RAP: RADIAL-ANGULAR RANDOM PROJECTION FOR HYPERBOLIC REPRESENTATIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Random projection is a basic compression primitive for Euclidean representation learning, with the Johnson–Lindenstrauss lemma giving a target dimension $k = O(\varepsilon^{-2} \log n)$ independent of the ambient dimension. As hyperbolic representations gain increasing attention in modern Transformers, vision-language models, and large language models, the same need for oblivious dimensionality reduction arises, but no analogous primitive yet exists for hyperbolic representations. We study this gap. Two natural attempts fail in dual ways: a tangent-space scheme gives a clean local guarantee but is chart-bound and pays a curvature penalty, while a fully hyperbolic projection in the hyperboloid model is geometrically native but mixes radial and angular error before feeding them into the conditioning of arcosh. We propose *radial-angular random projection* (RAP), which preserves the radial coordinate exactly and applies a Euclidean random projection only to the angular direction, lifting the result back to the target hyperboloid. RAP is hyperbolic-to-hyperbolic by construction, costs one Gaussian matrix-vector product, and reduces hyperbolic compression to a normalized angular JL question on $\mathbb{S}^{d-1}$. We prove a finite-set RAP guarantee with explicit $\sinh(\sqrt{c}\, r_x)\sinh(\sqrt{c}\, r_y)$ conditioning, complemented by a local tangent-space JL theorem and an arcosh-conditioning lemma for the fully hyperbolic comparator. Across an origin-centered broad-radius point cloud, a balanced tree, a clustered Gaussian, and the WordNet noun hierarchy, RAP attains the lowest mean relative distortion in every (n, k) cell, reducing distortion over the strongest hyperbolic baseline by roughly an order of magnitude on broad-radius data. RAP also scales to million-point datasets and improves retrieval on vision–language representations over hyperbolic projection baselines. Code is available in an anonymous repository.

1 INTRODUCTION

The Johnson–Lindenstrauss (JL) lemma states that any n-point set in Euclidean space admits a linear embedding into dimension $k = O(\varepsilon^{-2} \log n)$ with all pairwise distances preserved up to $1 \pm \varepsilon$ (Johnson & Lindenstrauss, 1984; Dasgupta & Gupta, 2003; Larsen & Nelson, 2017). The resulting random-projection primitive underlies image and text retrieval (Bingham & Mannila, 2001), approximate nearest-neighbor search (Ailon & Chazelle, 2009), randomized numerical linear algebra (Halko et al., 2011), the random-feature view of scalable attention (Choromanski et al., 2021), and more recent low-rank decomposition and projection schemes for compressing the weights and key-value caches of large language models (Saha et al., 2024; Chang et al., 2025).

Hyperbolic representations have become a standard tool wherever the data carries latent hierarchy. Poincaré and Lorentz embeddings achieve low distortion on taxonomies and trees (Nickel & Kiela, 2017; Sala et al., 2018), hyperbolic neural networks lift the standard differentiable toolkit onto the manifold (Ganea et al., 2018; Chami et al., 2019), and recent work extends this geometry to Transformers, vision-language models, and hyperbolic LLMs (Yang et al., 2024; He et al., 2025a). As these models grow, their hyperbolic embeddings become high-dimensional. JL gives Euclidean representations a generic answer to this regime, oblivious to the input set and independent of the ambient d; no analogous primitive yet exists for hyperbolic representations. The algorithmic question becomes unavoidable: *can high-dimensional hyperbolic embeddings be randomly compressed while preserving the distances that encode hierarchy?*

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.

Challenges. There are two challenges to developing a hyperbolic analogue of JL. **First**, hyperbolic space has no global linear chart, where a chart is a coordinate map from a manifold neighborhood to a Euclidean vector space; Euclidean random projections act on a single ambient vector space, but on a curved manifold a linear map only makes sense locally, inside a tangent space whose validity decays with radius. **Second**, the hyperbolic distance passes through arcosh, whose derivative behaves like $(z-1)^{-1/2}$ near $z=1$; small perturbations of Lorentz inner products are amplified at small distances and compressed at large ones, which defeats uniform multiplicative control. The two natural responses inherit these challenges in dual forms. A tangent-space scheme $\exp \circ \mathbf{P} \circ \log$ yields a clean local theorem but is chart-bound and pays an explicit cR^2 curvature penalty. A fully hyperbolic scheme that sketches spatial coordinates and renormalizes is geometrically native but mixes radial and angular error before feeding them into the ill-conditioned arcosh.

Our approach. To address these challenges, we propose *radial-angular random projection* (RAP), an oblivious hyperbolic-to-hyperbolic random projection that decomposes each point in polar form, preserves the radial coordinate exactly, and applies a Euclidean random projection only to the angular direction. Concretely, around a chosen origin $o \in \mathbb{H}^d_c$, every point factors as $x = (r_x, \mathbf{u}_x)$ with $r_x = d_c(o,x) \geq 0$ and $\mathbf{u}_x \in \mathbb{S}^{d-1}$, and the hyperbolic distance reads off through the law of cosines as a function of $(r_x, r_y, \langle \mathbf{u}_x, \mathbf{u}_y \rangle)$. Given a Gaussian matrix $\mathbf{P} \in \mathbb{R}^{k \times d}$, RAP keeps r_x unchanged, sketches $\mathbf{u}_x$ as $\mathbf{P}\mathbf{u}_x$, renormalizes the result on $\mathbb{S}^{k-1}$, and lifts the polar pair back to $\mathbb{H}^k_c$. The design is geometrically motivated: in a hyperbolic embedding of a hierarchy, r_x carries hierarchical depth and $\mathbf{u}_x$ carries within-level similarity, so confining compression to the angular factor preserves hierarchy by construction. Computationally, RAP requires one Gaussian matrix-vector product followed by a normalization, and it reduces the hyperbolic compression problem to a classical Euclidean JL question on unit vectors in $\mathbb{S}^{d-1}$. The hyperbolic law of cosines supplies the starting identity; the contribution is the oblivious normalized-angular map together with a simultaneous finite-set guarantee.

Contributions. The contributions of this paper are summarized as follows.

- We introduce RAP, an oblivious hyperbolic-to-hyperbolic random projection that preserves the radial coordinate exactly and reduces hyperbolic compression to a normalized Euclidean JL question on unit angular vectors in $\mathbb{S}^{d-1}$. The map requires a single Gaussian matrix-vector product followed by a normalization, and lands on the target hyperboloid by construction.
- We prove a finite-set RAP guarantee under bounded radii and pairwise separation, including hyperboloid validity, exact radial preservation, and an explicit $\sinh(\sqrt{c}\, r_x)\sinh(\sqrt{c}\, r_y)$ conditioning factor that governs how angular error transfers to hyperbolic distance. As complementary calibration, we establish a local tangent-space JL theorem with the explicit envelope $1 \pm \varepsilon \pm O(cR^2)$, and an arcosh-conditioning lemma for nearby pairs in the fully hyperbolic comparator.
- On four datasets (an origin-centered broad-radius point cloud, a balanced tree, a clustered Gaussian, and a WordNet noun subtree) and across $n \in \{128, 256, 512\}$ and $k \in \{16, 32, 64, 128\}$, RAP attains the lowest mean relative distortion in every (n,k) cell. RAP also reduces worst-pair distortion and scales to million-point datasets. On real-world vision–language representations, it improves macro-averaged retrieval scores by 5.5%–20.1% relative to the stronger of the two hyperbolic baselines while maintaining competitive classification accuracy.

2 Related Work

Hyperbolic representation learning. Poincaré embeddings first established hyperbolic geometry as a low-distortion host for taxonomies and tree-structured data (Nickel & Kiela, 2017; Sala et al., 2018). The differentiable toolkit on top of this foundation, including log/exp maps, Möbius operations, gyrovector arithmetic, and Riemannian optimization, was developed by hyperbolic neural networks (Ganea et al., 2018; Shimizu et al., 2021), extended to graph convolution (Chami et al., 2019), image classification (Khrulkov et al., 2020), and word and sequence embeddings (Tifrea et al., 2019). Probabilistic and optimization infrastructure has grown alongside it: wrapped and Lorentzian distributions support uncertainty modelling and metric learning (Nagano et al., 2019; Law et al., 2019), and Fréchet-mean differentiation closes the gap for objectives that need a hyperbolic centroid (Lou et al., 2020). A complementary, increasingly dominant line operates natively in the Lorentz model rather than passing through repeated tangent charts: fully hyperbolic neural networks and CNNs (Chen et al., 2022; Bdeir et al., 2024), hyperbolic Transformers (Yang et al., 2024), vision-language

and multimodal hyperbolic representations (Desai et al., 2023; Pal et al., 2025; Peng et al., 2025), hyperbolic LLMs and adaptation (He et al., 2025a; Yang et al., 2025), hyperbolic foundation-model libraries (He et al., 2025b), retrieval-augmented generation with hyperbolic indexes (Madhu et al., 2026), and recent hyperbolic vision Transformers and Busemann-style efficient layers (Alyoussef et al., 2026; van der Klis et al., 2026; Chen et al., 2026). All of these design *trainable* layers and end-to-end architectures; our contribution is orthogonal: an oblivious random-projection primitive that compresses learned hyperbolic embeddings without retraining.

Hyperbolic dimensionality reduction. Existing methods for reducing the dimension of hyperbolic representations are deterministic and data-dependent. Principal geodesic analysis (Fletcher et al., 2004) is the classical Riemannian generalization of PCA: it tangent-linearizes the data at a base point and applies Euclidean PCA in that chart. Nested hyperbolic spaces (Fan et al., 2022) parameterize a family of low-dimensional hyperbolic submanifolds and learn a projection onto the closest one, treating dimension reduction as a trainable component of the model. HoroPCA (Chami et al., 2021) formulates hyperbolic principal directions through horospherical projections in the Poincaré ball, producing a faithful hyperbolic-to-hyperbolic reduction that respects the manifold's boundary structure. Representation-tradeoff analyses for hyperbolic embeddings of trees and other discrete structures (Sala et al., 2018) clarify the achievable tradeoff between target dimension and worst-case distortion in the curved setting. None of these constructions is data-oblivious or carries a JL-style finite-set distance preservation guarantee, so they answer a different question than the one we study: they fit a low-dimensional hyperbolic representation to a particular dataset, whereas we ask for a single random map that preserves pairwise distances on any sufficiently regular finite point set. We therefore exclude these methods from the empirical comparison in Section 5; Appendix D details the distinction in problem statement and tabulates how HoroPCA, nested hyperbolic spaces, and RAP relate.

Classical JL and modern random sketching. Since Johnson & Lindenstrauss (1984), Euclidean random projections have accumulated a mature toolbox: streamlined probabilistic proofs (Dasgupta & Gupta, 2003), database-friendly and sparse variants (Achlioptas, 2003; Matoušek, 2008), fast linear transforms (Ailon & Chazelle, 2009), feature-hashing applications (Weinberger et al., 2009), and tight $k = \Theta(\varepsilon^{-2} \log n)$ lower bounds (Larsen & Nelson, 2017), with a broader metric-embedding context tracing back to Bourgain (1985). Random projection underlies image and text retrieval (Bingham & Mannila, 2001) and randomized numerical linear algebra (Halko et al., 2011); the random-feature view of attention provides scalable Transformer approximations (Choromanski et al., 2021), and recent work compresses the weights and key-value caches of large language models with low-rank/low-precision decompositions and low-rank projection schemes (Saha et al., 2024; Chang et al., 2025). Despite this breadth, every line in this literature is built on a global linear chart and Euclidean inner products, so its derivations do not transfer verbatim to a curved metric space; obtaining an oblivious random-projection primitive for hyperbolic representations therefore requires a construction tailored to the geometry, which is the gap the present paper closes.

3 Preliminaries

Let $\mathbb{H}^d_c$ denote d-dimensional hyperbolic space of constant curvature $-c < 0$, with hyperbolic distance $d_c(\cdot,\cdot)$, exponential map $\exp_q : T_a\mathbb{H}^d_c \to \mathbb{H}^d_c$, and logarithmic map $\log_a$ at $a \in \mathbb{H}^d_c$. The geometry is intrinsic: we instantiate $\mathbb{H}^d_c$ either as the Poincaré ball or the Lorentz hyperboloid when an explicit coordinate computation is needed, and the choice between the two does not affect distances or the maps $\exp_a, \log_a$ (Appendix A). We will use the polar form, the local Euclideanization through normal coordinates, and a standard distributional JL guarantee; the three together feed both the tangent-space comparator and RAP. A complete notation table is given in Appendix B.

Polar coordinates. Around an origin or anchor $o \in \mathbb{H}^d_c$, every point admits the polar representation $x = (r_x, \mathbf{u}_x)$ with $r_x = d_c(o,x) \geq 0$ and $\mathbf{u}_x \in \mathbb{S}^{d-1}$. The hyperbolic law of cosines reads

$$\cosh(\sqrt{c}\, d_c(x,y)) = \cosh(\sqrt{c}\, r_x)\cosh(\sqrt{c}\, r_y) - \sinh(\sqrt{c}\, r_x)\sinh(\sqrt{c}\, r_y)\langle \mathbf{u}_x, \mathbf{u}_y\rangle. \tag{1}$$

Here r_x encodes hierarchical depth and $\langle \mathbf{u}_x, \mathbf{u}_y\rangle$ encodes within-level similarity.

Local Euclideanization. In geodesic normal coordinates centered at a, the metric admits

$$g_{\mathbf{u}} = I + O(c\|\mathbf{u}\|^2), \tag{2}$$

uniformly on small balls. Concretely, for any $\mathbf{u}, \mathbf{v} \in T_a\mathbb{H}^d_c$ with $\|\mathbf{u}\|, \|\mathbf{v}\| \le R$ and $R\sqrt{c}$ small,

$$d_c(\exp_a \mathbf{u}, \exp_a \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|_2 \left(1 + O(cR^2)\right), \tag{3}$$

so on a sufficiently small ball the hyperbolic distance agrees with the Euclidean tangent distance up to a curvature correction. Equation (2) is what makes the tangent-space scheme analyzable, and it is also what confines that scheme to small balls.

Euclidean JL matrix. A standard JL matrix is $\mathbf{P} \in \mathbb{R}^{k\times d}$ with i.i.d. $\mathcal{N}(0, 1/k)$ entries. For any finite vector set $\{\mathbf{u}_i\}_{i=1}^n$, with $k = O(\varepsilon^{-2}\log(n/\delta))$,

$$(1-\varepsilon)\|\mathbf{u}_i - \mathbf{u}_j\| \le \|\mathbf{P}\mathbf{u}_i - \mathbf{P}\mathbf{u}_j\| \le (1+\varepsilon)\|\mathbf{u}_i - \mathbf{u}_j\| \quad \forall i, j \tag{4}$$

holds with probability at least $1 - \delta$. Tangent-chart projections in §4.2 are stated through orthonormal coordinate maps $U_a : T_a\mathbb{H}^d_c \to \mathbb{R}^d$ and $U_{\tilde{a}} : T_{\tilde{a}}\mathbb{H}^k_c \to \mathbb{R}^k$; the resulting maps are basis-independent.

4 Methodology

4.1 Problem Definition

Let $X = \{x_1, \dots, x_n\} \subset \mathbb{H}^d_c$ and let $k < d$ be a target dimension. We seek a randomized map $\Phi : \mathbb{H}^d_c \to \mathbb{H}^k_c$ such that

$$\Pr_{\Phi}\left[\max_{i<j}\left|\frac{d_c(\Phi(x_i), \Phi(x_j))}{d_c(x_i, x_j)} - 1\right| \le \varepsilon\right] \ge 1 - \delta, \tag{5}$$

with k growing polylogarithmically in n. We denote the worst-pair distortion of a realized map by

$$\Delta_X(\Phi) = \max_{i<j}\left|\frac{d_c(\Phi(x_i), \Phi(x_j))}{d_c(x_i, x_j)} - 1\right|. \tag{6}$$

Equation (5) is intentionally stronger than what is provable globally on a curved manifold, and the three schemes we study relax it in different ways. The tangent-space scheme delivers the bound only inside a small geodesic ball, paying a cR^2 curvature term outside the JL term itself. The fully hyperbolic scheme is global on the manifold but loses a clean multiplicative bound near coincident pairs because of the arcosh distance map. RAP keeps the radial coordinate exact, so Equation (5) reduces to a finite-set angular JL bound on $\mathbb{S}^{d-1}$ multiplied by an explicit hyperbolic conditioning factor; this is what we eventually prove in Appendix J. Figure 1 gives a unified schematic of these three design choices before we define them formally.

4.2 Vanilla Method I: Tangent-Space-Based Projection

Given anchors $a \in \mathbb{H}^d_c$ and $\tilde{a} \in \mathbb{H}^k_c$ and a Gaussian JL matrix $\mathbf{P} \in \mathbb{R}^{k\times d}$, the tangent-space scheme is

$$\Pi_{a,\mathbf{P}}(x) = \exp_{\tilde{a}}\left(U_{\tilde{a}}^{-1}\mathbf{P}U_a \log_a(x)\right). \tag{7}$$

The construction is local-by-design: $\log_a$ supplies a Euclidean chart, JL acts there, and $\exp_{\tilde{a}}$ returns to the manifold. Its guarantee inherits both this locality and a curvature penalty from Equation (2). We state it as a calibrated reference point: it shows that, inside a small enough ball, the standard JL distortion bound transfers to hyperbolic distances up to an explicit $O(cR^2)$ correction.

Theorem 1 (Local tangent-space JL). *Let $X = \{x_1, \dots, x_n\} \subset B(a, R) \subset \mathbb{H}^d_c$ with $R \le R_0/\sqrt{c}$. Let $\mathbf{P}$ be a Gaussian JL matrix with $k \ge C\varepsilon^{-2}\log(n/\delta)$, and assume the projected tangent cloud lies in a ball where the same comparison is valid. Then with probability at least $1 - \delta$, for all i, j,*

$$(1 - \varepsilon - C_1 cR^2)\, d_c(x_i, x_j) \le d_c(\Pi_{a,\mathbf{P}}x_i, \Pi_{a,\mathbf{P}}x_j) \le (1 + \varepsilon + C_1 cR^2)\, d_c(x_i, x_j). \tag{8}$$

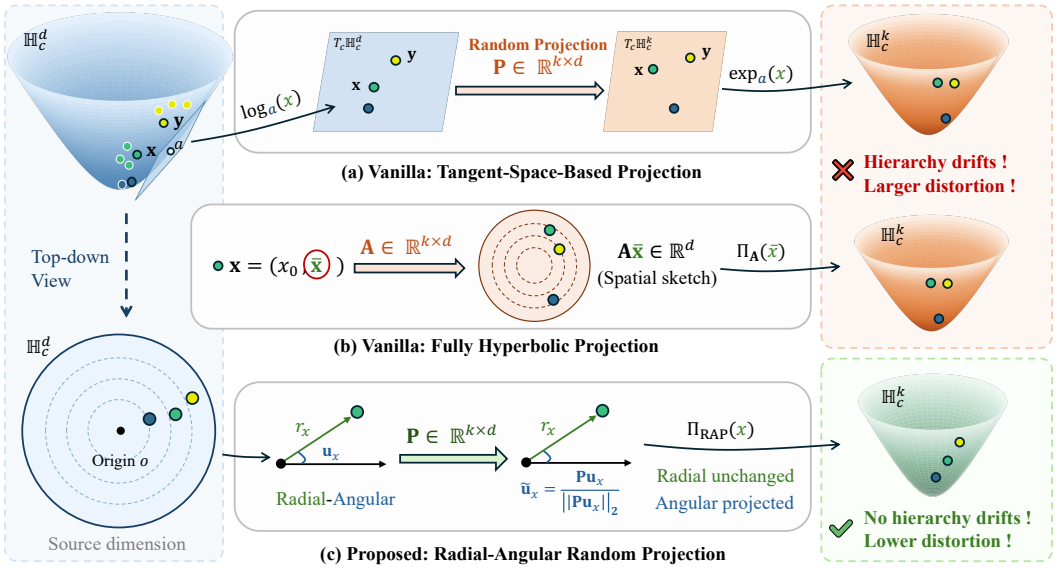


Figure 1: Unified schematic of the three hyperbolic random projection paradigms studied in this paper. Panel (a) shows the tangent-space baseline: points are sent to a Euclidean tangent chart by $\log_a$, sketched by a Gaussian JL matrix, and mapped back by an exponential map. Panel (b) shows the fully hyperbolic baseline: spatial coordinates are projected in the hyperboloid model and the time coordinate is recomputed, which keeps the output on the hyperboloid but couples radial and angular perturbations. Panel (c) shows RAP, our proposed radial-angular projection: the radius r_x is preserved exactly while only the angular direction $\mathbf{u}_x$ is projected and renormalized. This separation keeps hierarchy-carrying radial depth unchanged and confines compression error to angular information.

