## A. Geodesics in the Hyperboloid Model

The hyperboloid model is $(\mathbb{H}^n, \langle\cdot,\cdot\rangle_1)$, where $\mathbb{H}^n := \{x \in \mathbb{R}^{n,1} : \langle x, x\rangle_1 = -1,\ x_0 > 0\}$. The hyperboloid model can be viewed from the extrinsically as embedded in the pseudo-Riemannian manifold Minkowski space $(\mathbb{R}^{n,1}, \langle\cdot,\cdot\rangle_1)$ and inducing its metric. The Minkowski metric tensor $g^{\mathbb{R}^{n,1}}$ of signature $(n, 1)$ has the components

$$g^{\mathbb{R}^{n,1}} = \begin{bmatrix} -1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

The associated inner-product is $\langle x, y\rangle_1 := -x_0 y_0 + \sum_{i=1}^n x_i y_i$. Note that the hyperboloid model is a Riemannian manifold because the quadratic form associated with $g^{\mathbb{H}}$ is positive definite.

In the extrinsic view, the tangent space at $\mathbb{H}^n$ can be described as $T_x\mathbb{H}^n = \{v \in \mathbb{R}^{n,1} : \langle v, x\rangle_1 = 0\}$. See Robbin & Salamon (2011); Parkkonen (2013).

Geodesics of $\mathbb{H}^n$ are given by the following theorem (Eq (6.4.10) in Robbin & Salamon (2011)):

**Theorem 6.** *Let $x \in \mathbb{H}^n$ and $v \in T_x\mathbb{H}^n$ such that $\langle v, v\rangle = 1$. The unique unit-speed geodesic $\phi_{x,v} : [0, 1] \to \mathbb{H}^n$ with $\phi_{x,v}(0) = x$ and $\dot\phi_{x,v}(0) = v$ is*

$$\phi_{x,v}(t) = x\cosh(t) + v\sinh(t). \tag{38}$$

## B. Proof of Theorem 1

*Proof.* From theorem 6, appendix A, we know the expression of the unit-speed geodesics of the hyperboloid model $\mathbb{H}^n$. We can use the Egregium theorem to project the geodesics of $\mathbb{H}^n$ to the geodesics of $\mathbb{D}^n$. We can do that because we know an isometry $\psi : \mathbb{D}^n \to \mathbb{H}^n$ between the two spaces:

$$\psi(x) := (\lambda_x - 1, \lambda_x x), \quad \psi^{-1}(x_0, x') = \frac{x'}{1 + x_0} \tag{39}$$

Formally, let $x \in \mathbb{D}^n, v \in T_x\mathbb{D}^n$ with $g^{\mathbb{D}}(v, v) = 1$. Also, let $\gamma : [0, 1] \to \mathbb{D}^n$ be the unique unit-speed geodesic in $\mathbb{D}^n$ with $\gamma(0) = x$ and $\dot\gamma(0) = v$. Then, by Egregium theorem, $\phi := \psi \circ \gamma$ is also a unit-speed geodesic in $\mathbb{H}^n$. From theorem 6, we have that $\phi(t) = x'\cosh(t) + v'\sinh(t)$, for some $x' \in \mathbb{H}^n, v' \in T_{x'}\mathbb{H}^n$. One derives their expression:

$$x' = \psi \circ \gamma(0) = (\lambda_x - 1, \lambda_x x) \tag{40}$$
$$v' = \dot\phi(0) = \frac{\partial\psi(y_0, y)}{\partial y}\bigg|_{\gamma(0)} \dot\gamma(0) = \begin{bmatrix} \lambda_x^2\langle x, v\rangle \\ \lambda_x^2\langle x, v\rangle x + \lambda_x v \end{bmatrix}$$

Inverting once again, $\gamma(t) = \psi^{-1} \circ \phi(t)$, one gets the closed-form expression for $\gamma$ stated in the theorem. □

One can sanity check that indeed the formula from theorem 1 satisfies the conditions:

- $d_{\mathbb{D}}(\gamma(0), \gamma(t)) = t, \quad \forall t \in [0, 1]$
- $\gamma(0) = x$
- $\dot\gamma(0) = v$
- $\lim_{t\to\infty} \gamma(t) := \gamma(\infty) \in \partial\mathbb{D}^n$

## C. Proof of Corollary 1.1

*Proof.* Denote $u = \frac{1}{\sqrt{g_x^{\mathbb{D}}(v,v)}} v$. Using the notations from Thm. 1, one has $\exp_x(v) = \gamma_{x,u}(\sqrt{g_x^{\mathbb{D}}(v, v)})$. Using Eq. 3 and 6, one derives the result. □

## D. Proof of Corollary 1.2

*Proof.* For any geodesic $\gamma_{x,v}(t)$, consider the plane spanned by the vectors $x$ and $v$. Then, from Thm. 1, this plane contains all the points of $\gamma_{x,v}(t)$, i.e.

