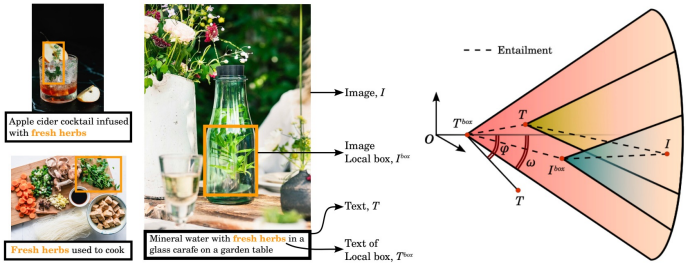
Hyperbolic Busemann Neural Networks

Ziheng Chen, Bernhard Schölkopf, Nicu Sebe

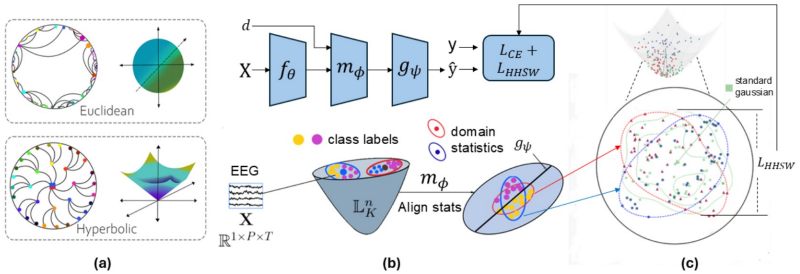
MOTIVATION

Applications of Hyperbolic Spaces

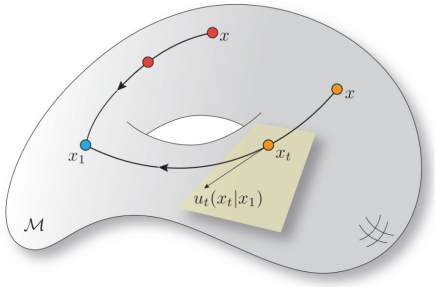
Vision-Language Models



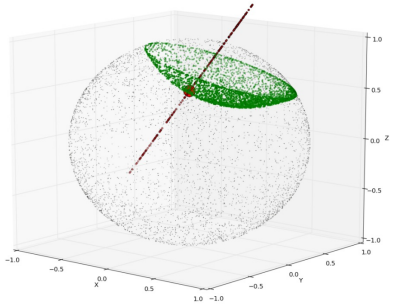
Brain-Computer Interfaces



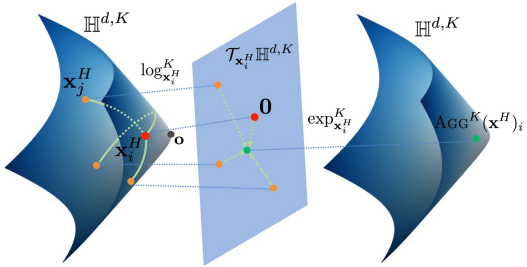
Generation



NLP



Graph



Busemann and Horosphere

Definition

$$\forall x \in \mathcal{H}_K^n, \quad B^\gamma(x) = \lim_{t\to\infty} \left(\mathrm{d}\left(x, \gamma(t)\right) - t\right).$$

$$H_\tau^v = \{x \in \mathcal{H}_K^n \mid B^v(x) = \tau\}$$

$$\forall x \in \mathbb{R}^n, \quad B^v(x) = -\langle x, v\rangle, \quad v \in \mathbb{S}^{n-1} = \{v \in \mathbb{R}^n \mid \|v\| = 1\}$$

Example

$$\mathbb{P}_K^n: \; B^v(x) = \frac{1}{\sqrt{-K}} \log\left(\frac{\left\|v - \sqrt{-K}x\right\|^2}{1 + K\|x\|^2}\right),$$

$$\mathbb{L}_K^n: \; B^v(x) = \frac{1}{\sqrt{-K}} \log\left(\sqrt{-K}\left(x_t - \langle x_s, v\rangle\right)\right).$$

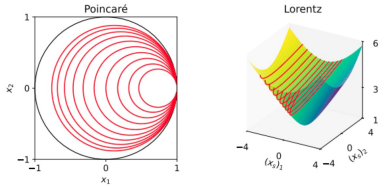


Figure 1. Illustration: red curves are different horospheres of B^v.

Generalization

Table 1. Generalization: Euclidean vs. hyperbolic.