The two error terms decouple. The ε term is the standard JL contribution and shrinks with the target dimension k; the $C_1 cR^2$ term is the curvature contribution and does not vanish with k. The constants R_0 and C_1 are universal: R_0 sets the locality range of the source ball (we take $R\sqrt{c} \le 1/2$), and $C_1 = 5C_0$ absorbs the source-side, target-side, and cross-term curvature errors composed in Appendix H, where the full proof is given. The locality of the bound is structural rather than an artifact of the proof: a uniform multiplicative bound on an unbounded geodesic ball would have to absorb the cR^2 term as $R \to \infty$, which the metric expansion in Equation (2) rules out. The local form is therefore the natural statement for any tangent-chart projection in negatively curved space.

4.3 Vanilla Method II: Fully Hyperbolic Projection

In the hyperboloid model a point is $x = (x_0, \bar{\mathbf{x}}) \in \mathbb{R}^{d+1}$ with time coordinate x_0 and spatial part $\bar{\mathbf{x}} \in \mathbb{R}^d$ satisfying $-x_0^2 + \|\bar{\mathbf{x}}\|_2^2 = -1/c$. A natural attempt at a hyperbolic-to-hyperbolic map is to sketch only the d spatial coordinates by $\mathbf{A} \in \mathbb{R}^{k\times d}$ and recompute the time coordinate from the hyperboloid constraint:

$$\Pi_{\mathbf{A}}(x) = \Big(\sqrt{1/c + \|\mathbf{A}\bar{\mathbf{x}}\|_2^2},\ \mathbf{A}\bar{\mathbf{x}}\Big). \tag{9}$$

The renormalization is forced: any $\mathbf{A}\bar{\mathbf{x}}$ alone leaves the hyperboloid, so recomputing the time coordinate returns the output to $\mathbb{H}_c^k$. The resulting map is intrinsic and chart-free, but distance is recovered through

$$d_c(x,y) = \tfrac{1}{\sqrt{c}} \operatorname{arcosh}\big(-c\langle x, y\rangle_L\big), \qquad \langle x, y\rangle_L = -x_0 y_0 + \langle \bar{\mathbf{x}}, \bar{\mathbf{y}}\rangle, \tag{10}$$

and $\operatorname{arcosh}'(z) \sim (z-1)^{-1/2}$ as $z \to 1^+$. Two effects follow. First, the renormalized time coordinate depends on the perturbed spatial norm $\|\mathbf{A}\bar{\mathbf{x}}\|_2$, so a single Euclidean sketch perturbs both radial magnitude and angular correlation simultaneously (Lemma 10, Appendix I). Second, the arcosh derivative amplifies inner-product perturbations on nearby pairs and damps them on distant ones, so a uniform multiplicative distance bound is unattainable across all distance scales.

4.4 Proposed Method: Radial-Angular Random Projection (RAP)

The hyperbolic law of cosines (Equation 1) writes every pairwise distance as a function of the two radii and the angular inner product. Preserving the radii exactly therefore reduces hyperbolic JL to controlling the angular inner product on $\mathbb{S}^{d-1}$, where standard Euclidean JL applies.

Fix an anchor $o \in \mathbb{H}^d_c$ and write $x = (r_x, \mathbf{u}_x)$ with $r_x = d_c(o, x)$ and $\mathbf{u}_x \in \mathbb{S}^{d-1}$. For a Gaussian $\mathbf{P} \in \mathbb{R}^{k\times d}$,

$$\widetilde{\mathbf{u}}_x = \frac{\mathbf{P}\mathbf{u}_x}{\|\mathbf{P}\mathbf{u}_x\|_2}, \qquad \Pi_{\mathrm{RAP}}(x) = \iota_k(r_x, \widetilde{\mathbf{u}}_x), \tag{11}$$

where ι_k is the polar parameterization of $\mathbb{H}^k_c$ around its canonical origin $\tilde{o} = (1/\sqrt{c}, \mathbf{0})$:

$$\iota_k(r, \mathbf{v}) = \frac{1}{\sqrt{c}}\big(\cosh(\sqrt{c}\, r),\, \sinh(\sqrt{c}\, r)\, \mathbf{v}\big), \qquad \mathbf{v} \in \mathbb{S}^{k-1}. \tag{12}$$

Computationally, RAP is one Gaussian matrix-vector product plus a renormalization on $\mathbb{S}^{k-1}$; no exponential map is needed at the target side.

Substituting Equation (11) into Equation (1), the angular-inner-product error transfers to the distance with conditioning weight $\sinh(\sqrt{c}r_x)\sinh(\sqrt{c}r_y)$.

Proposition 2 (RAP validity and angular reduction)**.** *Assume $\|\mathbf{P}\mathbf{u}_x\|_2 > 0$ for the non-anchor points considered. Then $\Pi_{\mathrm{RAP}}(x) \in \mathbb{H}^k_c$ and $d_c(o, \Pi_{\mathrm{RAP}}(x)) = r_x$. If $\mathbf{u}_x = \mathbf{u}_y$ then $d_c(\Pi_{\mathrm{RAP}}(x), \Pi_{\mathrm{RAP}}(y)) = |r_x - r_y| = d_c(x, y)$. For any pair x, y, if $|\langle\widetilde{\mathbf{u}}_x, \widetilde{\mathbf{u}}_y\rangle - \langle\mathbf{u}_x, \mathbf{u}_y\rangle| \le \eta$, then the corresponding perturbation of the law-of-cosines argument in Equation (1) is at most $\eta\, \sinh(\sqrt{c}r_x)\sinh(\sqrt{c}r_y)$.*

The RAP problem thus reduces to a normalized angular JL question on $\mathbb{S}^{d-1}$: under what conditions does $|\langle\widetilde{\mathbf{u}}_x, \widetilde{\mathbf{u}}_y\rangle - \langle\mathbf{u}_x, \mathbf{u}_y\rangle| \le \eta$ hold uniformly over a finite set of unit vectors? A standard JL norm-preservation event for the angular vectors and their pairwise differences is enough (Lemma 14). Combining this with the $\sinh(\sqrt{c}r_x)\sinh(\sqrt{c}r_y)$ factor and the arcosh inversion yields the finite-set RAP guarantee in Appendix J, conditioned on a bounded radius $r_i \le R$ and a pairwise separation $d_c(x_i, x_j) \ge \tau > 0$. The same appendix records the proof of Proposition 2.

Compared with the tangent scheme, RAP needs no small-ball restriction; compared with the fully hyperbolic scheme, RAP eliminates the radial-mixing term and only inherits the arcosh conditioning. Appendix K tabulates the three error envelopes and records the structural source of each term.

5 Experiments

Setup. We compare four projection rules: the tangent-space vanilla method (Section 4.2), the fully hyperbolic vanilla method (Section 4.3), RAP (Section 4.4), and a Euclidean JL diagnostic that sketches origin-log coordinates and measures Euclidean output distance against the original hyperbolic distance (Appendix C); the diagnostic is intentionally cross-geometry and is included only as a sanity check that a flattened Euclidean sketch is not a substitute for a hyperbolic-to-hyperbolic map. The two vanilla methods are not adversarially chosen: they instantiate the two design patterns that dominate trainable hyperbolic architectures (Sections 4.2 and 4.3). We evaluate on four datasets: an origin-centered *broad-radius* hyperbolic point cloud in ambient dimension 128, a balanced 4-ary depth-5 *tree* in $\mathbb{H}^{50}$, a *clustered Gaussian* dataset of eight hyperbolic components in $\mathbb{H}^{50}$, and the *WordNet* noun subtree rooted at `animal.n.01` (Miller, 1995); WordNet is the canonical hyperbolic-embedding benchmark introduced by Nickel & Kiela (2017) and is used here as a standard real-data probe in line with established practice in hyperbolic representation learning. The synthetic datasets sweep $n \in \{128, 256, 512\}$ and $k \in \{16, 32, 64, 128\}$ over 20 random seeds, with curvature fixed to $c = 1$. We report the mean relative distortion over nonzero-distance pairs; lower values indicate output distances closer to the original $d_c(x, y)$. Detailed dataset construction, the paired-comparison protocol, and the choice of mean over worst-pair distortion are detailed in Appendix E.

5.1 Main Results

Tables 1 and 2 report mean relative distortion across the four datasets and the full (n, k) sweep. RAP attains the lowest mean distortion in every (n, k) cell. The margin is roughly an order of magnitude

Table 1: Mean relative distortion across synthetic benchmarks, reported as mean±std. Lower is better; bold marks the best mean in each dataset–n block. The Euclidean JL baseline applies Gaussian JL in origin-log coordinates and compares Euclidean output to the original hyperbolic distances.

n	Scheme	$k = 16$	$k = 32$	$k = 64$	$k = 128$
Broad-radius benchmark					
128	Euclidean JL	0.375 ± 0.030	0.372 ± 0.022	0.365 ± 0.015	0.356 ± 0.011
	Local tangent-space	0.392 ± 0.034	0.405 ± 0.026	0.420 ± 0.018	0.433 ± 0.010
	Fully hyperbolic	0.232 ± 0.015	0.244 ± 0.015	0.266 ± 0.011	0.288 ± 0.009
	RAP (Ours)	$\mathbf{0.054} \pm 0.002$	$\mathbf{0.035} \pm 0.002$	$\mathbf{0.022} \pm 0.001$	$\mathbf{0.014} \pm 0.001$
	Δ impr.	$\mathbf{85.6\%} \uparrow$	$\mathbf{90.6\%} \uparrow$	$\mathbf{94.0\%} \uparrow$	$\mathbf{96.1\%} \uparrow$
256	Euclidean JL	0.256 ± 0.012	0.235 ± 0.008	0.217 ± 0.008	0.197 ± 0.008
	Local tangent-space	0.258 ± 0.014	0.244 ± 0.010	0.235 ± 0.010	0.230 ± 0.010
	Fully hyperbolic	0.169 ± 0.005	0.159 ± 0.005	0.156 ± 0.006	0.157 ± 0.006
	RAP (Ours)	$\mathbf{0.056} \pm 0.002$	$\mathbf{0.037} \pm 0.001$	$\mathbf{0.025} \pm 0.001$	$\mathbf{0.016} \pm 0.001$
	Δ impr.	$\mathbf{78.1\%} \uparrow$	$\mathbf{84.3\%} \uparrow$	$\mathbf{88.5\%} \uparrow$	$\mathbf{91.9\%} \uparrow$
512	Euclidean JL	0.202 ± 0.007	0.172 ± 0.006	0.151 ± 0.005	0.130 ± 0.003
	Local tangent-space	0.199 ± 0.009	0.171 ± 0.007	0.153 ± 0.006	0.139 ± 0.006
	Fully hyperbolic	0.140 ± 0.004	0.118 ± 0.003	0.105 ± 0.003	0.096 ± 0.003
	RAP (Ours)	$\mathbf{0.056} \pm 0.001$	$\mathbf{0.039} \pm 0.001$	$\mathbf{0.026} \pm 0.001$	$\mathbf{0.018} \pm 0.000$
	Δ impr.	$\mathbf{72.3\%} \uparrow$	$\mathbf{77.3\%} \uparrow$	$\mathbf{82.8\%} \uparrow$	$\mathbf{86.2\%} \uparrow$
Balanced tree					
128	Euclidean JL	0.146 ± 0.013	0.105 ± 0.009	0.075 ± 0.006	0.058 ± 0.005
	Local tangent-space	0.142 ± 0.010	0.099 ± 0.006	0.069 ± 0.004	0.049 ± 0.002
	Fully hyperbolic	0.133 ± 0.010	0.093 ± 0.006	0.064 ± 0.004	0.046 ± 0.002
	RAP (Ours)	$\mathbf{0.102} \pm 0.004$	$\mathbf{0.071} \pm 0.004$	$\mathbf{0.049} \pm 0.002$	$\mathbf{0.035} \pm 0.001$
	Δ impr.	$\mathbf{30.1\%} \uparrow$	$\mathbf{32.4\%} \uparrow$	$\mathbf{34.7\%} \uparrow$	$\mathbf{39.7\%} \uparrow$
256	Euclidean JL	0.144 ± 0.011	0.104 ± 0.008	0.074 ± 0.006	0.058 ± 0.005
	Local tangent-space	0.140 ± 0.009	0.099 ± 0.006	0.069 ± 0.004	0.049 ± 0.002
	Fully hyperbolic	0.131 ± 0.009	0.092 ± 0.005	0.064 ± 0.004	0.046 ± 0.002
	RAP (Ours)	$\mathbf{0.102} \pm 0.005$	$\mathbf{0.071} \pm 0.005$	$\mathbf{0.049} \pm 0.003$	$\mathbf{0.035} \pm 0.001$
	Δ impr.	$\mathbf{29.2\%} \uparrow$	$\mathbf{31.7\%} \uparrow$	$\mathbf{33.8\%} \uparrow$	$\mathbf{39.7\%} \uparrow$
512	Euclidean JL	0.144 ± 0.011	0.104 ± 0.008	0.074 ± 0.007	0.057 ± 0.005
	Local tangent-space	0.139 ± 0.009	0.098 ± 0.006	0.068 ± 0.004	0.049 ± 0.002
	Fully hyperbolic	0.131 ± 0.008	0.092 ± 0.005	0.064 ± 0.003	0.046 ± 0.002
	RAP (Ours)	$\mathbf{0.102} \pm 0.004$	$\mathbf{0.071} \pm 0.004$	$\mathbf{0.050} \pm 0.003$	$\mathbf{0.036} \pm 0.002$
	Δ impr.	$\mathbf{29.2\%} \uparrow$	$\mathbf{31.7\%} \uparrow$	$\mathbf{32.4\%} \uparrow$	$\mathbf{36.8\%} \uparrow$
Clustered Gaussian					
128	Euclidean JL	0.152 ± 0.024	0.112 ± 0.020	0.092 ± 0.018	0.076 ± 0.013
	Local tangent-space	0.135 ± 0.024	0.096 ± 0.018	0.072 ± 0.013	0.050 ± 0.009
	Fully hyperbolic	0.117 ± 0.017	0.083 ± 0.013	0.060 ± 0.009	0.041 ± 0.007
	RAP (Ours)	$\mathbf{0.090} \pm 0.007$	$\mathbf{0.063} \pm 0.006$	$\mathbf{0.044} \pm 0.005$	$\mathbf{0.031} \pm 0.003$
	Δ impr.	$\mathbf{40.8\%} \uparrow$	$\mathbf{43.8\%} \uparrow$	$\mathbf{52.2\%} \uparrow$	$\mathbf{59.2\%} \uparrow$
256	Euclidean JL	0.153 ± 0.025	0.113 ± 0.020	0.092 ± 0.017	0.076 ± 0.012
	Local tangent-space	0.135 ± 0.025	0.096 ± 0.018	0.071 ± 0.012	0.050 ± 0.009
	Fully hyperbolic	0.118 ± 0.018	0.083 ± 0.013	0.060 ± 0.009	0.041 ± 0.006
	RAP (Ours)	$\mathbf{0.091} \pm 0.008$	$\mathbf{0.063} \pm 0.006$	$\mathbf{0.044} \pm 0.004$	$\mathbf{0.031} \pm 0.003$
	Δ impr.	$\mathbf{40.5\%} \uparrow$	$\mathbf{44.2\%} \uparrow$	$\mathbf{52.2\%} \uparrow$	$\mathbf{59.2\%} \uparrow$
512	Euclidean JL	0.152 ± 0.024	0.113 ± 0.019	0.092 ± 0.017	0.076 ± 0.013
	Local tangent-space	0.134 ± 0.024	0.096 ± 0.018	0.071 ± 0.012	0.050 ± 0.009
	Fully hyperbolic	0.117 ± 0.017	0.083 ± 0.013	0.060 ± 0.009	0.041 ± 0.007
	RAP (Ours)	$\mathbf{0.091} \pm 0.007$	$\mathbf{0.063} \pm 0.006$	$\mathbf{0.044} \pm 0.004$	$\mathbf{0.031} \pm 0.003$
	Δ impr.	$\mathbf{40.1\%} \uparrow$	$\mathbf{44.2\%} \uparrow$	$\mathbf{52.2\%} \uparrow$	$\mathbf{59.2\%} \uparrow$

Note. Standard deviations are computed across 20 seeds. For broad-radius data, each seed first averages over radial scales $\{0.5, 1.0, 2.0\}$. Broad-radius data use ambient dimension 128; balanced-tree and clustered data use ambient dimension 50. Δ impr. denotes the relative distortion reduction of RAP over Euclidean JL, $100(D_{\mathrm{EJL}} - D_{\mathrm{RAP}})/D_{\mathrm{EJL}}$; green upward arrows indicate larger reductions.

on the broad-radius benchmark (e.g., n=512, k=32: RAP 0.039 vs fully hyperbolic 0.118) and a stable 20–30% on structured datasets (e.g., RAP 0.071 vs fully hyperbolic 0.094 on WordNet).

The non-RAP rows order themselves consistently: fully hyperbolic > tangent-space > Euclidean diagnostic. The tangent-space penalty is visible on broad-radius data and softens on the structured datasets, where most pairs lie in a comparatively narrow ball. The Euclidean diagnostic stays uniformly weakest, confirming that flattening followed by Euclidean compression of log-coordinates does not track hyperbolic distance.

Table 2: Mean relative distortion on the WordNet subtree rooted at `animal.n.01` as subtree depth and target dimension vary, reported as mean±std over 20 random seeds. Lower is better; bold marks the best mean in each combination of depth and target dimension. Euclidean JL uses the same origin-log baseline as Table 1.

Depth	n	Scheme	$k = 16$	$k = 32$	$k = 64$	$k = 128$
3	273	Euclidean JL	0.142 ± 0.010	0.102 ± 0.005	0.072 ± 0.005	0.052 ± 0.003
		Local tangent-space	0.141 ± 0.010	0.100 ± 0.006	0.071 ± 0.006	0.050 ± 0.004
		Fully hyperbolic	0.136 ± 0.009	0.097 ± 0.005	0.068 ± 0.005	0.048 ± 0.003
		RAP (Ours)	$\mathbf{0.099} \pm 0.004$	$\mathbf{0.070} \pm 0.002$	$\mathbf{0.049} \pm 0.002$	$\mathbf{0.034} \pm 0.002$
		Δ impr.	**30.3%** ↑	**31.4%** ↑	**31.9%** ↑	**34.6%** ↑
4	512	Euclidean JL	0.142 ± 0.011	0.103 ± 0.008	0.072 ± 0.007	0.053 ± 0.004
		Local tangent-space	0.140 ± 0.010	0.100 ± 0.008	0.070 ± 0.007	0.050 ± 0.004
		Fully hyperbolic	0.133 ± 0.009	0.095 ± 0.007	0.067 ± 0.006	0.047 ± 0.004
		RAP (Ours)	$\mathbf{0.099} \pm 0.004$	$\mathbf{0.069} \pm 0.002$	$\mathbf{0.048} \pm 0.002$	$\mathbf{0.034} \pm 0.002$
		Δ impr.	**30.3%** ↑	**33.0%** ↑	**33.3%** ↑	**35.8%** ↑
5	512	Euclidean JL	0.144 ± 0.012	0.104 ± 0.011	0.072 ± 0.008	0.055 ± 0.005
		Local tangent-space	0.140 ± 0.010	0.100 ± 0.009	0.069 ± 0.006	0.049 ± 0.004
		Fully hyperbolic	0.133 ± 0.009	0.094 ± 0.007	0.065 ± 0.006	0.046 ± 0.004
		RAP (Ours)	$\mathbf{0.102} \pm 0.004$	$\mathbf{0.071} \pm 0.004$	$\mathbf{0.049} \pm 0.003$	$\mathbf{0.034} \pm 0.002$
		Δ impr.	**29.2%** ↑	**31.7%** ↑	**31.9%** ↑	**38.2%** ↑
6	512	Euclidean JL	0.146 ± 0.014	0.106 ± 0.012	0.073 ± 0.009	0.058 ± 0.006
		Local tangent-space	0.141 ± 0.011	0.099 ± 0.009	0.069 ± 0.006	0.048 ± 0.004
		Fully hyperbolic	0.132 ± 0.011	0.092 ± 0.007	0.064 ± 0.005	0.045 ± 0.003
		RAP (Ours)	$\mathbf{0.103} \pm 0.005$	$\mathbf{0.070} \pm 0.004$	$\mathbf{0.049} \pm 0.003$	$\mathbf{0.034} \pm 0.002$
		Δ impr.	**29.5%** ↑	**34.0%** ↑	**32.9%** ↑	**41.4%** ↑

Note. We use the WordNet noun hierarchy rooted at `animal.n.01`. Depth is the maximum hyponym depth retained before deterministic subsampling; n is the number of evaluated nodes after capping each subtree at 512 nodes. Δ impr. denotes the relative distortion reduction of RAP over Euclidean JL.