$$\{\gamma_{x,v}(t) : t \in \mathbb{R}\} \subseteq \{ax + bv : a, b \in \mathbb{R}\} \tag{41}$$

□

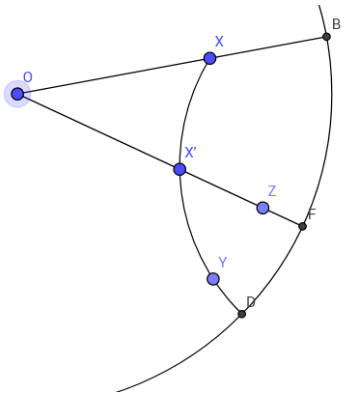
## E. Proof of Lemma 2

*Proof.* Assume the contrary and let $x \in \mathbb{D}^n \setminus \{0\}$ s.t. $\psi(\|x\|) > \frac{\pi}{2}$. We will show that transitivity implies that

$$\forall x' \in \partial\mathfrak{S}_x^{\psi(x)} : \quad \psi(\|x'\|) \le \frac{\pi}{2} \tag{42}$$

If the above is true, by moving $x'$ on any arbitrary (continuous) curve on the cone border $\partial\mathfrak{S}_x^{\psi(x)}$ that ends in $x$, one will get a contradiction due to the continuity of $\psi(\|\cdot\|)$.

We now prove the remaining fact, namely Eq. 42. Let any arbitrary $x' \in \partial\mathfrak{S}_x^{\psi(x)}$. Also, let $y \in \partial\mathfrak{S}_x^{\psi(x)}$ be any arbitrary point on the geodesic half-line connecting $x$ with $x'$ starting from $x'$ (i.e. excluding the segment from $x$ to $x'$). Moreover, let $z$ be any arbitrary point on the spoke through $x'$ radiating from $x'$, namely $z \in A_{x'}$ (notation from Eq. 15). Then, based on the properties of hyperbolic angles discussed before (based on Eq. 8), the angles $\angle yx'z$ and $\angle zx'x$ are well-defined.




From Cor. 1.2 we know that the points $O, x, x', y, z$ are coplanar. We denote this plane by $\mathcal{P}$. Furthermore, the metric of the Poincaré ball is conformal with the Euclidean metric. Given these two facts, we derive that

$$\angle yx'z + \angle zx'x = \angle(yx'x) = \pi \tag{43}$$

thus

$$\min(\angle yx'z, \angle zx'x) \leq \frac{\pi}{2} \tag{44}$$

It only remains to prove that

$$\angle yx'z \geq \psi(x') \quad \& \quad \angle zx'x \geq \psi(x') \tag{45}$$

Indeed, assume w.l.o.g. that $\angle yx'z < \psi(x')$. Since $\angle yx'z < \psi(x')$, there exists a point $t$ in the plane $\mathcal{P}$ such that

$$\angle Oxt < \angle Oxy \quad \& \quad \psi(x') \geq \angle tx'z > \angle yx'z \tag{46}$$

Then, clearly, $t \in \mathfrak{S}_{x'}^{\psi(x')}$, and also $t \notin \mathfrak{S}_{x}^{\psi(x)}$, which contradicts the transitivity property (Eq. 20). □

## F. Proof of Theorem 3

*Proof.* We first need to prove the following fact:

**Lemma 7.** *Transitivity implies that for all $x \in \mathbb{D}^n \setminus \{0\}$, $\forall x' \in \partial\mathfrak{S}_x^{\psi(x)}$:*

$$\sin(\psi(\|x'\|))\sinh(\|x'\|_{\mathbb{D}}) \leq \sin(\psi(\|x\|))\sinh(\|x\|_{\mathbb{D}}). \tag{47}$$

*Proof.* We will use the exact same figure and notations of points $y, z$ as in the proof of lemma 2. In addition, we assume w.l.o.g that

$$\angle yx'z \leq \frac{\pi}{2} \tag{48}$$

Further, let $b \in \partial\mathbb{D}^n$ be the intersection of the spoke through $x$ with the border of $\mathbb{D}^n$. Following the same argument as in the proof of lemma 2, one proves Eq. 45 which gives:

$$\angle yx'z \geq \psi(x') \tag{49}$$

In addition, the angle at $x'$ between the geodesics $xy$ and $Oz$ can be written in two ways:

$$\angle Ox'x = \angle yx'z \tag{50}$$

Since $x' \in \partial\mathfrak{S}_x^{\psi(x)}$, one proves

$$\angle Oxx' = \pi - \angle x'xb = \pi - \psi(x) \tag{51}$$

We apply hyperbolic law of sines (Eq. 10) in the hyperbolic triangle $Oxx'$:

$$\frac{\sin(\angle Oxx')}{\sinh(d_{\mathbb{D}}(O,x'))} = \frac{\sin(\angle Ox'x)}{\sinh(d_{\mathbb{D}}(O,x))} \tag{52}$$

