Coverage ↑	Gini ↓	Tail exp. ↑	Pop. lift ↓
Beauty	w/o UCC	0.8515	0.8045	0.1088	5.0070
	Full	**0.8670**	**0.7922**	**0.1125**	**4.8759**
Sports	w/o UCC	0.7662	0.8592	0.0758	7.3012
	Full	**0.7835**	**0.8492**	**0.0783**	**7.2497**
Toys	w/o UCC	0.8719	0.7345	0.1464	3.1366
	Full	**0.8845**	**0.7248**	**0.1483**	**3.1038**

G Additional Implementation Details

We use Qwen3-Embedding-0.6B as the semantic encoder. Language features are centered catalog-wide and normalized with $\epsilon_{\text{norm}} = 10^{-6}$. Tangent velocities use $W_v = 3$ and $\lambda = 0.8$; the local mean uses $W_m = 5$. The extrapolation scale is $\alpha = 0.75$, the geodesic radius is $r_{\text{geo}} = 1.25$, $\beta = 0.70$, and $\lambda_{\text{geo}} = 0.04$. UCC uses complex rank $r = 64$, source and target residual amplitude 0.30, particle readout geodesic coefficient $\kappa_{\text{ro}} \in (0.05, 0.40)$, phase radius at most $\pi/2$, and branch mass at most $m_{\max} = 0.55$. Particle transport uses $P = 4$, $d_p = 64$, and $S = 5$. Its learned entropy scale is $\epsilon_{\text{OT}} = 0.05 + 0.95\sigma(\eta_{\text{OT}})$, initialized with $\eta_{\text{OT}} = -1.0$; the off-diagonal penalty is $\lambda_{\text{in}} = \text{softplus}(\eta_{\text{in}})$, initialized with $\eta_{\text{in}} = -1.5$. Proposal, persistence, and readout gates lie in $(0.05, 0.50)$, $(0.70, 0.995)$, and $(0.05, 0.40)$, respectively. Complete seen-item masking is applied before Top-K extraction, and test evaluation is performed only after validation-selected training is complete.