Euclidean	Hyperbolic
Straight line	Geodesic
Parallel lines	Asymptotic geodesic rays
Inner product $\langle v, x\rangle$	Busemann function $-B^v(x)$
Hyperplane	Horosphere
Parallel hyperplanes with $v \in \mathbb{S}^{n-1}$	Horospheres with $v \in \mathbb{S}^{n-1} \cap T_e\mathcal{H}_K^n$

BUSEMANN MULTINOMIAL LOGISTIC REGRESSION

Euclidean MLR

$$\forall k,\ p(y=k \mid x) = \frac{\exp(\langle a_k, x\rangle + b_k)}{\sum_{j=1}^{C} \exp(\langle a_j, x\rangle + b_j)},$$

$$u_k(x) = \alpha_k \langle v_k, x\rangle + b_k.$$

$$v_k = \frac{a_k}{\|a_k\|} \in \mathbb{S}^{n-1} \quad \alpha_k = \|a_k\| > 0$$

$$\mathrm{d}\left(x, H_{a_k,b_k}\right) = \frac{|\langle a_k, x\rangle + b_k|}{\|a_k\|}$$

$$H_{a_k,b_k} = \{x \in \mathbb{R}^n : \langle a_k, x\rangle + b_k = 0\}$$

$$u_k(x) = \operatorname{sign}\left(\langle a_k, x\rangle + b_k\right) \|a_k\| \,\mathrm{d}\left(x, H_{a_k,b_k}\right),$$

BMLR

$$\forall k,\ p(y=k \mid x) = \frac{\exp(u_k(x))}{\sum_{j=1}^{C} \exp(u_j(x))},$$

$$u_k(x) = -\alpha_k B^{v_k}(x) + b_k,$$

with $\alpha_k > 0$, $v_k \in \mathbb{S}^{n-1}$, and $b_k \in \mathbb{R}$ as parameters.

$$\mathrm{d}\left(x, H_{v,\alpha,b}\right) = \frac{|-\alpha B^v(x) + b|}{\alpha}.$$

$$H_{v,\alpha,b} = \{x \in \mathcal{H}_K^n \mid -\alpha B^v(x) + b = 0\},$$

$$u_k(x) = \operatorname{sign}_k \alpha_k \,\mathrm{d}\left(x, H_{v_k,\alpha_k,b_k}\right)$$

$$\operatorname{sign}_k = \operatorname{sign}\left(-\alpha_k B^{v_k}(x) + b_k\right)$$

BUSEMANN FULLY CONNECTED LAYER

Euclidean FC

$$y = Ax + b \in \mathbb{R}^m$$

$$y_k = \langle a_k, x \rangle + b_k$$

Interpretation

Hyperbolic interpretation

$$\bar{\mathrm{d}}\left(y, H_{e_k,0}\right) = \langle a_k, x \rangle + b_k, \ \forall 1 \leq k \leq m,$$

$$\bar{\mathrm{d}}\left(y, H_{e_k,0}\right) = \operatorname{sign}\left(\langle e_k, y \rangle\right) \mathrm{d}\left(y, H_{e_k,0}\right)$$

Poincaré

$$y = \frac{\omega}{1 + \sqrt{1 - K \left\|\omega\right\|^2}}, \ \omega = \left[\frac{\sinh\left(\sqrt{-K} u_k(x)\right)}{\sqrt{-K}}\right]_{k=1}^{m},$$

Lorentz

$$y = \begin{bmatrix} y_t \\ y_s \end{bmatrix} = \begin{bmatrix} \sqrt{\frac{1}{-K} + \left\|y_s\right\|^2} \\ \frac{1}{\sqrt{-K}} \sinh\left(\sqrt{-K} u(x)\right) \end{bmatrix}$$

Generalization

$$\bar{\mathrm{d}}\left(y, H_{e_k,e}\right) = \phi\left(u_k(x)\right), \ \forall 1 \leq k \leq m.$$

$$\mathbb{L}_K^n \ni x \mapsto y = \begin{bmatrix} \sqrt{\frac{1}{-K} + \left\|y_s\right\|^2} \\ \frac{1}{\sqrt{-K}} \sinh\left(\sqrt{-K} u(x)\right) \end{bmatrix} \oplus_{\mathbb{L}} b \in \mathbb{L}_K^m,$$