Putting together Eqs. 48,49,50,51,52, and using the fact that $\sin(\cdot)$ is an increasing function on $[0, \frac{\pi}{2}]$, we derive the conclusion of this helper lemma. □

We now return to the proof of our theorem. Consider any arbitrary $r, r' \in (0,1) \cap \text{Dom}(\psi)$ with $r < r'$. Then, we claim that is enough to prove that

$$\exists x \in \mathbb{D}^n, x' \in \partial\mathfrak{S}_x^{\psi(x)} \quad s.t. \quad \|x\| = r, \|x'\| = r' \tag{53}$$

Indeed, if the above is true, then one can use the fact 5, i.e.

$$\sinh(\|x\|_{\mathbb{D}}) = \sinh\left(\ln\left(\frac{1+r}{1-r}\right)\right) = \frac{2r}{1-r^2} \tag{54}$$

and apply lemma 7 to derive

$$h(r') \leq h(r) \tag{55}$$

which is enough for proving the non-increasing property of function $h$.

We are only left to prove the fact 53. Let any arbitrary $x \in \mathbb{D}^n$ s.t. $\|x\| = r$. Also, consider any arbitrary geodesic $\gamma_{x,v} : \mathbb{R}_+ \to \partial\mathfrak{S}_x^{\psi(x)}$ that takes values on the cone border, i.e. $\angle(v, x) = \psi(x)$. We know that

$$\|\gamma_{x,v}(0)\| = \|x\| = r \tag{56}$$

and that this geodesic "ends" on the ball's border $\partial\mathbb{D}^n$, i.e.

$$\|\lim_{t\to\infty} \gamma_{x,v}(t)\| = 1 \tag{57}$$

Thus, because the function $\|\gamma_{x,v}(\cdot)\|$ is continuous, we obtain that for any $r' \in (r,1)$ there exists an $t' \in \mathbb{R}_+$ s.t. $\|\gamma_{x,v}(t')\| = r'$. By setting $x' := \gamma_{x,v}(t') \in \partial\mathfrak{S}_x^{\psi(x)}$ we obtain the desired result. □

## G. Proof of Theorem 5

*Proof.* For any $y \in \mathfrak{S}_x^{\psi(x)}$, the axial symmetry property implies that $\pi - \angle Oxy \leq \psi(x)$. Applying the hyperbolic cosine law in the triangle $Oxy$ and writing the above angle inequality in terms of the cosines of the two angles, one gets

$$\cos \angle Oxy = \frac{-\cosh(\|y\|_{\mathbb{D}}) + \cosh(\|x\|_{\mathbb{D}}) \cosh(d_{\mathbb{D}}(x,y))}{\sinh(\|x\|_{\mathbb{D}}) \sinh(d_{\mathbb{D}}(x,y))} \tag{58}$$

Eq. 28 is then derived from the above by an algebraic reformulation. □

## H. Training Details

For all methods except Order embeddings, we observe that initialization is very important. Being able to properly disentangle embeddings from different subparts of the graph in the initial learning stage is essential in order to train qualitative models. We conjecture that initialization is hard because these models are trained to minimize highly non-convex loss functions. In practice, we obtain our best results when initializing the embeddings corresponding to the hyperbolic cones using the Poincaré embeddings pre-trained for 100 epochs. The embeddings for the Euclidean cones are initialized using Simple Euclidean embeddings pre-trained also for 100 epochs. For the Simple Euclidean embeddings and Poincaré embeddings, we find the burn-in strategy of (Nickel & Kiela, 2017) to be essential for a good initial disentanglement. We also observe that the Poincaré embeddings are heavily collapsed to the unit ball border (as also pictured in Fig. 3) and so we rescale them by a factor of 0.7 before starting the training of the hyperbolic cones.

Each model is trained for 200 epochs after the initialization stage, except for order embeddings which were trained for 500 epochs. During training, 10 negative edges are generated per positive edge by randomly corrupting one of its end points. We use batch size of 10 for all models. For both cone models we use a margin of $\gamma = 0.01$.

All Euclidean models and baselines are trained using stochastic gradient descent. For the hyperbolic models, we do not find significant empirical improvements when using full Riemannian optimization instead of approximating it with a retraction map as done in (Nickel & Kiela, 2017). We thus use the retraction approximation since it is faster. For the cone models, we always project outside of the $\epsilon$ ball centered on the origin during learning as constrained by Eq. 26 and its Euclidean version. For both we use $\epsilon = 0.1$. A learning rate of 1e-4 is used for both Euclidean and hyperbolic cone models.