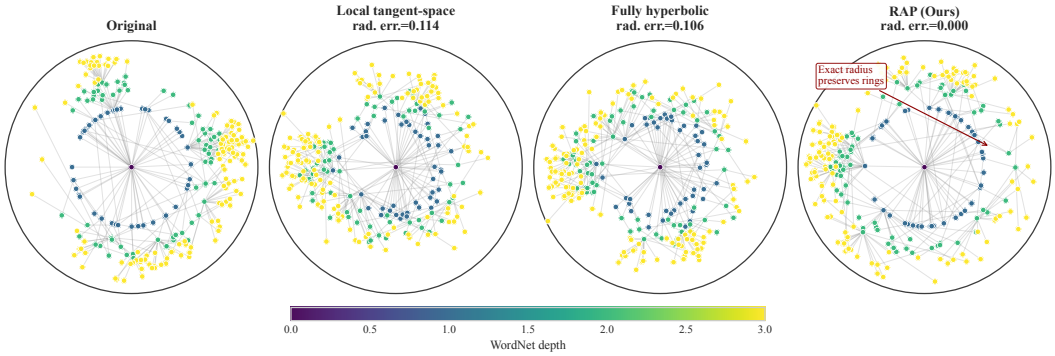


Figure 2: Qualitative WordNet subtree visualization at $k = 32$. RAP keeps the depth rings closest to the original view because its radial coordinate is preserved exactly, whereas the tangent-space and fully hyperbolic comparators exhibit visible radial drift.

5.2 Analysis and Discussion

RAP preserves the visible hierarchy structure. Figure 2 shows that RAP preserves the depth-separated radial structure of the original WordNet embeddings, whereas the two vanilla hyperbolic projections blur this structure through radial drift. This reflects RAP's exact preservation of radial coordinates; visualization details and per-node measurements are provided in Appendix F.2.

Distance-bin results support the predicted conditioning effect. Lemma 9 predicts greater sensitivity for nearby pairs, where $-c\langle x, y\rangle_L \approx 1$. Figure 3 groups nonzero pairs by original distance on broad-radius data at $k = 32$: the fully hyperbolic comparator's mean distortion falls from 0.186 in the nearest 20% to 0.062 in the farthest 20%. Euclidean JL and the tangent-space comparator remain essentially flat, as neither uses arcosh to compute distances. RAP has the lowest distortion in every bin; full results are in Appendix F.1.

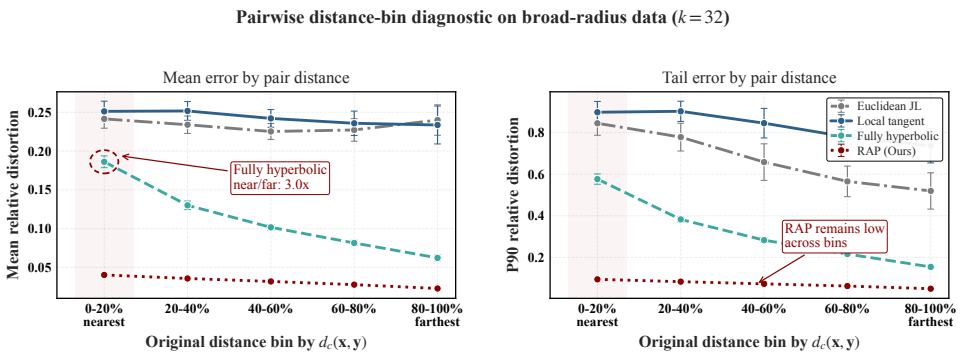


Figure 3: Pairwise-distance-bin diagnostic on broad-radius data at $k = 32$. Pairs are binned by the original hyperbolic distance $d_c(x, y)$ before distortion is measured. The fully hyperbolic comparator has pronounced nearest-pair sensitivity, whereas RAP keeps mean and tail distortion low and stable across all distance bins.

The RAP gap over the strongest non-RAP comparator is structural rather than scale-dependent. Figure 4 plots mean distortion on the origin-centered benchmark. RAP is best at every k, improving over the fully hyperbolic baseline from roughly $3\times$ at $k = 16$ to more than $6\times$ at $k = 64$. This persistent gap is diagnostic: increasing k reduces angular JL noise, but the tangent and fully hyperbolic baselines still perturb hierarchy-carrying radii. RAP removes this radial error floor, leaving angular distortion as the controlled residual error, exactly the mechanism isolated by the theory above.

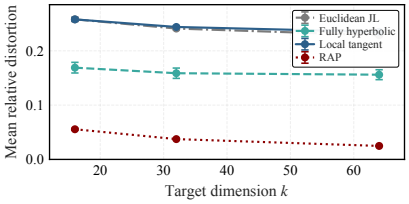


Figure 4: Origin-centered synthetic benchmark; 95% CIs over 20 seeds.

The gain is not only a distance-distortion artifact. Appendix G.1 further verifies complementary signals beyond mean distortion: neighborhood order, rank-based dimensionality-reduction quality, hierarchy/radius structure, and distance-kernel spectra. Tables 11 and 13 show that RAP gives the best mean rank correlation, kNN/trustworthiness-style stability, and spectral-profile errors on the tested hyperbolic representations; Table 12 is consistent with the exact-radius mechanism already discussed. These are empirical diagnostics rather than new JL guarantees, but they indicate that the same construction also preserves structural information used by hyperbolic representation learning.

Downstream performance on vision–language representations. On MERU and HyCoCLIP representations (Desai et al., 2023; Pal et al., 2025), RAP maintains competitive classification accuracy and improves retrieval scores by 5.5%–20.1% relative to the stronger of the two vanilla hyperbolic projections. Experimental protocols and detailed results appear in Appendix F.7 and Table 6.

Worst-pair distortion and scalability. RAP's advantage over the vanilla hyperbolic projections extends to exact worst-pair distortion and million-point compression. Appendix F.8 provides the all-pair and sampled-pair protocols and detailed results.

6 Conclusion

RAP fills the gap in data-oblivious hyperbolic-to-hyperbolic compression, with a simultaneous finite-set guarantee and explicit geometric conditioning. Preserving hierarchy-carrying radii avoids tangent-space curvature bias and radial-angular mixing using one Gaussian projection and normalization, without data-dependent fitting. Experiments demonstrate improved mean and exact worst-pair preservation, million-point scalability, and better vision–language retrieval than hyperbolic projection baselines. Its simplicity and geometric fidelity make RAP a promising building block for future hyperbolic foundation models. In the future, we plan to extend RAP to multi-curvature and product-manifold representations.

AI USE STATEMENT

Generative AI tools were used only for grammar, spelling, and typographical checks, as well as reference verification. The authors reviewed all suggested edits and take full responsibility for the final content, including all text, claims, proofs, experiments, and artifacts.

ETHICS STATEMENT

The method introduces no new data collection, human-subject interaction, or generative capability; its societal impacts depend on the downstream systems that use compressed hyperbolic representations. Biases and sensitive information in source representations may persist after compression, and our geometric and downstream evaluations do not establish fairness, privacy, or safety guarantees. Since compression may affect tasks and data subgroups differently, deployment benefits from application-specific evaluation and safeguards, particularly when outputs inform consequential decisions.

REPRODUCIBILITY STATEMENT

Section 4.4 specifies the RAP construction, and Appendix J provides its assumptions and proofs. Section 5 and Appendix E document the datasets, projection settings, random-seed protocol, evaluation metrics, and compute resources. Appendix F details the supplementary experiments, including the vision–language evaluation in Appendix F.7 and the exact worst-pair and scaling tests in Appendix F.8.

REFERENCES

Dimitris Achlioptas. Database-friendly random projections: Johnson–Lindenstrauss with binary coins. *Journal of Computer and System Sciences*, 66(4):671–687, 2003.

Nir Ailon and Bernard Chazelle. The fast Johnson–Lindenstrauss transform and approximate nearest neighbors. *SIAM Journal on Computing*, 39(1):302–322, 2009.

Haya Alyoussef, Ahmad Bdeir, Diego Coello de Portugal Mecke, Tom Hanika, Niels Landwehr, and Lars Schmidt-Thieme. HexFormer: Hyperbolic vision transformer with exponential map aggregation. *arXiv preprint arXiv:2601.19849*, 2026.

Richard G. Baraniuk and Michael B. Wakin. Random projections of smooth manifolds. *Foundations of Computational Mathematics*, 9(1):51–77, 2009. doi: 10.1007/s10208-007-9011-z.

Ahmad Bdeir, Kristian Schwethelm, and Niels Landwehr. Fully hyperbolic convolutional neural networks for computer vision. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=ekz1hN5QNh.

Ella Bingham and Heikki Mannila. Random projection in dimensionality reduction: Applications to image and text data. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 245–250, 2001.

Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. Food-101 – mining discriminative components with random forests. In *European Conference on Computer Vision*, 2014.

Jean Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel Journal of Mathematics*, 52(1–2):46–52, 1985.

Christos Boutsidis, Anastasios Zouzias, and Petros Drineas. Random projections for k-means clustering. In *Advances in Neural Information Processing Systems*, volume 23, 2010.

James W. Cannon, William J. Floyd, Richard Kenyon, and Walter R. Parry. Hyperbolic geometry. In Silvio Levy (ed.), *Flavors of Geometry*, volume 31 of *MSRI Publications*, pp. 59–115. Cambridge University Press, 1997.

Ines Chami, Zhitao Ying, Christopher Ré, and Jure Leskovec. Hyperbolic graph convolutional neural networks. In *Advances in Neural Information Processing Systems*, volume 32, pp. 4869–4880, 2019.

Ines Chami, Albert Gu, Dat P. Nguyen, and Christopher Ré. HoroPCA: Hyperbolic dimensionality reduction via horospherical projections. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 1419–1429. PMLR, 2021.

Chi-Chih Chang, Wei-Cheng Lin, Chien-Yu Lin, Chong-Yan Chen, Yu-Fang Hu, Pei-Shuo Wang, Ning-Chi Huang, Luis Ceze, Mohamed Abdelfattah, and Kai-Chiang Wu. Palu: KV-Cache compression with low-rank projection. In *International Conference on Learning Representations*, 2025.

Weize Chen, Xu Han, Yankai Lin, Hexu Zhao, Zhiyuan Liu, Peng Li, Maosong Sun, and Jie Zhou. Fully hyperbolic neural networks. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 5672–5686, Dublin, Ireland, 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.acl-long.389. URL https://aclanthology.org/2022.acl-long.389/.

Ziheng Chen, Bernhard Schölkopf, and Nicu Sebe. Hyperbolic busemann neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 42029–42038, 2026. URL https://openaccess.thecvf.com/content/CVPR2026/papers/Chen_Hyperbolic_Busemann_Neural_Networks_CVPR_2026_paper.pdf.

Krzysztof Choromanski, Valerii Likhosherstov, David Dohan, Xingyou Song, Georgiana Gane, Tamas Sarlos, Peter Hawkins, Jared Davis, Afroz Mohiuddin, Lukasz Kaiser, David Belanger, Lucy Colwell, and Adrian Weller. Rethinking attention with performers. In *International Conference on Learning Representations*, 2021.

Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, Sammy Mohamed, and Andrea Vedaldi. Describing textures in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014.

Sanjoy Dasgupta and Anupam Gupta. An elementary proof of the Johnson–Lindenstrauss lemma. *Random Structures & Algorithms*, 22(1):60–65, 2003.

Karan Desai, Maximilian Nickel, Tanmay Rajpurohit, Justin Johnson, and Shanmukha Ramakrishna Vedantam. Hyperbolic image-text representations. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 7694–7731. PMLR, 2023. URL https://proceedings.mlr.press/v202/desai23a.html.

Xiran Fan, Chun-Hao Yang, and Baba C. Vemuri. Nested hyperbolic spaces for dimensionality reduction and hyperbolic NN design. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 356–365, 2022.

Li Fei-Fei, Rob Fergus, and Pietro Perona. Learning generative visual models from few training examples. In *Workshop on Generative-Model Based Vision, IEEE Conference on Computer Vision and Pattern Recognition*, 2004.

P. Thomas Fletcher, Conglin Lu, Stephen M. Pizer, and Sarang Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging*, 23(8):995–1005, 2004.

Octavian-Eugen Ganea, Gary Bécigneul, and Thomas Hofmann. Hyperbolic neural networks. In *Advances in Neural Information Processing Systems*, 2018.

Nathan Halko, Per-Gunnar Martinsson, and Joel A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53 (2):217–288, 2011.

Neil He, Rishabh Anand, Hiren Madhu, Ali Maatouk, Smita Krishnaswamy, Leandros Tassiulas, Menglin Yang, and Rex Ying. HELM: Hyperbolic large language models via mixture-of-curvature experts. In *Advances in Neural Information Processing Systems*, volume 38, pp. 142604–142635. Curran Associates, Inc., 2025a. doi: 10.52202/085713-4774. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/d1e2f808a51842eedaf6ef0099d716c6-Abstract-Conference.html.

Neil He, Menglin Yang, and Rex Ying. HyperCore: The core framework for building hyperbolic foundation models with comprehensive modules. *arXiv preprint arXiv:2504.08912*, 2025b.

Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2019.

Justin Johnson, Bharath Hariharan, Laurens van der Maaten, Li Fei-Fei, C. Lawrence Zitnick, and Ross Girshick. CLEVR: A diagnostic dataset for compositional language and elementary visual reasoning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

William B. Johnson and Joram Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Contemporary Mathematics*, volume 26, pp. 189–206. American Mathematical Society, 1984.

Valentin Khrulkov, Leyla Mirvakhabova, Evgeniya Ustinova, Ivan Oseledets, and Victor Lempitsky. Hyperbolic image embeddings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.

Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009. URL https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf.

Kasper Green Larsen and Jelani Nelson. Optimality of the Johnson–Lindenstrauss lemma. In *Proceedings of the 58th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 633–638. IEEE, 2017.

Marc Law, Renjie Liao, Jake Snell, and Richard Zemel. Lorentzian distance learning for hyperbolic representations. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 3672–3681. PMLR, 2019.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

John A. Lee and Michel Verleysen. Quality assessment of dimensionality reduction: Rank-based criteria. *Neurocomputing*, 72(7–9):1431–1443, 2009. doi: 10.1016/j.neucom.2008.12.017.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common objects in context. In *European Conference on Computer Vision*, 2014.

Aaron Lou, Isay Katsman, Qingxuan Jiang, Serge Belongie, Ser-Nam Lim, and Christopher De Sa. Differentiating through the Fréchet mean. In *International Conference on Machine Learning*, 2020.

Hiren Madhu, Ngoc Bui, Ali Maatouk, Leandros Tassiulas, Smita Krishnaswamy, Menglin Yang, Sukanta Ganguly, Kiran Srinivasan, and Rex Ying. HypRAG: Hyperbolic dense retrieval for retrieval augmented generation. *arXiv preprint arXiv:2602.07739*, 2026.

Subhransu Maji, Esa Rahtu, Juho Kannala, Matthew Blaschko, and Andrea Vedaldi. Fine-grained visual classification of aircraft, 2013. URL https://arxiv.org/abs/1306.5151.

Jiří Matoušek. On variants of the Johnson–Lindenstrauss lemma. *Random Structures & Algorithms*, 33(2):142–156, 2008.

George A. Miller. WordNet: A lexical database for English. *Communications of the ACM*, 38(11): 39–41, 1995. doi: 10.1145/219717.219748.

Yoshihiro Nagano, Shoichiro Yamaguchi, Yasuhiro Fujita, and Masanori Koyama. A wrapped normal distribution on hyperbolic space for gradient-based learning. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 4693–4702. PMLR, 2019.

Maximilian Nickel and Douwe Kiela. Poincaré embeddings for learning hierarchical representations. In *Advances in Neural Information Processing Systems*, 2017.

Maximilian Nickel and Douwe Kiela. Learning continuous hierarchies in the Lorentz model of hyperbolic geometry. In *Proceedings of the 35th International Conference on Machine Learning*, 2018.

Maria-Elena Nilsback and Andrew Zisserman. Automated flower classification over a large number of classes. In *Indian Conference on Computer Vision, Graphics and Image Processing*, 2008.

Avik Pal, Max van Spengler, Guido Maria D'Amely di Melendugno, Alessandro Flaborea, Fabio Galasso, and Pascal Mettes. Compositional entailment learning for hyperbolic vision-language models. In *International Conference on Learning Representations*, 2025.

Saurabh Paul, Christos Boutsidis, Malik Magdon-Ismail, and Petros Drineas. Random projections for support vector machines. In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, volume 31 of *Proceedings of Machine Learning Research*, pp. 498–506. PMLR, 2013.

Zelin Peng, Zhengqin Xu, Qingyang Liu, Xiaokang Yang, and Wei Shen. HyperET: Efficient training in hyperbolic space for multi-modal large language models. In *Advances in Neural Information Processing Systems*, 2025.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8748–8763. PMLR, 2021. URL https://proceedings.mlr.press/v139/radford21a.html.

John G. Ratcliffe. *Foundations of Hyperbolic Manifolds*, volume 149 of *Graduate Texts in Mathematics*. Springer, 3 edition, 2019.

Olivier Roy and Martin Vetterli. The effective rank: A measure of effective dimensionality. In *Proceedings of the 15th European Signal Processing Conference*, pp. 606–610, 2007.

Rajarshi Saha, Naomi Sagan, Varun Srivastava, Andrea J. Goldsmith, and Mert Pilanci. Compressing large language models using low rank and low precision decomposition. In *Advances in Neural Information Processing Systems*, 2024.

Frederic Sala, Christopher De Sa, Albert Gu, and Christopher Ré. Representation tradeoffs for hyperbolic embeddings. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 4460–4469. PMLR, 2018.

Ryohei Shimizu, Yusuke Mukuta, and Tatsuya Harada. Hyperbolic neural networks++. In *International Conference on Learning Representations*, 2021.

Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 1631–1642. Association for Computational Linguistics, 2013. URL https://aclanthology.org/D13-1170/.

Alexandru Tifrea, Gary Bécigneul, and Octavian-Eugen Ganea. Poincaré GloVe: Hyperbolic word embeddings. In *International Conference on Learning Representations*, 2019.

Robert van der Klis, Ricardo Chávez Torres, Max van Spengler, Yuhui Ding, Thomas Hofmann, and Pascal Mettes. Fast and geometrically grounded lorentz neural networks. *arXiv preprint arXiv:2601.21529*, 2026.

Bastiaan S. Veeling, Jasper Linmans, Jim Winkens, Taco Cohen, and Max Welling. Rotation equivariant CNNs for digital pathology. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, volume 11071 of *Lecture Notes in Computer Science*, pp. 210–218. Springer International Publishing, 2018. doi: 10.1007/978-3-030-00934-2_24. URL https://doi.org/10.1007/978-3-030-00934-2_24.