COMPARISON

MLR

Real distance · Compact parameters · Batch-efficient

Method	Logit $u_k(x)$, $\forall k \in \{1, \dots, C\}$	Space	Dist	#Params	Compact params	FLOPs	Batch efficiency
Euclidean MLR	$\langle a_k, x \rangle + b_k$, with $a_k \in \mathbb{R}^n, b_k \in \mathbb{R}$	$\mathbb{R}^n$	Real	$C(n+1)$	✓	$C(2n)$	✓
Poincaré MLR [23, Eq. (25)]	$\frac{\lambda_{p_k}^K \|a_k\|}{\sqrt{-K}} \sinh^{-1}\left(\frac{2\sqrt{-K}\langle -p_k \oplus_M x, a_k\rangle}{\left(1 + K\|-p_k \oplus_M x\|^2\right)\|a_k\|}\right)$, with $p_k \in \mathbb{P}_K^n, a_k \in T_{p_k}\mathbb{P}_K^n$	$\mathbb{P}_K^n$	Real	$C(2n)$	✗	$C(19n+29)$	✗
Poincaré MLR [51, Eq. (6)]	$\frac{2}{\sqrt{-K}} \alpha_k \sinh^{-1}(\alpha - \beta)$, $\alpha = \lambda_x^K \sqrt{-K}\langle x, v_k\rangle \cosh\left(2\sqrt{-K}b_k\right)$, $\beta = \left(\lambda_x^K - 1\right)\sinh\left(2\sqrt{-K}b_k\right)$, with $\alpha_k > 0, v_k \in \mathbb{S}^{n-1}, b_k \in \mathbb{R}$	$\mathbb{P}_K^n$	Real	$C(n+2)$	✓	$C(4n+52)$	✓
Pseudo-Busemann MLR [45, Cor. 4.3]	$-d(x, p_k)\frac{B^{v_k}(-p_k \oplus_M x)}{\|-p_k \oplus_M x\|}$, with $p_k \in \mathbb{P}_K^n, v_k \in \mathbb{S}^{n-1}$	$\mathbb{P}_K^n$	Pseudo	$C(2n)$	✗	$C(19n+34)$	✗
Lorentz MLR [3, Eq. (12)]	$\frac{1}{\sqrt{-K}}\text{sign}(\alpha)\beta\left\|\sinh^{-1}\left(\sqrt{-K}\frac{\alpha}{\beta}\right)\right\|$, $\alpha = \cosh\left(\sqrt{-K}b_k\right)\langle z_k, x_s\rangle - \sinh\left(\sqrt{-K}b_k\right)$, $\beta = \sqrt{\left\|\cosh(\sqrt{-K}b_k)z_k\right\|^2 - \left(\sinh(\sqrt{-K}b_k)\|z_k\|\right)^2}$, with $z_k \in \mathbb{R}^n, b_k \in \mathbb{R}$	$\mathbb{L}_K^n$	Real	$C(n+1)$	✓	$C(4n+52)$	✓
BMLR	$-\alpha_k B^{v_k}(x) + b_k$, with $\alpha_k > 0, v_k \in \mathbb{S}^{n-1}, b_k \in \mathbb{R}$	$\mathbb{P}_K^n$ $\mathbb{L}_K^n$	Real	$C(n+2)$	✓	$\mathbb{P}_K^n : C(6n+12)$ $\mathbb{L}_K^n : C(2n+12)$	✓