Kilian Q. Weinberger, Anirban Dasgupta, John Langford, Alexander J. Smola, and Josh Attenberg. Feature hashing for large scale multitask learning. In *International Conference on Machine Learning*, 2009.

Menglin Yang, Harshit Verma, Delvin Ce Zhang, Jiahong Liu, Irwin King, and Rex Ying. Hypformer: Exploring efficient transformer fully in hyperbolic space. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 3770–3781. ACM, 2024. doi: 10.1145/3637528.3672039.

Menglin Yang, Ram Samarth B B, Aosong Feng, Bo Xiong, Jiahong Liu, Irwin King, and Rex Ying. Hyperbolic fine-tuning for large language models. In *Advances in Neural Information Processing Systems*, 2025.

Peter Young, Alice Lai, Micah Hodosh, and Julia Hockenmaier. From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions. *Transactions of the Association for Computational Linguistics*, 2:67–78, 2014. doi: 10.1162/tacl_a_00166. URL https://aclanthology.org/Q14-1006/.

Appendix

A Hyperbolic Geometry Background **17**
A.1 Models of Hyperbolic Geometry 17
A.2 Tangent Spaces, Exponential and Logarithmic Maps 18
A.3 Polar Coordinates and the Law of Cosines 18

B Notation **19**

C Euclidean JL Baseline **19**

D Comparison with Learned Hyperbolic Dimensionality Reduction **20**

E Experimental Protocol Details **20**

F Supplementary Robustness Experiments **21**
F.1 Pairwise-Distance Conditioning of the Fully Hyperbolic Comparator 21
F.2 Per-Node Radial Preservation on WordNet 21
F.3 Local-Radius Behavior of the Tangent-Space Comparator 22
F.4 Tangent-Space Curvature-Scaling Calibration 23
F.5 Anchor Sensitivity of the Tangent-Space Comparator 23
F.6 Sample-Size Behavior of Mean Distortion 24
F.7 Modern Vision–Language Evaluation Details 25
F.8 Exact Worst-Pair and Scaling Audits 25

G Property Preservation Beyond Distance Distortion **27**
G.1 Quantitative Analysis of Preserved Properties 27
G.2 Concrete WordNet Hierarchy Case Study 27

H Proofs for the Tangent-Space Comparator **28**
H.1 Source-Side Metric Comparison 28
H.2 JL with the Origin Included 30
H.3 Target-Side Comparison 30
H.4 Error Composition 31

I Fully Hyperbolic Comparator Conditioning **31**

J RAP Radial Preservation and Angular Sensitivity **33**
J.1 Well-Definedness and the Hyperboloid Constraint 33
J.2 Normalized Angular JL 34
J.3 From Angular Error to Hyperbolic Distance Error 34
J.4 A Finite-Set RAP Guarantee 35

K Theoretical Comparison of the Three Schemes **36**

A Hyperbolic Geometry Background

This appendix collects the model definitions, distance formulas, and tangent-space maps used implicitly throughout the paper. The presentation is self-contained and follows standard treatments of hyperbolic geometry (Cannon et al., 1997; Ratcliffe, 2019) together with the machine-learning conventions popularized by hyperbolic neural networks (Ganea et al., 2018; Nickel & Kiela, 2018). We write $\mathbb{H}^d_c$ for d-dimensional hyperbolic space of constant sectional curvature $-c$ with $c > 0$, and d_c for its geodesic distance.

A.1 Models of Hyperbolic Geometry

Hyperbolic space admits several equivalent realizations as Riemannian submanifolds of Euclidean spaces; any two such realizations are connected by an isometry, so geometric quantities such as d_c are model-independent. Different models expose different aspects of the geometry, and the analysis in this paper draws on two of them.

The *Poincaré ball model* $\mathbb{D}^d_c$ is the open ball $\{\mathbf{x} \in \mathbb{R}^d : c\|\mathbf{x}\|_2^2 < 1\}$ equipped with the conformal metric

$$g^{\mathbb{D}}_{\mathbf{x}} = (\lambda^c_{\mathbf{x}})^2 \, g^E, \qquad \lambda^c_{\mathbf{x}} = \frac{2}{1 - c\|\mathbf{x}\|_2^2}, \tag{13}$$

where g^E is the Euclidean inner product. The conformal factor $\lambda^c_{\mathbf{x}}$ blows up at the boundary, so the boundary sphere $\{c\|\mathbf{x}\|_2^2 = 1\}$ sits at infinite geodesic distance from any interior point. Geodesic distance has the closed form

$$d_c(\mathbf{x}, \mathbf{y}) = \frac{1}{\sqrt{c}} \operatorname{arcosh}\!\left(1 + \frac{2c\|\mathbf{x} - \mathbf{y}\|_2^2}{(1 - c\|\mathbf{x}\|_2^2)(1 - c\|\mathbf{y}\|_2^2)}\right). \tag{14}$$

The model is conformal: angles between tangent vectors agree with their Euclidean angles, which makes it convenient for visualizing low-dimensional embeddings. The price is that straight lines through the origin are the only Euclidean-straight geodesics; all others are circular arcs orthogonal to the boundary.

The *Lorentz* or *hyperboloid model* $\mathbb{L}^d_c$ realizes hyperbolic space as the upper sheet of a two-sheeted hyperboloid in $\mathbb{R}^{d+1}$. Writing a point as $x = (x_0, \bar{\mathbf{x}})$ with $x_0 \in \mathbb{R}$ and $\bar{\mathbf{x}} \in \mathbb{R}^d$, and equipping $\mathbb{R}^{d+1}$ with the Lorentzian inner product

$$\langle x, y \rangle_L = -x_0 y_0 + \langle \bar{\mathbf{x}}, \bar{\mathbf{y}} \rangle, \tag{15}$$

the model is

$$\mathbb{L}^d_c = \{x \in \mathbb{R}^{d+1} : \langle x, x \rangle_L = -1/c, \; x_0 > 0\}. \tag{16}$$

The Riemannian metric is the restriction of $\langle \cdot, \cdot \rangle_L$ to the tangent bundle, which is positive definite on each tangent space. The geodesic distance reads

$$d_c(x, y) = \frac{1}{\sqrt{c}} \operatorname{arcosh}\!\big(-c \, \langle x, y \rangle_L\big). \tag{17}$$

Geodesics are intersections of $\mathbb{L}^d_c$ with two-dimensional linear subspaces of $\mathbb{R}^{d+1}$, so the model is convenient for closed-form manipulations and for the spatial-coordinate sketch in Equation (9).

The Poincaré ball and Lorentz models are related by a stereographic projection from the point $(-1/\sqrt{c}, \mathbf{0})$:

$$\pi : \mathbb{L}^d_c \to \mathbb{D}^d_c, \qquad \pi(x_0, \bar{\mathbf{x}}) = \frac{\sqrt{c}\,\bar{\mathbf{x}}}{1 + \sqrt{c}\, x_0}. \tag{18}$$

The map π is a Riemannian isometry, and identities such as the law of cosines (Equation (1)) are stable under Equation (18). We therefore state intrinsic results in terms of the model-free $\mathbb{H}^d_c$ and switch to the Lorentz model only when an explicit spatial-coordinate computation is needed (Section 4.3 and Appendix I).

A.2 Tangent Spaces, Exponential and Logarithmic Maps

For any base point $a \in \mathbb{H}^d_c$, the tangent space $T_a\mathbb{H}^d_c$ is a d-dimensional Euclidean inner-product space with the metric tensor g_a. The geometry around a is connected to its tangent space by the Riemannian exponential map

$$\exp_a : T_a\mathbb{H}^d_c \to \mathbb{H}^d_c, \qquad \exp_a(\mathbf{v}) = \gamma_{\mathbf{v}}(1), \tag{19}$$

where $\gamma_{\mathbf{v}}$ is the unique unit-speed geodesic with $\gamma_{\mathbf{v}}(0) = a$ and $\dot\gamma_{\mathbf{v}}(0) = \mathbf{v}/\|\mathbf{v}\|_{g_a}$ emanating from a. Because $\mathbb{H}^d_c$ is a Cartan–Hadamard manifold, $\exp_a$ is a diffeomorphism onto $\mathbb{H}^d_c$ and admits a global inverse, the logarithmic map

$$\log_a : \mathbb{H}^d_c \to T_a\mathbb{H}^d_c, \qquad \|\log_a(x)\|_{g_a} = d_c(a, x). \tag{20}$$

The pair $(\exp_a, \log_a)$ furnishes the geodesic normal chart used throughout the paper: it sends a neighborhood of a to a Euclidean neighborhood of the origin in $T_a\mathbb{H}^d_c$, with norms in the chart matching geodesic distances to a exactly along radial directions and agreeing with Euclidean distances up to an $O(c\|\mathbf{u}\|^2)$ correction for nearby pairs (Equation (2)).

In the Lorentz model the maps admit closed forms. Tangent vectors at $x \in \mathbb{L}^d_c$ are characterized by $\langle x, \mathbf{v}\rangle_L = 0$, their norm is $\|\mathbf{v}\|_L = \sqrt{\langle \mathbf{v}, \mathbf{v}\rangle_L}$, and

$$\exp^L_x(\mathbf{v}) = \cosh\big(\sqrt{c}\,\|\mathbf{v}\|_L\big)\, x + \frac{\sinh\big(\sqrt{c}\,\|\mathbf{v}\|_L\big)}{\sqrt{c}\,\|\mathbf{v}\|_L}\,\mathbf{v}, \tag{21}$$

$$\log^L_x(y) = \frac{\operatorname{arcosh}\big(-c\,\langle x, y\rangle_L\big)}{\sinh\big(\operatorname{arcosh}(-c\,\langle x, y\rangle_L)\big)}\,\big(y + c\,\langle x, y\rangle_L\, x\big). \tag{22}$$

Equation (21) reduces to a hyperbolic rotation of the pair $(x, \mathbf{v}/\|\mathbf{v}\|_L)$ inside the two-dimensional plane that they span; Equation (22) expresses $\log^L_x(y)$ as the component of y orthogonal to x in the Lorentz inner product, rescaled to have $\|\log^L_x(y)\|_L = d_c(x, y)$.

In the Poincaré ball, the corresponding formulas are most compactly written through Möbius addition, the curved analogue of vector addition in the ball (Ganea et al., 2018):

$$\mathbf{x} \oplus_c \mathbf{y} = \frac{(1 + 2c\langle \mathbf{x}, \mathbf{y}\rangle + c\|\mathbf{y}\|_2^2)\mathbf{x} + (1 - c\|\mathbf{x}\|_2^2)\mathbf{y}}{1 + 2c\langle \mathbf{x}, \mathbf{y}\rangle + c^2\|\mathbf{x}\|_2^2\|\mathbf{y}\|_2^2}. \tag{23}$$

The exponential and logarithmic maps at $\mathbf{x} \in \mathbb{D}^d_c$ then take the form

$$\exp^{\mathbb{D}}_{\mathbf{x}}(\mathbf{v}) = \mathbf{x} \oplus_c \tanh\left(\frac{\sqrt{c}\,\lambda^c_{\mathbf{x}}\|\mathbf{v}\|_2}{2}\right)\frac{\mathbf{v}}{\sqrt{c}\,\|\mathbf{v}\|_2}, \tag{24}$$

$$\log^{\mathbb{D}}_{\mathbf{x}}(\mathbf{y}) = \frac{2}{\sqrt{c}\,\lambda^c_{\mathbf{x}}}\operatorname{arctanh}\big(\sqrt{c}\,\|-\mathbf{x} \oplus_c \mathbf{y}\|_2\big)\frac{-\mathbf{x} \oplus_c \mathbf{y}}{\|-\mathbf{x} \oplus_c \mathbf{y}\|_2}, \tag{25}$$

specializing at the origin to $\exp^{\mathbb{D}}_{\mathbf{0}}(\mathbf{v}) = \tanh(\sqrt{c}\|\mathbf{v}\|_2)\mathbf{v}/(\sqrt{c}\|\mathbf{v}\|_2)$ and $\log^{\mathbb{D}}_{\mathbf{0}}(\mathbf{y}) = \operatorname{arctanh}(\sqrt{c}\|\mathbf{y}\|_2)\mathbf{y}/(\sqrt{c}\|\mathbf{y}\|_2)$. The two model-specific exponentials are intertwined by $\pi \circ \exp^L_x = \exp^{\mathbb{D}}_{\pi(x)} \circ d\pi_x$ on the appropriate tangent vectors, which is why the analyses in the paper can be conducted in either model without changing the underlying distortion quantities.

A.3 Polar Coordinates and the Law of Cosines

Fixing an origin $o \in \mathbb{H}^d_c$ and identifying $T_o\mathbb{H}^d_c$ with $\mathbb{R}^d$ through a Riemannian isometry, every point $x \neq o$ has a unique decomposition

$$\log_o(x) = r_x\,\mathbf{u}_x, \qquad r_x = d_c(o, x) \in (0, \infty), \qquad \mathbf{u}_x \in \mathbb{S}^{d-1}, \tag{26}$$

which we abbreviate as $x = (r_x, \mathbf{u}_x)$ in polar form. In Lorentz coordinates this corresponds to $x = (\cosh(\sqrt{c}\,r_x)/\sqrt{c}, \sinh(\sqrt{c}\,r_x)\mathbf{u}_x/\sqrt{c})$ with $o = (1/\sqrt{c}, \mathbf{0})$, as used in the polar parameterization ι_k of Equation (11). The hyperbolic law of cosines (Equation (1)) follows directly by computing $\langle x, y\rangle_L$ in these coordinates and applying the distance formula (Equation (17)). Decoupling r_x from $\mathbf{u}_x$ in this way is what makes the proposed radial-angular projection in Section 4.4 a natural construction: the radius is preserved verbatim, and the entire compression budget is spent on $\mathbf{u}_x$.

B NOTATION

Table 3 collects the symbols used in the main text and the proofs that follow. Throughout, italic letters denote points on the hyperbolic manifold, while bold letters denote Euclidean, tangent, angular, or spatial coordinate vectors and matrices.

Table 3: Core notation used throughout the paper.

Symbol	Meaning
$\mathbb{H}_c^d$	d-dimensional hyperbolic space with constant curvature $-c$.
c	Positive curvature magnitude, so the manifold curvature is $-c$.
$d_c(x, y)$	Hyperbolic geodesic distance between manifold points x and y.
$X = \{x_1, \dots, x_n\}$	Finite input point set used in the finite-set preservation problem.
d, k, n	Ambient dimension, target dimension, and number of evaluated points.
ε, δ	Target distortion level and failure probability in the JL-style objective.
$x, y, a, \tilde{a}, o, \tilde{o}$	Points on the hyperbolic manifold.
o	Origin or anchor used to define polar coordinates.
$r_x, \mathbf{u}_x$	Hyperbolic radius $d_c(o, x)$ and angular direction in $\mathbb{S}^{d-1}$.
$\widetilde{\mathbf{u}}_x$	Compressed and renormalized angular direction produced by RAP.
$\bar{\mathbf{x}}$	Spatial Lorentz coordinate vector of a hyperboloid point $x = (x_0, \bar{\mathbf{x}})$.
$U_a, U_{\tilde{a}}$	Orthonormal coordinate identifications from tangent spaces to Euclidean coordinate spaces.
$\mathbf{P}$	Euclidean JL matrix used to sketch tangent or angular coordinates.
$\mathbf{A}$	Spatial-coordinate matrix used by the fully hyperbolic comparator.
Φ	Generic randomized compression map in the problem definition.
$\Delta_X(\Phi)$	Worst-pair relative distortion of a realized map Φ on the finite set X.
Π_{RAP}	Proposed radial-angular projection $(r_x, \mathbf{u}_x) \mapsto (r_x, \mathbf{P}\mathbf{u}_x/\lVert\mathbf{P}\mathbf{u}_x\rVert)$.
$\Pi_{a,\mathbf{P}}$	Tangent-space comparator $\exp_{\tilde{a}}(U_{\tilde{a}}^{-1}\mathbf{P}U_a \log_a(\cdot))$.
$\Pi_{\mathbf{A}}$	Fully hyperbolic comparator that sketches spatial coordinates and renormalizes to the hyperboloid.
$\mathbf{z}_x$	Origin-log Euclidean coordinate vector $U_o \log_o(x)$ used by the Euclidean JL baseline.
Φ_{Euc}	Euclidean JL baseline map $\Phi_{\mathrm{Euc}}(x) = \mathbf{P}\mathbf{z}_x$, whose output lives in $\mathbb{R}^k$.
$d_{\mathrm{Euc},\mathbf{P}}$	Euclidean output distance used by the Euclidean JL baseline.
η	Angular inner-product perturbation used in the RAP sensitivity statement.

C EUCLIDEAN JL BASELINE

Fix the origin or anchor $o \in \mathbb{H}_c^d$ and let $\mathbf{P} \in \mathbb{R}^{k\times d}$ be a Gaussian JL matrix. For each $x \in \mathbb{H}_c^d$ write $\mathbf{z}_x = U_o \log_o(x) \in \mathbb{R}^d$. The Euclidean JL diagnostic is

$$\Phi_{\mathrm{Euc}}(x) = \mathbf{P}\mathbf{z}_x \in \mathbb{R}^k, \qquad d_{\mathrm{Euc},\mathbf{P}}(x, y) = \lVert\Phi_{\mathrm{Euc}}(x) - \Phi_{\mathrm{Euc}}(y)\rVert_2, \tag{27}$$

and its reported error is $|d_{\mathrm{Euc},\mathbf{P}}(x, y)/d_c(x, y) - 1|$. The output lives in $\mathbb{R}^k$, not in $\mathbb{H}_c^k$, so the diagnostic is intentionally cross-geometry: it asks whether a single logarithmic chart followed by an ordinary Euclidean sketch already approximates hyperbolic distance. The tangent-space comparator $\Pi_{o,\mathbf{P}}$ from Section 4.2 differs in that it composes the same Euclidean JL matrix with $\exp_{\tilde{o}}$ and evaluates distortion in $\mathbb{H}_c^k$.

The two are numerically close inside a small ball: for $x, y \in B(o, R)$ with $R\sqrt{c}$ small, normal-coordinate comparison gives $d_c(x, y) = \lVert\mathbf{z}_x - \mathbf{z}_y\rVert(1 + O(cR^2))$, and the target-side exponential adds another $O(cR'^2)$ term. As the radius grows, the logarithmic chart ceases to be a faithful linearization, and the Euclidean diagnostic increasingly mismeasures the underlying hyperbolic distance, while the tangent-space comparator at least evaluates its output in the correct geometry.

D Comparison with Learned Hyperbolic Dimensionality Reduction

HoroPCA and nested hyperbolic spaces are the closest hyperbolic dimensionality-reduction methods to our setting, but they solve a different problem from an oblivious JL-style sketch. HoroPCA learns hyperbolic principal directions using ideal points, Busemann coordinates, and horospherical projections (Chami et al., 2021). Nested hyperbolic spaces learn or select a low-dimensional embedded hyperboloid and use the resulting projection as both a dimensionality-reduction primitive and a trainable component of hyperbolic neural networks (Fan et al., 2022). Both methods can be strong reconstruction or visualization baselines when fitted to a particular dataset, but their fitted map changes with the dataset. RAP, in contrast, samples a single random map independently of the data and asks for finite-set distance preservation.

Table 4: Positioning relative to data-dependent hyperbolic dimensionality reduction. "JL-style" means an oblivious finite-set distance-preservation statement for a randomly sampled map.

Method	Fitted objective / map	Oblivious JL-style?	Role relative to RAP
Tangent PCA / PGA (Fletcher et al., 2004)	Variance or reconstruction in a tangent/geodesic model fitted at a base point	No: data-dependent principal directions	Useful classical reference, but inherits locality and base-point dependence.
HoroPCA (Chami et al., 2021)	Hyperbolic PCA via fitted ideal points, Busemann coordinates, and horospherical projections	No: fits directions to the dataset	Strong learned dimensionality-reduction baseline for reconstruction or visualization, but not an oblivious random projection.
Nested hyperbolic spaces (Fan et al., 2022)	Reconstruction through a learned lower-dimensional hyperboloid embedded in the original one	No: learns projection parameters	Closest learned hyperbolic-to-hyperbolic comparator, but evaluates a fitted submanifold objective.
RAP (ours)	Finite-set distance preservation by keeping radius and randomly sketching angular directions	Yes: no fitting, one random map	Targets the JL-style primitive for arbitrary sufficiently regular finite sets.