FC

Busemann-based · Comparable complexity

Method	$\mathcal{F} : \mathcal{H}_K^n \ni x \mapsto y \in \mathcal{H}_K^m$	Space	Methodology	Parameters	#Params	FLOPs
Möbius [23, Eq. 27]	$\frac{1}{\sqrt{-K}}\tanh\left(\frac{\|Wx\|}{\|x\|}\tanh^{-1}\left(\sqrt{-K}\|x\|\right)\right)\frac{Wx}{\|Wx\|}$	$\mathbb{P}_K^n$	Tangent	$W \in \mathbb{R}^{m \times n}$	mn	$2nm + 2n$ $+2m + 24$
Poincaré FC [51, Eq. (7)]	$y = \frac{\omega}{1 + \sqrt{1 - K\|\omega\|^2}}$, $\omega_k = \frac{\sinh\left(\sqrt{-K}u_k(x)\right)}{\sqrt{-K}}$, with $u_k(x)$ in Tab. 2	$\mathbb{P}_K^n$	Poincaré geometry	$\alpha_k > 0,\ v_k \in \mathbb{S}^{n-1}$, $b_k \in \mathbb{R}$, for $k = 1, \dots, m$	$m(n+2)$	$4nm + 71m + 4$
Lorentz FC [11, Eq. (3)]	$y = \begin{bmatrix}\sqrt{\|\psi(Wx, v)\|^2 - 1/K} \\ \psi(Wx, v)\end{bmatrix}$, $\psi(Wx, v) = \lambda\sigma\left(v^\top x + b'\right)\frac{W\phi(x) + b}{\|W\phi(x) + b\|}$ with ϕ and σ as the activation and sigmoid	$\mathbb{L}_K^n$	Ambient Minkowski	$W \in \mathbb{R}^{m \times (n+1)}$, $v \in \mathbb{R}^{n+1}, b \in \mathbb{R}^m$, $b' \in \mathbb{R}, \lambda > 0$	$m(n+1) + m$ $+(n+1) + 2$	$2nm + 8m$ $+2n + 10$
BFC	$y = \frac{\omega}{1 + \sqrt{1 - K\|\omega\|^2}}$, $\omega = \frac{\sinh\left(\sqrt{-K}u(x)\right)}{\sqrt{-K}}$; $y_s = \frac{1}{\sqrt{-K}}\sinh\left(\sqrt{-K}u(x)\right)$, $y_t = \sqrt{\frac{1}{-K} + \|y_s\|^2}$, with $u_k(x) = \phi(-\alpha_k B^{v_k}(x) + b_k)$	$\mathbb{P}_K^n$ $\mathbb{L}_K^n$	Busemann	$\alpha_k > 0,\ v_k \in \mathbb{S}^{n-1}$, $b_k \in \mathbb{R}$, for $k = 1, \dots, m$	$m(n+2)$	$6nm + 29m + 4$ $2nm + 30m + 2$

EXPERIMENTS

Image

Image classification

Space	Method	CIFAR-10 (Num. classes: 10)			CIFAR-100 (Num. classes: 100)			Tiny-ImageNet (Num. classes: 200)			ImageNet-1k (Num. classes: 1000)		
		Acc	Fit Time	#Params	Acc	Fit Time	#Params	Acc	Fit Time	#Params	Acc	Fit Time	#Params
$\mathbb{R}^n$	MLR	95.14 ± 0.12	10.66	5.13K	77.72 ± 0.15	10.60	51.30K	65.19 ± 0.12	69.17	102.60K	71.87	2263.12	513K
$\mathbb{P}_K^n$	PMLR	95.04 ± 0.13	11.94	5.14K	77.19 ± 0.50	12.11	51.40K	64.93 ± 0.38	71.90	102.80K	71.77	2300.11	514K
	PBMLR-P	95.23 ± 0.08	**21.92**	**10.24K**	77.78 ± 0.15	**76.84**	**102.40K**	65.43 ± 0.27	**336.58**	**204.80K**	71.46	**3907.12**	**1024K**
	BMLR-P	**95.32 ± 0.14**	12.01	5.14K	**78.10 ± 0.35**	12.13	51.40K	**66.16 ± 0.19**	71.98	102.80K	**73.36**	2300.77	514K
$\mathbb{L}_K^n$	LMLR	94.98 ± 0.12	11.55	5.13K	78.03 ± 0.21	11.72	51.30K	65.63 ± 0.10	69.27	102.60K	72.46	2277.17	513K
	BMLR-L	**95.25 ± 0.02**	**11.08**	5.14K	**78.07 ± 0.26**	**11.22**	51.40K	**65.99 ± 0.14**	**69.19**	102.80K	**73.24**	**2276.53**	514K