This distinction matters for empirical comparisons. If HoroPCA or nested hyperbolic spaces are fitted separately on each benchmark, they are allowed to use the entire point cloud to choose a low-dimensional representation; RAP and the other projection baselines in Section 5 are not. Such an experiment would answer a useful but different reconstruction question. On the other hand, forcing HoroPCA or nested hyperbolic spaces to use random unfitted directions would no longer evaluate the methods as proposed. We therefore treat them as related data-dependent dimensionality-reduction methods rather than main baselines for the JL-style problem.

E Experimental Protocol Details

Datasets. The *broad-radius benchmark* samples origin-centered hyperbolic points in $\mathbb{H}^{128}$ with radial scales $\{0.5, 1.0, 2.0\}$; for each seed we average over the three scales before computing the reported mean and standard deviation. The *balanced tree* embeds a 4-ary depth-5 tree in $\mathbb{H}^{50}$ and subsamples $n \in \{128, 256, 512\}$ nodes uniformly at random per seed. The *clustered Gaussian* dataset places eight cluster centers at different radii and angular directions in $\mathbb{H}^{50}$, with $n/8$ points per cluster sampled from a wrapped hyperbolic Gaussian around each center. The *WordNet* dataset uses the noun hierarchy rooted at `animal.n.01`, varying the retained hyponym depth and capping each subtree at 512 nodes via deterministic subsampling. WordNet is distributed under the Princeton WordNet 3.0 license; we use only synset identifiers and hyponym edges.

Comparison protocol. Each method uses an independent Gaussian projection matrix per seed; all four methods on a given seed share the same source point set so that the comparison is paired.

Curvature is fixed to $c = 1$ throughout the main experiments; the curvature sensitivity of the local tangent-space comparator is treated separately below in Appendix F.

Choice of mean over worst-pair distortion. The worst-pair distortion $\Delta_X(\Phi)$ defined in Section 4.1 is the natural target of the JL-style guarantee, but it is sensitive to a single nearly-coincident pair under finite sampling. We therefore report the mean relative distortion over nonzero-distance pairs in Tables 1 and 2. Appendix F.8 reports exact all-pair maxima for the three hyperbolic methods, while Appendix F.4 provides a separate curvature-scaling calibration for the local tangent comparator.

Compute resources. The core synthetic and WordNet evaluations are lightweight CPU workloads and involve no model fitting. The largest stress test projects one million points from $d = 4096$ to $k = 128$ on one A40 GPU; one projection-and-evaluation run takes about 11 seconds and peaks at 15.8 GiB. The vision–language study uses frozen representations, so its reported cost is projection and metric evaluation rather than training.

F Supplementary Robustness Experiments

Across all robustness checks reported in this appendix, the ranking of the four projection rules established in Tables 1 and 2 is preserved, and RAP's advantage extends past the regimes the theory directly characterizes. Specifically: (i) the fully hyperbolic comparator has a pronounced nearest-pair sensitivity consistent with the arcosh conditioning mechanism (Appendix F.1); (ii) RAP preserves the per-node radial coordinate exactly on the WordNet subtree, while the comparators introduce systematic drift (Appendix F.2); (iii) the tangent-space mean distortion decreases with k while its tail grows with local radius (Appendix F.3); (iv) a curvature-scaling calibration separates the nominal $5cR^2$ trend from small-radius near-pair sensitivity (Appendix F.4); (v) the tangent-space comparator is stable under anchor shifts up to a geodesic distance of order $1/\sqrt{c}$ (Appendix F.5); (vi) the mean relative distortion concentrates as n grows, consistent with the standard JL averaging behavior (Appendix F.6). Empirical robustness therefore extends past the strict envelope of Theorem 1 and the explicit conditioning factors in the finite-set RAP guarantee.

F.1 Pairwise-Distance Conditioning of the Fully Hyperbolic Comparator

This appendix records the experimental setup and full numerical detail behind Figure 3. We run the broad-radius synthetic benchmark with $n = 256$, ambient dimension 128, target dimension $k = 32$, curvature $c = 1$, and radial scale 2.0. For each of 20 seeds, we sort all nonzero pairs by their original distance $d_c(x, y)$ and divide them into five equal-frequency bins, then compute mean and P90 relative distortion within each bin. P90 is the 90th percentile of $|d_c(\Pi x_i, \Pi x_j)/d_c(x_i, x_j) - 1|$ within a bin, computed separately for each seed before aggregation. The pattern in Figure 3 matches Lemma 9: nearby pairs have $-c\langle x, y\rangle_L = \cosh(\sqrt{c}\, d_c(x, y))$ close to 1 and therefore lie in the high-sensitivity region of arcosh, so small Lorentz-product perturbations get amplified into large distance errors specifically on those pairs. Table 5 reports the nearest-bin and farthest-bin mean distortion for every scheme.

Table 5: Nearest-pair versus farthest-pair mean distortion. Standard deviations are computed over 20 seeds.

Scheme	Nearest pairs	Farthest pairs
Euclidean JL	0.241 ± 0.026	0.240 ± 0.042
Local tangent	0.251 ± 0.028	0.234 ± 0.052
Fully hyperbolic	0.186 ± 0.016	0.062 ± 0.008
RAP (Ours)	$\mathbf{0.040} \pm 0.002$	$\mathbf{0.023} \pm 0.001$

F.2 Per-Node Radial Preservation on WordNet

This appendix provides visualization details and per-node measurements for the WordNet comparison in Section 5.2. We use the same WordNet subtree at $k = 32$.

In the two-dimensional Poincare view in Figure 2, each node's distance from the origin is set to its full hyperbolic radius, and its angular position is given by the leading uncentered principal-plane direction of the projected embedding. In the two comparator panels, nodes from distinct hyponym depths overlap in radius, blurring the depth coloring across rings. The RAP panel reproduces the original concentric layering and isolates depth to a single shell. The black circle marks the display zoom window rather than the unit Poincare boundary. The Euclidean diagnostic is omitted because its output does not lie in $\mathbb{H}^k_c$.

Figure 5 plots each node's original hyperbolic radius against its projected radius, with points colored by WordNet depth. RAP lies on the diagonal by construction, whereas the tangent-space and fully hyperbolic comparators introduce systematic radial drift across depths. Their mean relative radius errors are 0.114 and 0.106, respectively, compared with 0.000 for RAP.

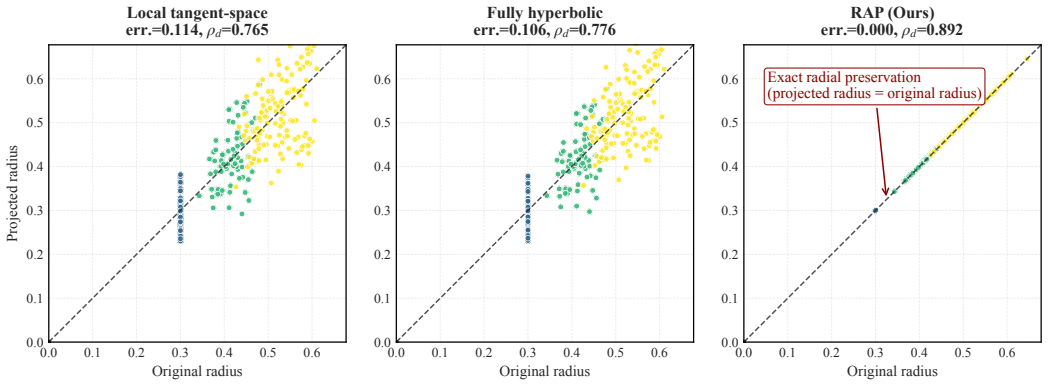


Figure 5: Radius preservation on the WordNet subtree at $k = 32$. Points are colored by WordNet depth. The diagonal indicates exact radial preservation.

F.3 Local-Radius Behavior of the Tangent-Space Comparator

This diagnostic isolates the tangent-space tail behavior. We sample synthetic balls of increasing local radius R and evaluate the local tangent-space projection across target dimensions. Figure 6 shows that mean distortion decreases with the target dimension k, while tail distortion grows as R increases. Both effects are consistent with Theorem 1: the JL term shrinks with k, but the curvature term cR^2 is independent of k and tracks the local radius. The figure is therefore a calibration of the local-JL guarantee rather than a cross-method comparison.

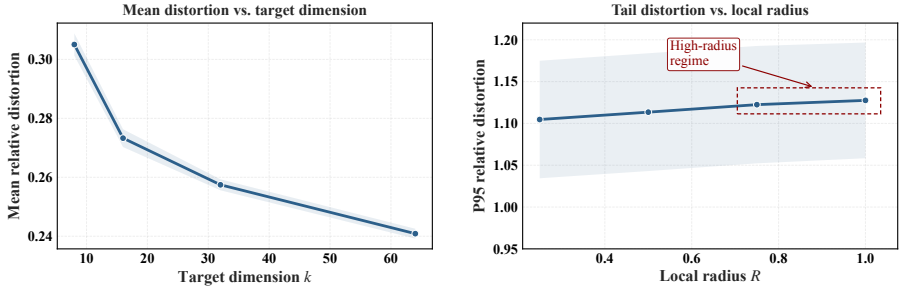


Figure 6: Tangent-space tail distortion as a function of the local radius R. Shaded bands show 95% confidence intervals (t-distribution) over 20 seeds. Mean distortion decreases with target dimension; tail distortion rises with local radius, matching the $\varepsilon + O(cR^2)$ envelope of Theorem 1.

F.4 TANGENT-SPACE CURVATURE-SCALING CALIBRATION

We evaluate the tangent-space scheme on synthetic balls of varying local radius R and compare the observed worst-pair distortion with the nominal curve $\varepsilon + 5cR^2$ derived in Appendix H. This plot does not certify the theorem seed by seed: the theorem also conditions on the realized tangent-space JL event and on source and target locality. We therefore use Figure 7 only to calibrate the cR^2 trend. The observed maximum grows more slowly than the nominal curve at moderate radii; at small radii it is dominated by near-coincident pairs and separation regularity. The constant $C_1 = 5C_0$ is deliberately conservative.

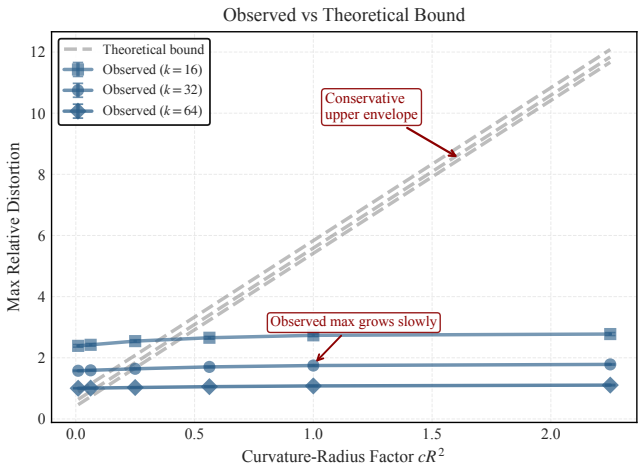


Figure 7: Observed maximum distortion and the nominal curvature-scaling curve $\varepsilon + 5cR^2$ (dashed gray) for $k \in \{16, 32, 64\}$ over 20 seeds. The curve is a calibration reference, not a certified per-run bound; the small-radius maximum is sensitive to near-pair effects.

F.5 ANCHOR SENSITIVITY OF THE TANGENT-SPACE COMPARATOR

This experiment probes whether the tangent-space comparator is sensitive to the choice of anchor, which is the only externally specified geometric parameter besides the JL matrix. We sweep the anchor's geodesic distance from the origin from 0.0 to 2.0, evaluate at three local radii $R \in \{0.25, 0.5, 0.75\}$, and fix $k = 32$. Figure 8 shows that mean distortion is essentially flat up to anchor distance ~ 1.0 (the curvature length scale $1/\sqrt{c} = 1$ at $c = 1$) and grows only modestly at distance 2.0. The pattern matches the local-JL prediction that the bound depends on the local-radius parameter R rather than on the anchor's absolute position. This experiment does not replace a RAP-specific comparison among the canonical origin, a Fréchet mean, and a semantic root; we treat that comparison as future work.

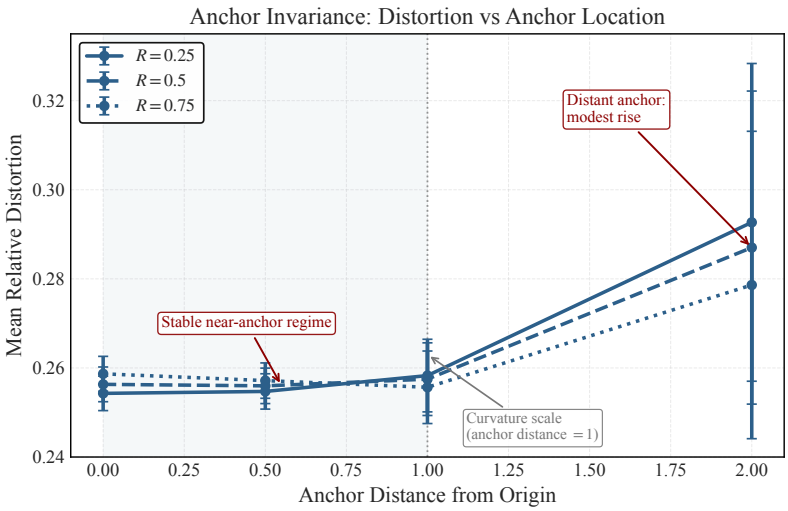


Figure 8: Anchor invariance: mean distortion as a function of the anchor's geodesic distance from the origin, for local radii $R \in \{0.25, 0.5, 0.75\}$ and $k = 32$ over 20 seeds. Distortion remains stable for anchor distances up to 1.0 and increases modestly at distance 2.0.

F.6 Sample-Size Behavior of Mean Distortion

This experiment isolates the tangent-space comparator on local balls of radius $R = 0.5$ and varies the number of evaluated points n and the target dimension $k \in \{8, 16, 32, 64\}$. Figure 9 shows that the mean relative distortion reported in the main tables decreases as n grows. The effect is a concentration effect of the reported sample average rather than a strengthening of the underlying guarantee: the worst-pair distortion, which the theorem actually controls, is by definition non-decreasing in n, while the mean over $\binom{n}{2}$ pair-wise relative errors concentrates around its expectation as more pairs are averaged. The diagnostic clarifies the metric used in the main tables and shows that the n-trend is the standard JL averaging behavior, not an artifact specific to hyperbolic projection.

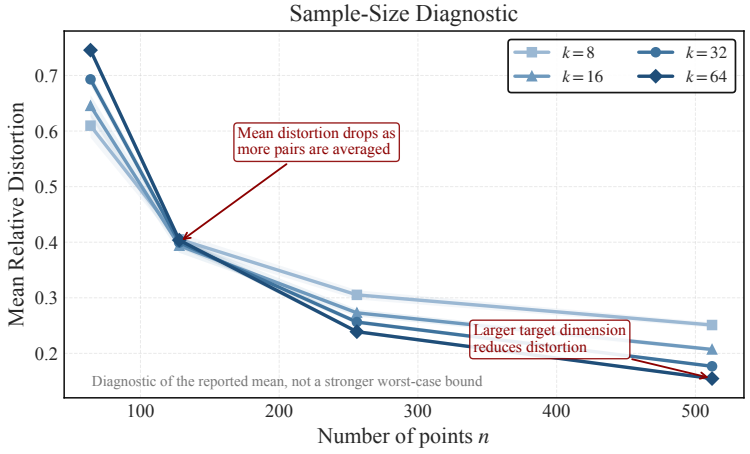


Figure 9: Supplementary sample-size diagnostic for the tangent-space comparator. Mean distortion is shown for $k \in \{8, 16, 32, 64\}$ with $R = 0.5$ over 20 seeds. The downward trend reflects concentration of the reported sample average over more pairs, rather than a stronger worst-case preservation guarantee.

F.7 Modern Vision–Language Evaluation Details

The frozen-representation evaluation uses released embeddings from MERU ViT-S, ViT-B, and ViT-L (Desai et al., 2023) and HyCoCLIP ViT-S and ViT-B (Pal et al., 2025). Each representation contains 512 Lorentz spatial coordinates at curvature $c = 0.1$; the time coordinate is reconstructed from the hyperboloid constraint, giving a 513-coordinate point. The input audit covers all 150 feature files and gives a maximum hyperboloid-constraint error of 2.9×10^{-6}.

We evaluate 13 classification datasets: FGVC-Aircraft (Maji et al., 2013), Caltech-101 (Fei-Fei et al., 2004), CIFAR-10 and CIFAR-100 (Krizhevsky, 2009), CLEVR-Counts (Johnson et al., 2017), Country211 (Radford et al., 2021), Describable Textures Dataset (DTD) (Cimpoi et al., 2014), EuroSAT (Helber et al., 2019), Oxford Flowers-102 (Nilsback & Zisserman, 2008), Food-101 (Bossard et al., 2014), MNIST (LeCun et al., 1998), PatchCamelyon (PCam) (Veeling et al., 2018), and Rendered SST-2 (Socher et al., 2013; Radford et al., 2021). We also evaluate bidirectional image–text retrieval on COCO (Lin et al., 2014) and Flickr30K (Young et al., 2014). Across the five backbones, these datasets produce 65 classification and 10 retrieval backbone–dataset cells.

RAP and the two vanilla hyperbolic projections use target spatial dimensions $k \in \{64, 128, 256\}$ and five paired projection seeds. All methods receive the same frozen features in each cell. Classification uses the original zero-shot accuracy interface. Retrieval reports the mean of text-to-image and image-to-text Recall@1, Recall@5, and Recall@10. Table 6 compares RAP with the tangent-space vanilla method (Section 4.2) and the fully hyperbolic vanilla method (Section 4.3). For each task, method, target dimension, and seed, we take an equally weighted mean over the backbone–dataset cells. We report the mean and sample standard deviation of these five seed-level macro scores, rather than the variation across datasets. Relative gains use the ratio of unrounded macro means, with the larger mean of the two hyperbolic baselines as the denominator; they are not averages of per-cell relative gains. The fully hyperbolic baseline has the higher macro mean in all six task–dimension settings. RAP's relative classification gains are 1.74%, 0.43%, and 0.49% at $k = 64, 128, 256$, respectively; the corresponding retrieval gains are 20.14%, 11.00%, and 5.50%.

Table 6: Frozen-representation evaluation against vanilla hyperbolic projections. Classification reports accuracy (%); retrieval averages Recall@$\{1, 5, 10\}$ (%) in both directions. Entries are mean ± sample standard deviation over five paired projection seeds, each macro-averaging 65 classification or 10 retrieval backbone–dataset cells. Higher is better; bold marks the best compressed mean. Relative gain is $100(S_{\mathrm{RAP}}/S_{\mathrm{best}} - 1)$, where S_{best} is the larger unrounded macro mean of the two hyperbolic baselines. Uncompressed reference scores are 26.21 for classification and 31.44 for retrieval.

Task	k	Tangent-space vanilla	Fully hyperbolic vanilla	RAP	Rel. gain
Classification	64	20.12 ± 0.34	20.14 ± 0.35	$\mathbf{20.49} \pm 0.28$	+1.74%
Classification	128	22.64 ± 0.31	22.65 ± 0.31	$\mathbf{22.74} \pm 0.35$	+0.43%
Classification	256	24.23 ± 0.19	24.24 ± 0.20	$\mathbf{24.36} \pm 0.18$	+0.49%
Retrieval	64	15.17 ± 0.63	15.24 ± 0.63	$\mathbf{18.31} \pm 0.14$	+20.14%
Retrieval	128	21.03 ± 0.41	21.10 ± 0.41	$\mathbf{23.42} \pm 0.19$	+11.00%
Retrieval	256	25.36 ± 0.43	25.41 ± 0.43	$\mathbf{26.81} \pm 0.27$	+5.50%

The 75 backbone–dataset cells, three hyperbolic projection methods, three target dimensions, and five seeds yield 3,375 compressed evaluations, all of which are present in the evaluation and geometry audits. The largest output hyperboloid-constraint error is 5.8×10^{-6}. RAP's radial-distance drift is at most 1.8×10^{-7} across the drift audit, consistent with exact radius preservation up to floating-point precision. These are frozen-embedding compression experiments rather than end-to-end retraining results.

F.8 Exact Worst-Pair and Scaling Audits

Exact worst-pair audit. We repeat the broad-radius setting at $n = 256$, $d = 128$, $k = 32$, radial scale 2.0, and $c = 1$ with independent data and projection random streams. Every one of the 32,640 unordered pairs is evaluated for each of 20 paired seeds. The maximum column in Table 7 is the

mean and standard deviation of the per-seed global maxima, rather than a maximum over sampled pairs.

Table 7: Exact all-pair distortion audit. Lower is better.

Method	Mean	P95	P99	Exact maximum
Local tangent	0.098	0.241	0.316	0.483 ± 0.062
Fully hyperbolic	0.059	0.161	0.232	0.428 ± 0.055
RAP (Ours)	**0.034**	**0.099**	**0.149**	$\mathbf{0.335} \pm 0.038$

RAP lowers the exact maximum by 21.8% relative to the fully hyperbolic comparator. The gain therefore extends beyond average distortion: RAP also reduces the largest pairwise error in this all-pair evaluation, improving both typical and worst-pair distance preservation. The realized minimum input separation is 0.0197 ± 0.0133 across seeds, with a minimum of 0.00228, making explicit the near-pair conditioning behind the tail metric.

The worst-pair advantage also holds in both distance regimes (Table 8). For each seed, we sort all pairs by their original hyperbolic distance and take the nearest and farthest 20%, giving $6{,}528$ pairs per bin. RAP attains the lowest exact maximum in both bins, showing that its advantage is not confined to either nearby or distant pairs.

Table 8: Exact worst-pair relative distortion within the nearest and farthest 20% of input pairs, using the all-pair setup in Table 7. Entries are mean $\pm$ sample standard deviation of the per-seed bin maxima over 20 paired seeds. Lower is better; bold marks the best mean.

Method	Near exact max. ↓	Far exact max. ↓
Local tangent	0.459 ± 0.070	0.389 ± 0.041
Fully hyperbolic	0.426 ± 0.056	0.225 ± 0.029
RAP (Ours)	$\mathbf{0.328} \pm 0.038$	$\mathbf{0.193} \pm 0.031$

Ambient-dimension and point-count scaling. Table 9 fixes $k = 128$. At $n = 512$, all $130{,}816$ unordered pairs are evaluated over 20 paired seeds. For $n \geq 10^5$, every point is projected and one million distinct pairs are sampled per seed over five paired seeds. The sampled rows therefore support large-scale mean and tail estimates, not claims about the global worst pair among all $O(n^2)$ pairs.

Table 9: High-dimensional and large-scale mean relative distortion (lower is better). "All" denotes exact evaluation of every unordered pair; "1M sampled" denotes one million distinct pairs per seed.

d	n	RAP	Fully hyp.	Reduction	Pair evaluation
768	512	**0.0194**	0.0379	48.7%	All
1,024	512	**0.0195**	0.0380	48.7%	All
4,096	512	**0.0197**	0.0378	48.0%	All
1,024	100,000	**0.0196**	0.0381	48.5%	1M sampled
1,024	1,000,000	**0.0196**	0.0381	48.5%	1M sampled
4,096	1,000,000	**0.0196**	0.0381	48.5%	1M sampled

Table 10 reports exact worst-pair distortion for the full-pair $n = 512$ runs. For each seed and method, we take the maximum over all pairs and all three radial scales $\{0.5, 1.0, 2.0\}$ before aggregating across seeds. RAP attains the lowest exact maximum at every tested ambient dimension, extending its worst-pair advantage to high-dimensional representations.

Table 10: Exact worst-pair relative distortion at increasing ambient dimension d, with $n = 512$ and $k = 128$. Each radial scale evaluates all 130,816 unordered pairs. Entries are mean $\pm$ sample standard deviation of the per-seed maxima over all pairs and radial scales across 20 paired seeds. Lower is better; bold marks the best mean.

Method	$d = 768$	$d = 1{,}024$	$d = 4{,}096$
Local tangent	0.268 ± 0.017	0.274 ± 0.016	0.277 ± 0.028
Fully hyperbolic	0.257 ± 0.013	0.265 ± 0.018	0.272 ± 0.028
RAP (Ours)	$\mathbf{0.190} \pm 0.012$	$\mathbf{0.195} \pm 0.019$	$\mathbf{0.189} \pm 0.009$

G Property Preservation Beyond Distance Distortion

G.1 Quantitative Analysis of Preserved Properties

The main experiments report mean relative distance distortion. We also ask whether the same projections preserve geometric properties that make Euclidean random projections useful beyond distance compression: neighborhood rankings, hierarchy/radius structure, and low-dimensional spectral signatures. These evaluations are motivated by analyses of random projections for manifold-modeled data, clustering, and margin-based learning (Baraniuk & Wakin, 2009; Boutsidis et al., 2010; Paul et al., 2013), together with rank-based quality criteria for dimensionality reduction (Lee & Verleysen, 2009). All results below use $k = 32$ and report mean$\pm$std over 20 random projection seeds.

For neighborhood structure, we compare pairwise distance rankings by per-point Spearman correlation, kNN@10 recall, trustworthiness, and continuity. For hierarchy structure, we measure radial error, the Spearman correlation between tree depth and hyperbolic radius, whether the parent of a node remains among its ten nearest neighbors, and the mean rank of the parent. For spectral structure, we form the distance kernel $\exp(-d/\tau)$, normalize its eigenvalues, and compare effective rank, participation ratio, and the top-20 spectral profile; effective rank follows the entropy-based notion of Roy & Vetterli (2007). Euclidean JL is included as a cross-geometry baseline, but it has no hyperbolic radius after projection, so radius-based entries are omitted.

Table 11: Neighborhood and rank preservation at target dimension $k = 32$. Higher values are better; bold marks the best mean in each dataset–metric block.

Dataset	Scheme	Rank $\rho\uparrow$	kNN@10 $\uparrow$	Trust. $\uparrow$	Cont. $\uparrow$
Balanced tree	Euclidean JL	0.770 ± 0.028	0.762 ± 0.013	0.988 ± 0.003	0.991 ± 0.002
	Local tangent-space	0.772 ± 0.028	0.767 ± 0.014	0.989 ± 0.003	0.992 ± 0.002
	Fully hyperbolic	0.782 ± 0.028	0.770 ± 0.014	0.989 ± 0.003	$\mathbf{0.992} \pm 0.002$
	RAP (Ours)	$\mathbf{0.825} \pm 0.026$	$\mathbf{0.775} \pm 0.015$	$\mathbf{0.990} \pm 0.003$	$\mathbf{0.992} \pm 0.002$
Clustered Gaussian	Euclidean JL	0.882 ± 0.037	0.595 ± 0.012	0.987 ± 0.001	$\mathbf{0.988} \pm 0.001$
	Local tangent-space	0.911 ± 0.031	0.598 ± 0.011	$\mathbf{0.987} \pm 0.001$	$\mathbf{0.988} \pm 0.001$
	Fully hyperbolic	0.925 ± 0.023	0.600 ± 0.011	$\mathbf{0.987} \pm 0.001$	$\mathbf{0.988} \pm 0.001$
	RAP (Ours)	$\mathbf{0.951} \pm 0.014$	$\mathbf{0.602} \pm 0.012$	$\mathbf{0.988} \pm 0.001$	$\mathbf{0.988} \pm 0.001$
WordNet subtree	Euclidean JL	0.798 ± 0.023	0.605 ± 0.015	0.968 ± 0.003	0.975 ± 0.003
	Local tangent-space	0.802 ± 0.023	0.604 ± 0.015	0.969 ± 0.003	0.975 ± 0.003
	Fully hyperbolic	0.810 ± 0.022	0.607 ± 0.015	0.970 ± 0.003	0.975 ± 0.003
	RAP (Ours)	$\mathbf{0.864} \pm 0.013$	$\mathbf{0.623} \pm 0.014$	$\mathbf{0.976} \pm 0.002$	$\mathbf{0.979} \pm 0.002$

Overall, RAP is strongest on hierarchy-aware analyses: exact radial preservation gives near-zero radius error and consistently high depth–radius correlation. The neighborhood and spectral analyses are not formal JL guarantees, but they give a useful sanity check that RAP's lower distance distortion also translates into more stable local order and kernel-spectrum behavior on the tested hyperbolic representations.

G.2 Concrete WordNet Hierarchy Case Study

Figure 10 gives a concrete visual case study for the hierarchy/radius interpretation. We take a small union of real WordNet hypernym paths rooted at `animal.n.01`, embed this path union with the

Table 12: Hierarchy and radius preservation on tree-structured data at target dimension $k = 32$. Euclidean JL has no hyperbolic radius, so radius-based entries are omitted.

Dataset	Scheme	Radius err. ↓	Depth–radius ρ ↑	Parent@10 ↑	Parent rank ↓
Balanced tree	Euclidean JL	–	–	$\mathbf{1.000} \pm 0.001$	1.609 ± 0.045
	Local tangent-space	0.101 ± 0.011	0.559 ± 0.048	$\mathbf{1.000} \pm 0.001$	1.565 ± 0.044
	Fully hyperbolic	0.091 ± 0.010	0.578 ± 0.045	$\mathbf{1.000} \pm 0.001$	1.562 ± 0.047
	RAP (Ours)	$\mathbf{0.000} \pm 0.000$	$\mathbf{0.720} \pm 0.000$	$\mathbf{1.000} \pm 0.000$	$\mathbf{1.555} \pm 0.039$
WordNet subtree	Euclidean JL	–	–	0.987 ± 0.005	1.740 ± 0.067
	Local tangent-space	0.101 ± 0.010	0.764 ± 0.033	$\mathbf{0.989} \pm 0.005$	1.673 ± 0.071
	Fully hyperbolic	0.094 ± 0.009	0.775 ± 0.030	$\mathbf{0.988} \pm 0.005$	1.664 ± 0.068
	RAP (Ours)	$\mathbf{0.000} \pm 0.000$	$\mathbf{0.892} \pm 0.000$	0.987 ± 0.004	$\mathbf{1.626} \pm 0.056$

Table 13: Spectral-signature preservation for the distance kernel $\exp(-d/\tau)$ at target dimension $k = 32$. Lower values are better.

Dataset	Scheme	Eff. rank err. ↓	PR err. ↓	Top-20 L_1 ↓
Balanced tree	Euclidean JL	0.159 ± 0.035	0.105 ± 0.041	0.049 ± 0.008
	Local tangent-space	0.104 ± 0.040	0.067 ± 0.042	0.038 ± 0.009
	Fully hyperbolic	0.103 ± 0.036	0.064 ± 0.039	0.037 ± 0.008
	RAP (Ours)	$\mathbf{0.055} \pm 0.032$	$\mathbf{0.026} \pm 0.023$	$\mathbf{0.030} \pm 0.006$
Clustered Gaussian	Euclidean JL	0.199 ± 0.025	0.107 ± 0.035	0.059 ± 0.010
	Local tangent-space	0.067 ± 0.035	0.044 ± 0.032	0.040 ± 0.011
	Fully hyperbolic	0.065 ± 0.026	0.037 ± 0.025	0.038 ± 0.009
	RAP (Ours)	$\mathbf{0.049} \pm 0.025$	$\mathbf{0.015} \pm 0.011$	$\mathbf{0.031} \pm 0.005$
WordNet subtree	Euclidean JL	0.142 ± 0.036	0.082 ± 0.037	0.048 ± 0.005
	Local tangent-space	0.108 ± 0.039	0.064 ± 0.031	0.041 ± 0.005
	Fully hyperbolic	0.107 ± 0.036	0.061 ± 0.030	0.041 ± 0.005
	RAP (Ours)	$\mathbf{0.072} \pm 0.017$	$\mathbf{0.018} \pm 0.011$	$\mathbf{0.032} \pm 0.003$

same hyperbolic tree routine used in the WordNet benchmark, and apply the three projection rules to a two-dimensional target for visualization. This is not a replacement for the full WordNet sweep in Table 2; rather, it makes the qualitative meaning of radial depth explicit through familiar words such as dog, eagle, salmon, and bee.

The qualitative WordNet visualization in Figure 2 complements these property tables by showing the hierarchy rings directly. Figure 5 in the main text isolates the radial component by plotting each node's original radius against its projected radius. The RAP points lie on the diagonal by construction, while the two comparators introduce small but systematic radial displacement.

H Proofs for the Tangent-Space Comparator

We state and prove each step as a self-contained lemma.

H.1 Source-Side Metric Comparison

Lemma 3 (Local metric comparison in constant curvature)**.** *Let $\mathbf{u}, \mathbf{v} \in T_a\mathbb{H}^d_c$ with $\|\mathbf{u}\|, \|\mathbf{v}\| \le R$ and $R\sqrt{c} \le R_0$ for a sufficiently small universal constant $R_0 \le 1$. Then:*

1. *(Global lower bound) $\|\mathbf{u} - \mathbf{v}\| \le d_c(\exp_a \mathbf{u}, \exp_a \mathbf{v})$.*
2. *(Local upper bound) $d_c(\exp_a \mathbf{u}, \exp_a \mathbf{v}) \le (1 + C_0 c R^2)\|\mathbf{u} - \mathbf{v}\|$ where C_0 depends only on R_0.*

Proof. Lower bound. By the Cartan–Hadamard theorem, in a simply connected manifold of nonpositive sectional curvature, the exponential map at any point is a distance-expanding map: $d_c(\exp_a \mathbf{u}, \exp_a \mathbf{v}) \ge \|\mathbf{u} - \mathbf{v}\|$. This holds globally, not just for small R.

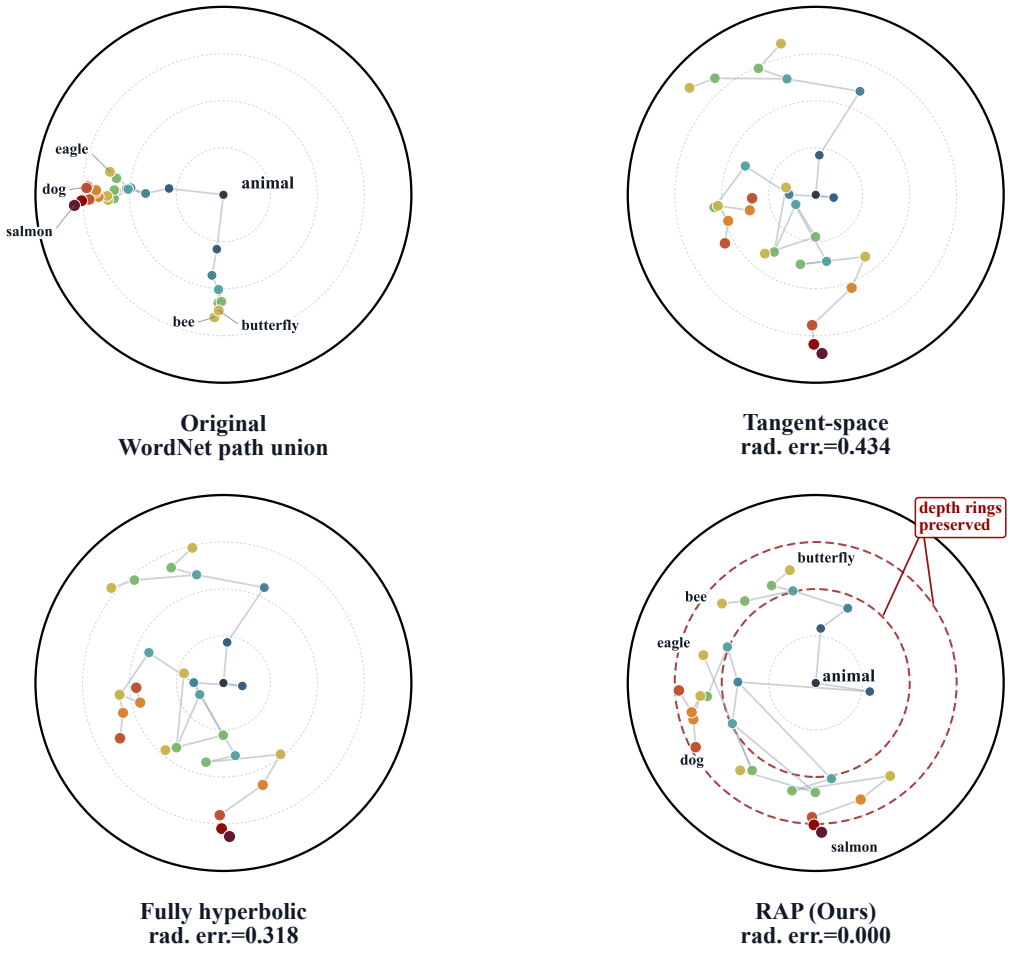


Figure 10: Concrete WordNet projection case study in Poincare disks. Nodes follow selected hypernym paths under `animal.n.01`; node color encodes WordNet depth, so same-level nodes share the same color. The original panel is a radial two-dimensional view of the input tree embedding, while the remaining panels are actual $k = 2$ hyperbolic projection outputs. The tangent-space and fully hyperbolic projections introduce visible radial drift; RAP preserves the depth-carrying radial coordinate exactly, with red dashed rings highlighting the preserved depth rings.

Upper bound. Write $\mathbf{u} = r\boldsymbol{\xi}$ and $\mathbf{v} = s\boldsymbol{\eta}$ with $\|\boldsymbol{\xi}\| = \|\boldsymbol{\eta}\| = 1$, and let $\theta \in [0, \pi]$ be the angle between them ($\cos\theta = \langle \boldsymbol{\xi}, \boldsymbol{\eta} \rangle$). The hyperbolic law of cosines in constant curvature $-c$ gives

$$\cosh(\sqrt{c}\, d_c(\exp_a \mathbf{u}, \exp_a \mathbf{v})) = \cosh(\sqrt{c}\, r)\cosh(\sqrt{c}\, s) - \sinh(\sqrt{c}\, r)\sinh(\sqrt{c}\, s)\cos\theta. \tag{28}$$

Taylor-expand around $\sqrt{c}\, r$ and $\sqrt{c}\, s$ small. Using $\cosh(t) = 1 + \frac{t^2}{2} + \frac{t^4}{24} + \cdots$ and $\sinh(t) = t + \frac{t^3}{6} + \cdots$, the right-hand side becomes

$$\begin{aligned}
&\left(1 + \frac{cr^2}{2} + O(c^2 r^4)\right)\left(1 + \frac{cs^2}{2} + O(c^2 s^4)\right) \\
&\quad - \left(\sqrt{c}\, r + O(c^{3/2} r^3)\right)\left(\sqrt{c}\, s + O(c^{3/2} s^3)\right)\cos\theta \\
&= 1 + \frac{c}{2}(r^2 + s^2) - cs\, r\cos\theta + O(c^2 R^4) \\
&= 1 + \frac{c}{2}(r^2 + s^2 - 2rs\cos\theta) + O(c^2 R^4) \\
&= 1 + \frac{c}{2}\|\mathbf{u} - \mathbf{v}\|^2 + O(c^2 R^4),
\end{aligned}$$

where we used $\|\mathbf{u}-\mathbf{v}\|^2 = r^2 + s^2 - 2rs\cos\theta$.

Now set $D = d_c(\exp_a \mathbf{u}, \exp_a \mathbf{v})$ and $L = \|\mathbf{u}-\mathbf{v}\|$. We have

$$\cosh(\sqrt{c}\, D) = 1 + \frac{c}{2}L^2 + E, \qquad |E| \le C' c^2 R^4, \tag{29}$$

for a universal C'. Using $\operatorname{arcosh}(1+z) = \sqrt{2z}\,(1+O(z))$ for small $z > 0$, and noting that $z = \frac{c}{2}L^2 + E$ with $L \le 2R$ gives $z \le 2cR^2 + C'c^2R^4 \le 3cR^2$ (for $R_0 \le 1$), we obtain

$$D = L\sqrt{1 + \frac{2E}{cL^2}}\,(1 + O(cR^2)). \tag{30}$$

If $L = 0$ (i.e., $\mathbf{u} = \mathbf{v}$), then $D = 0$ as well, so the bound $D \le (1 + C_0 cR^2)L$ holds trivially. For $L > 0$ we handle two cases. If $L \ge c^{1/2}R^2$, then $|2E/(cL^2)| \le 2C'c^2R^4/(c \cdot cR^4) = 2C'$, which is $O(1)$ and the square root contributes only an $O(cR^2)$ factor.

If $L < c^{1/2}R^2$, a tighter estimate of the error E is needed. Returning to the Taylor expansion of the right-hand side of Equation (28), we collect all terms beyond the leading $1 + \frac{c}{2}L^2$ more carefully. The next-order contribution is

$$E = \frac{c^2}{4} r^2 s^2 \sin^2\theta + O(c^3R^6), \tag{31}$$

where $\sin^2\theta = 1 - \cos^2\theta$. We claim $r^2s^2\sin^2\theta \le R^2L^2$. Indeed, $rs|\sin\theta|$ is twice the area of the triangle with sides r, s and included angle θ, while $L = \|\mathbf{u}-\mathbf{v}\|$ is the third side, so $rs|\sin\theta| = rL\sin\alpha$ for some angle α, giving $rs|\sin\theta| \le rL \le RL$. Hence

$$|E| \le \tfrac{c^2R^2}{4} L^2 + O(c^3R^6). \tag{32}$$

Substituting back, $\frac{2|E|}{cL^2} \le \frac{cR^2}{2} + O(c^2R^4)$, which is $O(cR^2)$ uniformly for $R\sqrt{c} \le R_0 \le 1$, regardless of the size of L. With this refined bound, the identity $\cosh(\sqrt{c}\,D) = 1 + \frac{c}{2}L^2(1+O(cR^2))$ holds in both regimes. Applying $\operatorname{arcosh}(1+z) = \sqrt{2z}\,(1+O(z))$ with $z = \frac{c}{2}L^2(1+O(cR^2)) = O(cR^2)$ gives $D = L(1+O(cR^2))$, hence $D \le (1 + C_0cR^2)L$ with a universal C_0 (which absorbs both the refined error E from the Taylor expansion and the remainder from the arcosh inversion). □

H.2 JL with the Origin Included

Lemma 4 (Euclidean JL with norm control)**.** *Let $\{\mathbf{u}_1, \dots, \mathbf{u}_n\} \subset \mathbb{R}^d$ and let $\mathbf{P} \in \mathbb{R}^{k\times d}$ be a standard JL matrix with $k \ge C\varepsilon^{-2}\log((n+1)/\delta)$. Then with probability at least $1-\delta$, the JL event holds simultaneously for the augmented set $\{0, \mathbf{u}_1, \dots, \mathbf{u}_n\}$. In particular, for all i:*

$$\|\mathbf{P}\mathbf{u}_i\| = \|\mathbf{P}\mathbf{u}_i - \mathbf{P}\cdot 0\| \le (1+\varepsilon)\|\mathbf{u}_i\|. \tag{33}$$

Proof. The augmented set has $n+1$ points, hence $\binom{n+1}{2} \le n^2$ pairs. A standard JL matrix with $k \ge C\varepsilon^{-2}\log(n^2/\delta) = C\varepsilon^{-2}(2\log n + \log(1/\delta))$ preserves all pairwise distances with probability $\ge 1-\delta$, by a union bound over all pairs. The norm bound follows from the pair $(0, \mathbf{u}_i)$. □

H.3 Target-Side Comparison

Lemma 5 (Target-side metric comparison)**.** *Under the assumption $R' = (1+\varepsilon)R \le R_1/\sqrt{c}$, the same bounds as Lemma 3 apply in $\mathbb{H}^k_c$ around $\tilde{a}$ for vectors of norm at most R':*

$$\|\mathbf{w}-\mathbf{w}'\| \le d_c(\exp_{\tilde{a}}\mathbf{w}, \exp_{\tilde{a}}\mathbf{w}') \le (1 + C_0 cR'^2)\|\mathbf{w}-\mathbf{w}'\|. \tag{34}$$

Proof. Lemma 3 applies to any hyperbolic space of constant curvature $-c$, regardless of dimension. The target space $\mathbb{H}^k_c$ has the same curvature, so the same derivation holds with R' in place of R. □

Remark 6 (Explicit values of C_0 and R_0). Tracing through the Taylor expansion in Lemma 3, the universal constant C' in the error bound $|E| \le \frac{c^2R^2}{4}L^2 + O(c^3R^6)$ satisfies $C' \le 1$ for $R_0 \le 1$. The arcosh inversion step introduces a further $(1 + O(cR^2))$ factor. Taking $R_0 = 1/(2\sqrt{c})$ (i.e.,

$R\sqrt{c} \le 1/2$) keeps all higher-order remainders below their leading terms, and one may verify that $C_0 = 1$ suffices in this regime. Concretely, the condition $R\sqrt{c} \le 1/2$ means the geodesic ball radius is at most half the curvature length scale $1/\sqrt{c}$. For the experiments in this paper ($c = 1$, $R \le 1.5$), the value $R\sqrt{c} = 1.5$ exceeds this strict threshold; nonetheless the bound remains empirically conservative, which is consistent with the constants being pessimistic. The factor $C_1 = 5C_0 = 5$ used in the bound-validation diagnostic arises from the error composition lemma: the source side contributes $C_0 c R^2$, the target side contributes $C_0 c R'^2 \le 4C_0 c R^2$ (since $R' \le 2R$ for $\varepsilon \le 1$), and the cross-term adds at most $C_0 c R^2$, yielding $C_1 = 1 + 4 + \text{cross} \le 5C_0$ as a safe upper envelope. In the code this appears as the hardcoded multiplier 5.0.

H.4 Error Composition

Lemma 7 (Abstract composition of distortion factors)**.** *Let $\alpha, \beta \ge 0$ with $\alpha \le 1$ and $\beta \le 1$. If*

$$\frac{1-\alpha}{1+\beta} \le \frac{d_c(\Pi x_i, \Pi x_j)}{d_c(x_i, x_j)} \le (1+\alpha)(1+\beta), \tag{35}$$

then

$$1 - \alpha - \beta \le \frac{d_c(\Pi x_i, \Pi x_j)}{d_c(x_i, x_j)} \le 1 + \alpha + \beta + \alpha\beta. \tag{36}$$

In particular, if $\alpha \le 1$, then $1 + \alpha + \beta + \alpha\beta \le 1 + \alpha + 2\beta$.

Proof. Lower bound. For $\alpha, \beta \in [0, 1]$, $\frac{1-\alpha}{1+\beta} \ge 1 - \alpha - \beta$ because $(1 - \alpha - \beta)(1 + \beta) = 1 - \alpha - \beta^2 - \alpha\beta \le 1 - \alpha$.

Upper bound. $(1 + \alpha)(1 + \beta) = 1 + \alpha + \beta + \alpha\beta$, which gives the first form. For the simplified bound, $\alpha\beta \le \beta$ when $\alpha \le 1$, so $1 + \alpha + \beta + \alpha\beta \le 1 + \alpha + 2\beta$. □

Corollary 8 ($C_1 = 5C_0$)**.** *In the proof of Theorem 1, the source-side curvature error is $\beta_{\text{src}} = C_0 c R^2$ (Lemma 3) and the target-side curvature error is $\beta_{\text{tgt}} = C_0 c R'^2$ with $R' = (1+\varepsilon)R \le 2R$ (Lemma 5), so $\beta_{\text{tgt}} \le 4C_0 c R^2$. Applying Lemma 7 with $\alpha = \varepsilon$ and $\beta = \beta_{\text{src}} + \beta_{\text{tgt}} \le 5C_0 c R^2$ (upper-bounding the composition of the source and target distortion by their sum plus cross-term, as in the proof of Theorem 1) yields $C_1 = 5C_0$.*

Combining Lemmas 3–7 completes the proof of Theorem 1. The constant C_1 in the theorem equals $5C_0$, where C_0 is the metric comparison constant from Lemma 3.

I Fully Hyperbolic Comparator Conditioning

The fully hyperbolic comparator stays on the hyperboloid by renormalization, so its weakness is not loss of the manifold constraint but that hyperbolic distance is read off through arcosh of a Lorentz inner product. Two lemmas isolate the resulting conditioning issue.

Lemma 9 (Arcosh conditioning)**.** *Let $z \ge 1 + \rho$ and let $|\Delta| \le \rho/2$ for some $\rho > 0$. Then*

$$|\operatorname{arcosh}(z + \Delta) - \operatorname{arcosh}(z)| \le \frac{|\Delta|}{\sqrt{\rho(2+\rho)}} \left(1 + O\left(\frac{|\Delta|}{\rho}\right)\right). \tag{37}$$

In particular, when $z = 1 + \rho$ with $\rho \ll 1$, the first-order sensitivity is of order $|\Delta|/\sqrt{\rho}$.

Proof. The derivative of arcosh is

$$\frac{d}{dz} \operatorname{arcosh}(z) = \frac{1}{\sqrt{z-1}\sqrt{z+1}}. \tag{38}$$

For t between z and $z + \Delta$, the assumption $|\Delta| \le \rho/2$ gives $t - 1 \ge \rho/2$ and $t + 1 \ge 2 + \rho/2$. The mean value theorem therefore gives a Lipschitz bound of order $|\Delta|/\sqrt{\rho(2+\rho)}$. Expanding the denominator around $z = 1 + \rho$ yields the displayed first-order expression. For nearby pairs in hyperbolic space, $z = -c\langle x, y\rangle_L = \cosh(\sqrt{c}\, d_c(x, y)) = 1 + O(c\, d_c(x, y)^2)$, so small Lorentz-product errors are amplified by roughly the inverse distance scale. □

Lemma 10 (Renormalized fully hyperbolic sketches mix radial and angular errors)**.** *For the fully hyperbolic comparator*

$$\Pi_{\mathbf{A}}(x) = \left(\sqrt{\frac{1}{c} + \|\mathbf{A}\bar{\mathbf{x}}\|_2^2},\ \mathbf{A}\bar{\mathbf{x}} \right), \tag{39}$$

the perturbed Lorentz product between two projected points is

$$-c\langle \Pi_{\mathbf{A}}(x), \Pi_{\mathbf{A}}(y)\rangle_L = c\sqrt{\frac{1}{c} + \|\mathbf{A}\bar{\mathbf{x}}\|_2^2}\sqrt{\frac{1}{c} + \|\mathbf{A}\bar{\mathbf{y}}\|_2^2} - c\langle \mathbf{A}\bar{\mathbf{x}}, \mathbf{A}\bar{\mathbf{y}}\rangle. \tag{40}$$

Thus the same spatial sketch controls both the radial time coordinate and the spatial angular term.

Proof. The identity follows by substituting the renormalized coordinates into the Lorentz product $\langle x, y\rangle_L = -x_0 y_0 + \langle \bar{\mathbf{x}}, \bar{\mathbf{y}}\rangle$. The square-root terms depend on the sketched norms $\|\mathbf{A}\bar{\mathbf{x}}\|_2$ and $\|\mathbf{A}\bar{\mathbf{y}}\|_2$, while the final term depends on the sketched Euclidean inner product. Hence a single Euclidean sketch perturbs radial magnitude and angular correlation simultaneously. Lemma 9 then explains why these perturbations are difficult to turn into a uniform multiplicative distance guarantee across all distance scales. □

Composing the two lemmas yields a multiplicative envelope for the fully hyperbolic comparator on a finite point set. The statement parallels the finite-set RAP guarantee in Appendix J and isolates the same $1/[\sqrt{c}\,\tau \sinh(\sqrt{c}\,\tau)]$ conditioning factor.

Corollary 11 (Fully hyperbolic comparator multiplicative envelope)**.** *Let $X = \{x_1, \dots, x_n\} \subset \mathbb{H}_c^d$ have polar coordinates $x_i = (r_i, \mathbf{u}_i)$ with $r_i \le R$ for all i and $d_c(x_i, x_j) \ge \tau > 0$ for all $i \ne j$. Fix $\varepsilon \in (0, 1]$ and let $\mathbf{A} \in \mathbb{R}^{k\times d}$ be a Gaussian sketch with $k \ge C\varepsilon^{-2}\log(n/\delta)$. With probability at least $1 - \delta$, the squared-norm JL event holds simultaneously for the spatial parts and their pairwise differences:*

$$\left| \|\mathbf{A}v\|_2^2 - \|v\|_2^2 \right| \le \varepsilon \|v\|_2^2 \qquad \text{for every} \quad v \in \{\bar{\mathbf{x}}_i\}_{i=1}^n \cup \{\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j\}_{i<j}. \tag{41}$$

If in addition $6\,\varepsilon\,\sinh^2(\sqrt{c}\,R) \le \frac{1}{2}\big(\cosh(\sqrt{c}\,\tau) - 1\big)$, then on the event in Equation (41),

$$\max_{i<j} \left| \frac{d_c(\Pi_{\mathbf{A}} x_i, \Pi_{\mathbf{A}} x_j)}{d_c(x_i, x_j)} - 1 \right| \le \frac{12\,\varepsilon\,\sinh^2(\sqrt{c}\,R)}{\sqrt{c}\,\tau\,\sinh(\sqrt{c}\,\tau)}. \tag{42}$$

The numerator splits into two contributions of comparable order: a radial term of order $\varepsilon\big(\sinh^2(\sqrt{c}\,r_i) + \sinh^2(\sqrt{c}\,r_j)\big)$ from the renormalization (Lemma 10) and a matching angular term of the same form from the spatial inner-product sketch.

Proof. Let $T_i = \sqrt{1/c + \|\bar{\mathbf{x}}_i\|_2^2} = \cosh(\sqrt{c}\,r_i)/\sqrt{c}$ and $\widehat{T}_i = \sqrt{1/c + \|\mathbf{A}\bar{\mathbf{x}}_i\|_2^2}$, and abbreviate $c_\bullet = \cosh(\sqrt{c}\,r_\bullet)$, $s_\bullet = \sinh(\sqrt{c}\,r_\bullet)$. By Lemma 10,

$$-c\langle \Pi_{\mathbf{A}} x_i, \Pi_{\mathbf{A}} x_j\rangle_L = c\,\widehat{T}_i\widehat{T}_j - c\langle \mathbf{A}\bar{\mathbf{x}}_i, \mathbf{A}\bar{\mathbf{x}}_j\rangle. \tag{43}$$

Equation (41) gives $c\,\widehat{T}_i^2 = 1 + c\|\mathbf{A}\bar{\mathbf{x}}_i\|_2^2 = c_i^2 + \varepsilon_i s_i^2$ with $|\varepsilon_i| \le \varepsilon$. Squaring,

$$\big(c\,\widehat{T}_i\widehat{T}_j\big)^2 = (c_i^2 + \varepsilon_i s_i^2)(c_j^2 + \varepsilon_j s_j^2) = (c_i c_j)^2 + \varepsilon_i c_j^2 s_i^2 + \varepsilon_j c_i^2 s_j^2 + \varepsilon_i\varepsilon_j s_i^2 s_j^2. \tag{44}$$

Since $c\,\widehat{T}_i\widehat{T}_j + c_i c_j \ge c_i c_j$ and $|\varepsilon_\bullet| \le \varepsilon \le 1$, dividing the squared identity by $c\widehat{T}_i\widehat{T}_j + c_i c_j$ yields

$$|\Delta_{ij}^{\mathrm{rad}}| := \big|c\,\widehat{T}_i\widehat{T}_j - c_i c_j\big| \le \varepsilon\Big(\tfrac{c_j s_i^2}{c_i} + \tfrac{c_i s_j^2}{c_j}\Big) + \varepsilon^2\,\tfrac{s_i^2 s_j^2}{c_i c_j}. \tag{45}$$

The function $r \mapsto \sinh(r)\tanh(r)$ is non-decreasing on $r \ge 0$, so $\frac{c_j s_i^2}{c_i} + \frac{c_i s_j^2}{c_j} \le s_i^2 + s_j^2$ for all $r_i, r_j \ge 0$. Combining with $\frac{s_i^2 s_j^2}{c_i c_j} \le s_i s_j \le \frac{1}{2}(s_i^2 + s_j^2)$,

$$\begin{aligned} |\Delta_{ij}^{\mathrm{rad}}| &\le \varepsilon\big(\sinh^2(\sqrt{c}\,r_i) + \sinh^2(\sqrt{c}\,r_j)\big) \\ &\quad + \tfrac{\varepsilon^2}{2}\big(\sinh^2(\sqrt{c}\,r_i) + \sinh^2(\sqrt{c}\,r_j)\big) \\ &\le \tfrac{3\varepsilon}{2}\big(\sinh^2(\sqrt{c}\,r_i) + \sinh^2(\sqrt{c}\,r_j)\big), \end{aligned} \tag{46}$$

where the last inequality uses $\varepsilon \le 1$.

For the angular term, polarization $\langle \mathbf{A}\bar{\mathbf{x}}_i, \mathbf{A}\bar{\mathbf{x}}_j\rangle = \big(\|\mathbf{A}\bar{\mathbf{x}}_i\|_2^2 + \|\mathbf{A}\bar{\mathbf{x}}_j\|_2^2 - \|\mathbf{A}(\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j)\|_2^2\big)/2$ combined with Equation (41) and $\|\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j\|_2^2 \le 2(\|\bar{\mathbf{x}}_i\|_2^2 + \|\bar{\mathbf{x}}_j\|_2^2)$ gives

$$\big|\langle \mathbf{A}\bar{\mathbf{x}}_i, \mathbf{A}\bar{\mathbf{x}}_j\rangle - \langle \bar{\mathbf{x}}_i, \bar{\mathbf{x}}_j\rangle\big| \le \tfrac{\varepsilon}{2}\big(\|\bar{\mathbf{x}}_i\|_2^2 + \|\bar{\mathbf{x}}_j\|_2^2 + \|\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j\|_2^2\big) \le \tfrac{3\varepsilon}{2}\big(\|\bar{\mathbf{x}}_i\|_2^2 + \|\bar{\mathbf{x}}_j\|_2^2\big). \tag{47}$$

Multiplying by c and using $\|\bar{\mathbf{x}}_i\|_2^2 = s_i^2/c$,

$$|\Delta_{ij}^{\mathrm{ang}}| := c\big|\langle \mathbf{A}\bar{\mathbf{x}}_i, \mathbf{A}\bar{\mathbf{x}}_j\rangle - \langle \bar{\mathbf{x}}_i, \bar{\mathbf{x}}_j\rangle\big| \le \tfrac{3\varepsilon}{2}\big(\sinh^2(\sqrt{c}\, r_i) + \sinh^2(\sqrt{c}\, r_j)\big). \tag{48}$$

Subtracting the unperturbed identity $c_i c_j - s_i s_j \cos\theta_{ij} = \cosh(\sqrt{c}\, d_c(x_i, x_j))$, the total perturbation $\Delta_{ij} = (c\widehat{T}_i\widehat{T}_j - c_i c_j) - c\big(\langle \mathbf{A}\bar{\mathbf{x}}_i, \mathbf{A}\bar{\mathbf{x}}_j\rangle - \langle \bar{\mathbf{x}}_i, \bar{\mathbf{x}}_j\rangle\big)$ of the arcosh argument satisfies (non-asymptotically)

$$|\Delta_{ij}| \le |\Delta_{ij}^{\mathrm{rad}}| + |\Delta_{ij}^{\mathrm{ang}}| \le 3\varepsilon\big(\sinh^2(\sqrt{c}\, r_i) + \sinh^2(\sqrt{c}\, r_j)\big) \le 6\varepsilon \sinh^2(\sqrt{c}\, R). \tag{49}$$

The hypothesis ensures $|\Delta_{ij}| \le \frac{1}{2}(\cosh(\sqrt{c}\,\tau) - 1) \le \frac{1}{2}(\cosh(\sqrt{c}\, d_c(x_i, x_j)) - 1)$, so the mean-value bound for arcosh along the segment between $-c\langle x_i, x_j\rangle_L$ and $-c\langle \Pi_{\mathbf{A}} x_i, \Pi_{\mathbf{A}} x_j\rangle_L$ (the same bound used in the proof of Lemma 15) yields $|d_c(\Pi_{\mathbf{A}} x_i, \Pi_{\mathbf{A}} x_j) - d_c(x_i, x_j)| \le 2|\Delta_{ij}|/[\sqrt{c}\sinh(\sqrt{c}\,\tau)]$. Dividing by $d_c(x_i, x_j) \ge \tau$ gives Equation (42). □

J RAP Radial Preservation and Angular Sensitivity

We prove Proposition 2 and the finite-set RAP guarantee. The argument has three steps: well-definedness of the projected point on the target hyperboloid, normalized JL preservation of angular inner products, and transfer of the angular error into hyperbolic distance error through the law of cosines.

J.1 Well-Definedness and the Hyperboloid Constraint

We work in the hyperboloid model

$$\mathbb{H}_c^k = \left\{ z = (z_0, \bar{\mathbf{z}}) \in \mathbb{R}^{k+1} : -z_0^2 + \|\bar{\mathbf{z}}\|_2^2 = -\frac{1}{c},\ z_0 > 0 \right\}, \tag{50}$$

with curvature $-c$. The origin is $\tilde{o} = (1/\sqrt{c}, \mathbf{0})$.

Lemma 12 (Polar coordinates give valid hyperboloid points)**.** *For any $r \ge 0$ and any $\mathbf{v} \in \mathbb{S}^{k-1}$, define*

$$\iota_k(r, \mathbf{v}) = \frac{1}{\sqrt{c}}\big(\cosh(\sqrt{c}\, r), \sinh(\sqrt{c}\, r)\mathbf{v}\big). \tag{51}$$

Then $\iota_k(r, \mathbf{v}) \in \mathbb{H}_c^k$ and $d_c(\tilde{o}, \iota_k(r, \mathbf{v})) = r$.

Proof. Let $z = \iota_k(r, \mathbf{v})$. Since $\|\mathbf{v}\|_2 = 1$,

$$-z_0^2 + \|\bar{\mathbf{z}}\|_2^2 = -\frac{1}{c}\cosh^2(\sqrt{c}\, r) + \frac{1}{c}\sinh^2(\sqrt{c}\, r) = -\frac{1}{c}, \tag{52}$$

and $z_0 > 0$. Thus $z \in \mathbb{H}_c^k$. Moreover, $-c\langle \tilde{o}, z\rangle_L = \cosh(\sqrt{c}\, r)$, so

$$d_c(\tilde{o}, z) = \frac{1}{\sqrt{c}}\operatorname{arcosh}\big(\cosh(\sqrt{c}\, r)\big) = r. \tag{53}$$

□

Corollary 13 (RAP is a valid hyperbolic random map on finite sets)**.** *Let $X = \{x_1, \dots, x_n\} \subset \mathbb{H}_c^d$ be fixed, and write $x_i = (r_i, \mathbf{u}_i)$ around an origin o, with an arbitrary direction chosen when $r_i = 0$. If $\|\mathbf{P}\mathbf{u}_i\|_2 > 0$ for every non-anchor point, then RAP maps every x_i into $\mathbb{H}_c^k$ and preserves the anchor distance $d_c(o, x_i) = r_i$ exactly. If $\mathbf{P}$ is Gaussian with entries $\mathcal{N}(0, 1/k)$, this nonzero condition holds for the fixed finite set X with probability one.*

Proof. For each non-anchor point, RAP sets $\widetilde{\mathbf{u}}_i = \mathbf{P}\mathbf{u}_i/\|\mathbf{P}\mathbf{u}_i\|_2$, hence $\widetilde{\mathbf{u}}_i \in \mathbb{S}^{k-1}$. Lemma 12 then implies that $\Pi_{\mathrm{RAP}}(x_i) = \iota_k(r_i, \widetilde{\mathbf{u}}_i)$ lies in $\mathbb{H}^k_c$ and has radius r_i. If $r_i = 0$, RAP maps $x_i = o$ to $\tilde{o}$, so the conclusion is immediate.

For a fixed nonzero $\mathbf{u}_i$, the random vector $\mathbf{P}\mathbf{u}_i$ is a nondegenerate Gaussian vector in $\mathbb{R}^k$, so $\Pr[\mathbf{P}\mathbf{u}_i = \mathbf{0}] = 0$. A union bound over the finite set of non-anchor points gives the probability-one statement. □

J.2 Normalized Angular JL

RAP does not use the projected vector $\mathbf{P}\mathbf{u}$ directly; it renormalizes it back to the target sphere. The next lemma records that the usual JL norm-preservation event is sufficient to preserve angular inner products after this normalization.

Lemma 14 (Normalized JL preserves angular inner products)**.** *Let $\mathbf{u}_1, \dots, \mathbf{u}_n \in \mathbb{S}^{d-1}$ and let $0 < \varepsilon < 1$. Suppose a matrix $\mathbf{P} \in \mathbb{R}^{k\times d}$ satisfies, for all i, j,*

$$\big|\|\mathbf{P}\mathbf{u}_i\|_2^2 - 1\big| \le \varepsilon, \qquad \big|\|\mathbf{P}(\mathbf{u}_i - \mathbf{u}_j)\|_2^2 - \|\mathbf{u}_i - \mathbf{u}_j\|_2^2\big| \le \varepsilon\|\mathbf{u}_i - \mathbf{u}_j\|_2^2. \tag{54}$$

Define $\widetilde{\mathbf{u}}_i = \mathbf{P}\mathbf{u}_i/\|\mathbf{P}\mathbf{u}_i\|_2$. Then, for all i, j,

$$|\langle\widetilde{\mathbf{u}}_i, \widetilde{\mathbf{u}}_j\rangle - \langle\mathbf{u}_i, \mathbf{u}_j\rangle| \le 4\varepsilon. \tag{55}$$

Moreover, if $\mathbf{P}$ is a Gaussian JL matrix, then the displayed hypotheses hold with probability at least $1 - \delta$ whenever $k \ge C\varepsilon^{-2}\log(n/\delta)$ for a universal constant C.

Proof. Write $a_i = \|\mathbf{P}\mathbf{u}_i\|_2^2$, $a_j = \|\mathbf{P}\mathbf{u}_j\|_2^2$, $D_{ij} = \|\mathbf{u}_i - \mathbf{u}_j\|_2^2$, and $\widehat{D}_{ij} = \|\mathbf{P}(\mathbf{u}_i - \mathbf{u}_j)\|_2^2$. By polarization,

$$\langle\mathbf{P}\mathbf{u}_i, \mathbf{P}\mathbf{u}_j\rangle = \frac{a_i + a_j - \widehat{D}_{ij}}{2}, \qquad \langle\mathbf{u}_i, \mathbf{u}_j\rangle = \frac{2 - D_{ij}}{2}. \tag{56}$$

Since $D_{ij} \le 4$, the assumptions imply

$$|\langle\mathbf{P}\mathbf{u}_i, \mathbf{P}\mathbf{u}_j\rangle - \langle\mathbf{u}_i, \mathbf{u}_j\rangle| \le \frac{\varepsilon + \varepsilon + \varepsilon D_{ij}}{2} \le 3\varepsilon. \tag{57}$$

Also $a_i, a_j \in [1-\varepsilon, 1+\varepsilon]$, so $\sqrt{a_i a_j} \in [1-\varepsilon, 1+\varepsilon]$. Because $|\langle\mathbf{P}\mathbf{u}_i, \mathbf{P}\mathbf{u}_j\rangle| \le \sqrt{a_i a_j}$,

$$\left|\frac{\langle\mathbf{P}\mathbf{u}_i, \mathbf{P}\mathbf{u}_j\rangle}{\sqrt{a_i a_j}} - \langle\mathbf{P}\mathbf{u}_i, \mathbf{P}\mathbf{u}_j\rangle\right| \le |1 - \sqrt{a_i a_j}| \le \varepsilon. \tag{58}$$

Combining the two estimates gives the 4ε bound.

For the probability statement, apply the standard JL norm guarantee to the finite collection

$$\mathcal{V} = \{\mathbf{u}_i : 1 \le i \le n\} \cup \{\mathbf{u}_i - \mathbf{u}_j : 1 \le i < j \le n,\ \mathbf{u}_i \ne \mathbf{u}_j\}. \tag{59}$$

This collection has at most $n + n^2$ vectors, so $k \ge C\varepsilon^{-2}\log(n/\delta)$ suffices after a union bound. □

J.3 From Angular Error to Hyperbolic Distance Error

For polar points $x = (r_x, \mathbf{u}_x)$ and $y = (r_y, \mathbf{u}_y)$, define

$$Z(x, y) = \cosh(\sqrt{c}\, r_x)\cosh(\sqrt{c}\, r_y) - \sinh(\sqrt{c}\, r_x)\sinh(\sqrt{c}\, r_y)\langle\mathbf{u}_x, \mathbf{u}_y\rangle. \tag{60}$$

Then $Z(x, y) = \cosh(\sqrt{c}\, d_c(x, y))$.

Lemma 15 (Angular error transfers through the hyperbolic law of cosines)**.** *Let $x, y \in \mathbb{H}^d_c$ have distance $d_{xy} = d_c(x, y) > 0$, and suppose RAP is defined on both points. If*

$$|\langle\widetilde{\mathbf{u}}_x, \widetilde{\mathbf{u}}_y\rangle - \langle\mathbf{u}_x, \mathbf{u}_y\rangle| \le \eta, \tag{61}$$

then

$$|Z(\Pi_{\mathrm{RAP}}x, \Pi_{\mathrm{RAP}}y) - Z(x, y)| \le \eta\, \sinh(\sqrt{c}\, r_x)\sinh(\sqrt{c}\, r_y). \tag{62}$$

If the right-hand side is at most $\frac{1}{2}\big(\cosh(\sqrt{c}\, d_{xy}) - 1\big)$, then

$$|d_c(\Pi_{\mathrm{RAP}}x, \Pi_{\mathrm{RAP}}y) - d_{xy}| \le \frac{2\eta\, \sinh(\sqrt{c}\, r_x)\sinh(\sqrt{c}\, r_y)}{\sqrt{c}\, \sinh(\sqrt{c}\, d_{xy})}. \tag{63}$$

Consequently,

$$\left|\frac{d_c(\Pi_{\mathrm{RAP}}x, \Pi_{\mathrm{RAP}}y)}{d_c(x, y)} - 1\right| \le \frac{2\eta\, \sinh(\sqrt{c}\, r_x)\sinh(\sqrt{c}\, r_y)}{\sqrt{c}\, d_{xy}\sinh(\sqrt{c}\, d_{xy})}. \tag{64}$$

Proof. RAP leaves r_x and r_y unchanged, so the only changed term in $Z(x, y)$ is the angular inner product. This gives the first bound.

Let $Z = Z(x, y)$ and $\widetilde{Z} = Z(\Pi_{\mathrm{RAP}}x, \Pi_{\mathrm{RAP}}y)$. By the mean value theorem,

$$\left|\operatorname{arcosh}(\widetilde{Z}) - \operatorname{arcosh}(Z)\right| \le |\widetilde{Z} - Z| \sup_{t \text{ between } Z \text{ and } \widetilde{Z}} \frac{1}{\sqrt{t-1}\sqrt{t+1}}. \tag{65}$$

The assumed small-perturbation condition gives $t - 1 \ge (Z - 1)/2$ and $t + 1 \ge (Z + 1)/2$ along the segment. Hence the derivative is at most $2/\sqrt{(Z-1)(Z+1)} = 2/\sinh(\sqrt{c}\, d_{xy})$. Multiplying by $1/\sqrt{c}$ converts arcosh error to hyperbolic distance error and proves the displayed inequalities. □

Proposition 16 (Pairs on the same radial ray are exact)**.** *If RAP is defined on $x = (r_x, \mathbf{u})$ and $y = (r_y, \mathbf{u})$, then*

$$d_c(\Pi_{\mathrm{RAP}}x, \Pi_{\mathrm{RAP}}y) = d_c(x, y) = |r_x - r_y|. \tag{66}$$

More generally, if $\langle \widetilde{\mathbf{u}}_x, \widetilde{\mathbf{u}}_y \rangle = \langle \mathbf{u}_x, \mathbf{u}_y \rangle$, then RAP preserves the distance between x and y exactly.

Proof. If $\mathbf{u}_x = \mathbf{u}_y = \mathbf{u}$ and $\mathbf{P}\mathbf{u} \neq \mathbf{0}$, then the normalized projected directions are equal: $\widetilde{\mathbf{u}}_x = \widetilde{\mathbf{u}}_y = \mathbf{P}\mathbf{u}/\|\mathbf{P}\mathbf{u}\|_2$. Therefore the angular inner product remains 1. The hyperbolic law of cosines gives

$$\cosh(\sqrt{c}\, d_c(x, y)) = \cosh(\sqrt{c}\, r_x)\cosh(\sqrt{c}\, r_y) - \sinh(\sqrt{c}\, r_x)\sinh(\sqrt{c}\, r_y) = \cosh(\sqrt{c}\, |r_x - r_y|), \tag{67}$$

and the same identity holds after RAP. The more general statement follows because the law-of-cosines expression is unchanged whenever the angular inner product is unchanged. □

J.4 A Finite-Set RAP Guarantee

The following theorem is not a universal hyperbolic JL theorem. Instead, it is the finite-set guarantee naturally suggested by the RAP mechanism: if radii are bounded, pairs are not arbitrarily close, and the angular directions satisfy a JL event, then RAP preserves all distances with an explicit distortion bound.

Theorem 17 (Finite-set RAP guarantee under bounded radius and separation)**.** *Let $X = \{x_1, \dots, x_n\} \subset \mathbb{H}^d_c$ have polar coordinates $x_i = (r_i, \mathbf{u}_i)$ around o. Assume $r_i \le R$ for all i and $d_c(x_i, x_j) \ge \tau > 0$ for all $i \neq j$. Let $\mathbf{P}$ be a Gaussian JL matrix and let $0 < \varepsilon < 1$. If*

$$4\varepsilon \sinh^2(\sqrt{c}\, R) \le \frac{1}{2}\big(\cosh(\sqrt{c}\, \tau) - 1\big), \tag{68}$$

and $k \ge C\varepsilon^{-2}\log(n/\delta)$, then with probability at least $1 - \delta$, RAP is defined on X, maps X into $\mathbb{H}^k_c$, preserves all anchor radii exactly, and satisfies

$$\max_{i<j} \left| \frac{d_c(\Pi_{\mathrm{RAP}}x_i, \Pi_{\mathrm{RAP}}x_j)}{d_c(x_i, x_j)} - 1 \right| \le \frac{8\varepsilon \sinh^2(\sqrt{c}\, R)}{\sqrt{c}\, \tau \sinh(\sqrt{c}\, \tau)}. \tag{69}$$

More precisely, each pair obeys the sharper bound

$$\left| \frac{d_c(\Pi_{\mathrm{RAP}}x_i, \Pi_{\mathrm{RAP}}x_j)}{d_c(x_i, x_j)} - 1 \right| \le \frac{8\varepsilon \sinh(\sqrt{c}\, r_i)\sinh(\sqrt{c}\, r_j)}{\sqrt{c}\, d_c(x_i, x_j)\sinh(\sqrt{c}\, d_c(x_i, x_j))}. \tag{70}$$

Proof. By Lemma 14, with probability at least $1 - \delta$,

$$|\langle \widetilde{\mathbf{u}}_i, \widetilde{\mathbf{u}}_j \rangle - \langle \mathbf{u}_i, \mathbf{u}_j \rangle| \le 4\varepsilon \tag{71}$$

for all pairs. The same event implies $\|\mathbf{P}\mathbf{u}_i\|_2^2 \ge 1 - \varepsilon > 0$, so RAP is defined on all non-anchor points; Corollary 13 gives the hyperboloid constraint and exact anchor-radius preservation.

For a fixed pair, Lemma 15 applies with $\eta = 4\varepsilon$. Since $r_i, r_j \le R$ and $d_c(x_i, x_j) \ge \tau$, the theorem's small-perturbation condition ensures

$$4\varepsilon \sinh(\sqrt{c}\, r_i)\sinh(\sqrt{c}\, r_j) \le 4\varepsilon \sinh^2(\sqrt{c}\, R) \le \frac{1}{2}\big(\cosh(\sqrt{c}\, d_c(x_i, x_j)) - 1\big). \tag{72}$$

Therefore Lemma 15 yields the sharper pairwise bound. Replacing r_i, r_j by R and $d_c(x_i, x_j)$ by τ gives the uniform maximum bound. □

Proof of Proposition 2. The statement that RAP lands in $\mathbb{H}_c^k$ and preserves $d_c(o, x) = r_x$ follows from Corollary 13. The exactness for pairs with the same angular direction follows from Proposition 16. The law-of-cosines perturbation bound is the first conclusion of Lemma 15. □

K Theoretical Comparison of the Three Schemes

This appendix collects the multiplicative distortion envelopes proved for the three constructions and arranges them under a common comparison. The underlying statements are Theorem 1 for the tangent scheme, Corollary 11 (Appendix I) for the fully hyperbolic scheme, and Theorem 17 (Appendix J) for RAP. All three use the same JL parameter ε and target dimension $k \geq C\varepsilon^{-2}\log(n/\delta)$.

Table 14: Multiplicative distortion envelope of the three projection schemes under their respective regularity conditions: tangent uses anchor $a \in \mathbb{H}_c^d$ with radius bound $R \leq R_0/\sqrt{c}$; the fully hyperbolic scheme and RAP use bounded radius $r_i \leq R$ and pairwise separation $d_c(x_i, x_j) \geq \tau$.

Scheme	Distortion envelope	Error decomposition
Tangent (Theorem 1)	$\varepsilon + C_1\, cR^2$	JL term ε shrinks with k; curvature term cR^2 persists and forces the local ball $R \leq R_0/\sqrt{c}$.
Fully hyperbolic (Corollary 11)	$\dfrac{12\,\varepsilon\,\sinh^2(\sqrt{c}\,R)}{\sqrt{c}\,\tau\,\sinh(\sqrt{c}\,\tau)}$	A radial contribution from the renormalization (Lemma 10) and a comparable angular contribution from the spatial inner-product sketch; both pass through the arcosh conditioning factor at small distances.
RAP (Theorem 17)	$\dfrac{8\,\varepsilon\,\sinh(\sqrt{c}\,r_x)\sinh(\sqrt{c}\,r_y)}{\sqrt{c}\,d_c(x,y)\sinh(\sqrt{c}\,d_c(x,y))}$	A single angular contribution; radii are preserved exactly and the radial source of error is removed by construction.

The envelopes for the fully hyperbolic scheme and RAP share the same denominator: both constructions read distance through arcosh and pay the conditioning factor $1/[\sqrt{c}\,d\sinh(\sqrt{c}\,d)]$ for nearby pairs. The numerators differ in structure. The fully hyperbolic numerator carries two contributions of comparable order, one from the renormalized time coordinate and one from the spatial inner-product sketch, and both are bounded only by the worst-case factor $\sinh^2(\sqrt{c}\,R)$. RAP eliminates the radial source by construction: the radius is preserved exactly, only the angular inner product is perturbed, and the per-pair numerator collapses to $\sinh(\sqrt{c}\,r_x)\sinh(\sqrt{c}\,r_y)$ with no cross-term. For a pair with one near-anchor endpoint, $\sinh(\sqrt{c}\,r)$ is small at that endpoint and the RAP numerator is much smaller than $\sinh^2(\sqrt{c}\,R)$, while the fully hyperbolic numerator is not.

The tangent envelope is multiplicatively cleanest inside its small ball but contains a non-vanishing cR^2 term, so it does not extend to large radii without an additional global argument. The envelopes for the fully hyperbolic scheme and RAP are stated for the finite-set regime with bounded radius and pairwise separation, mirroring the finite-set form of classical Euclidean JL. The empirical hierarchy in Section 5 tracks this theoretical ordering.