Genome

Genome sequence learning

Benchmark	Task	Dataset	Num. classes	$\mathbb{P}_K^n$ PMLR	$\mathbb{P}_K^n$ PBMLR-P	$\mathbb{P}_K^n$ BMLR	$\mathbb{L}_K^n$ LMLR	$\mathbb{L}_K^n$ BMLR-L
TEB	Retrotransposons	LTR Copia	2	75.34 ± 1.02	74.37 ± 1.48	**76.73 ± 1.08**	73.01 ± 1.07	**75.86 ± 1.52**
		LINEs	2	85.54 ± 0.61	85.92 ± 0.65	**86.05 ± 1.08**	83.14 ± 0.80	**86.72 ± 0.58**
		SINEs	2	95.30 ± 0.85	95.34 ± 1.58	**95.99 ± 0.74**	**96.70 ± 0.87**	96.29 ± 0.59
	DNA transposons	CMC-EnSpm	2	83.39 ± 0.56	83.62 ± 1.00	**84.03 ± 0.71**	81.78 ± 1.05	**84.15 ± 1.00**
		hAT-Ac	2	89.38 ± 0.90	**89.86 ± 0.54**	89.62 ± 0.74	88.94 ± 0.69	**90.70 ± 0.51**
	Pseudogenes	processed	2	72.45 ± 1.49	71.99 ± 2.04	**73.09 ± 1.66**	**73.71 ± 1.76**	73.32 ± 1.65
		unprocessed	2	75.37 ± 2.27	71.99 ± 1.47	**75.71 ± 1.89**	74.54 ± 1.98	**76.15 ± 1.61**
GUE	Core Promoter Detection	tata	2	**80.95 ± 1.47**	79.32 ± 2.44	80.29 ± 1.63	80.90 ± 1.15	**81.76 ± 1.16**
		notata	2	70.02 ± 0.52	**70.60 ± 0.75**	70.48 ± 0.35	**71.26 ± 0.56**	70.43 ± 0.39
		all	2	67.64 ± 0.77	68.02 ± 0.63	**68.50 ± 0.61**	67.63 ± 0.56	**68.36 ± 1.07**
	Promoter Detection	tata	2	80.30 ± 1.59	80.27 ± 2.71	**82.83 ± 1.69**	**83.27 ± 1.95**	82.55 ± 1.54
		notata	2	92.63 ± 0.36	**93.05 ± 0.32**	92.75 ± 0.51	91.74 ± 0.57	**92.60 ± 0.49**
		all	2	90.53 ± 0.50	**90.79 ± 0.77**	90.20 ± 0.65	89.34 ± 0.40	**89.82 ± 0.45**
	Covid Variant Classification	Covid	9	**74.09 ± 0.25**	70.84 ± 0.80	73.40 ± 0.30	64.07 ± 0.51	**72.45 ± 0.21**
	Species Classification	Virus	20	67.24 ± 2.10	59.17 ± 3.32	**77.12 ± 1.23**	71.34 ± 2.05	**77.21 ± 1.04**
		Fungi	25	15.06 ± 1.32	18.75 ± 1.77	**30.01 ± 0.76**	15.07 ± 1.88	**30.14 ± 2.48**

EXPERIMENTS

Graph

Node classification

Space	Method	Disease $\delta = 0$	Airport $\delta = 1$	PubMed $\delta = 3.5$	Cora $\delta = 11$
$\mathbb{P}^n_K$	HGCN	86.87 ± 2.58	85.34 ± 1.16	76.29 ± 0.98	76.56 ± 0.81
	HGCN-PMLR	88.98 ± 1.96	84.78 ± 1.48	76.02 ± 1.09	77.47 ± 1.15
	HGCN-PBMLR-P	89.05 ± 0.78	85.04 ± 0.97	75.89 ± 0.78	77.90 ± 1.00
	HGCN-BMLR-P	**92.45 ± 0.96**	**86.02 ± 0.53**	**77.36 ± 0.73**	**78.48 ± 1.52**
$\mathbb{L}^n_K$	HGCN	87.83 ± 0.77	84.94 ± 1.40	76.49 ± 0.88	77.37 ± 1.72
	HGCN-LMLR	89.72 ± 1.51	82.61 ± 1.01	75.44 ± 1.17	69.91 ± 3.61
	HGCN-BMLR-L	**90.80 ± 1.15**	**85.27 ± 1.17**	**77.30 ± 0.41**	**77.65 ± 2.10**

Graph for FC

Link prediction

Table 7. Comparison of hyperbolic FC layers on link prediction. Best results within each hyperbolic model are in **bold**.

Space	Method	Methodology	Disease $\delta = 0$	Airport $\delta = 1$	PubMed $\delta = 3.5$	Cora $\delta = 11$
$\mathbb{P}^n_K$	Möbius	Tangent	76.35 ± 1.83	93.31 ± 0.41	**94.93 ± 0.06**	90.80 ± 0.56
	Poincaré FC	Poincaré geometry	79.45 ± 1.01	94.31 ± 0.16	94.24 ± 0.25	88.21 ± 0.72
	BFC-P	Busemann	**80.45 ± 0.93**	**94.88 ± 0.39**	94.85 ± 0.07	**91.94 ± 0.32**
$\mathbb{L}^n_K$	LTFC	Tangent	71.32 ± 5.36	92.68 ± 0.35	94.85 ± 0.17	89.37 ± 0.64
	Lorentz FC	Ambient Minkowski	72.78 ± 2.04	92.99 ± 0.33	94.20 ± 0.10	92.06 ± 0.62
	BFC-L	Busemann	**78.36 ± 0.51**	**95.37 ± 0.17**	**94.90 ± 0.04**	**92.28 ± 0.12**

Thanks

Homepage

LinkedIn

UNIVERSITÀ DI TRENTO

MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS