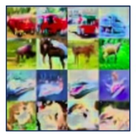
Neural Networks on Symmetric Spaces of Noncompact Type

Xuan Son Nguyen Shuo Yang Aymeric Histace

ETIS, UMR 8051, CY Cergy Paris University, ENSEA, CNRS, Cergy, France

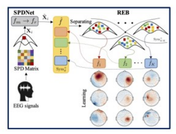
Applications of Riemannian Symmetric Spaces

Image generation



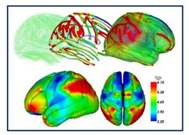
Huang et al., 2019

Brain-Computer Interfaces



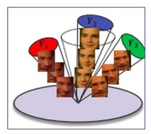
Suh and Kim, 2021

Medical imaging



Pennec, 2006

Face recognition



Huang et al., 2015

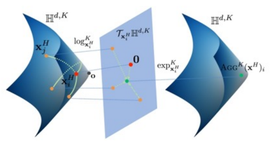
Robot learning

Calinon, 2020

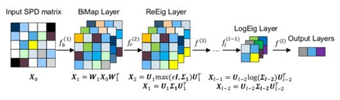
Some Applications of Symmetric Positive Definite (SPD) and Grassmann Manifolds

Problem And Related Works

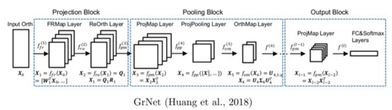
- Problem: Generalize Euclidean neural networks to the symmetric space setting.
- Existing approaches focus on building neural networks on hyperbolic spaces and matrix manifolds.
- Existing approaches either work for SPD manifolds associated with special families of Riemannian metrics or require rich algebraic structures of the considered spaces, which limits their generality
- It is not trivial for existing works to generalize many traditional machine learning models to symmetric spaces due to the lack of concepts in Euclidean spaces for the considered spaces, e.g., hyperplanes and the distance from a point to a hyperplane.



HGCN (Chami et al., 2019)



SPDNet (Huang et al., 2017)



GrNet (Huang et al., 2018)

Proposed Approach

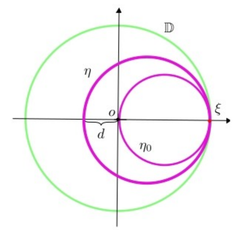
Contributions:

- Construct the point-to-hyperplane distance in a noncompact symmetric space.
- Derive an expression for the point-to-hyperplane distance in a noncompact symmetric space equipped with a G-invariant
- Riemannian metric.
- Fully-connected layers and an attention mechanism for neural networks on noncompact symmetric spaces.

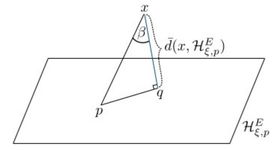
Let $\delta : [0, \infty) \to X$ be a (unit-speed) geodesic ray and $\xi = \delta(\infty) \in \partial X$. One defines the Busemann coordinate of a point $x \in X$ in the direction of ξ as

$$B_\xi(x) = \lim_{t \to \infty} (d(x, \delta(t)) - t).$$

The function $B_\xi : X \to \mathbb{R}$ is called the Busemann function associated to the geodesic ray δ.



The boundary $\partial\mathbb{D}$ of the Poincaré disk model $\mathbb{D}$ is illustrated by the green circle. The boundary point ξ is normal to both the horocycle η and the basic horocycle η_0. The distance between the origin o and the horocycle ξ is $-d$.



The distance between a point x and a hyperplane $\mathcal{H}^E_{\xi,p}$.

Definition (Hyperplanes on a Symmetric Space). *For $p \in X$ and $\xi \in \partial X$, hyperplanes on X are defined as*

$$\mathcal{H}_{\xi,p} = \{x \in X : B_\xi(\ominus p \oplus x) = 0\},$$

where $\ominus$ and $\oplus$ are the inverse and binary operations on X, respectively.

Definition (Point-to-Hyperplane Distance on a Symmetric Space). *Let $\mathcal{H}_{\xi,p}$ be a hyperplane, and let $\|\cdot\|_{\mathbb{S}}$ be a norm on X. Then the (signed) distance $\bar{d}(x, \mathcal{H}_{\xi,p})$ between a point $x \in X$ and $\mathcal{H}_{\xi,p}$ is defined as*

$$\bar{d}(x, \mathcal{H}_{\xi,p}) = d(x, p) \frac{B_\xi(\ominus p \oplus x)}{\| \ominus p \oplus x\|_{\mathbb{S}}}.$$

Corollary. *Let $\ominus$ and $\oplus$ be the Möbius subtraction $\ominus_M$ and Möbius addition $\oplus_M$ in $\mathbb{B}_m$, respectively, and let $\|\cdot\|_{\mathbb{S}}$ be the Euclidean norm $\|\cdot\|$. Let $p \in \mathbb{B}_m$, $\xi \in \partial\mathbb{B}_m$, and let $\mathcal{H}_{\xi,p}$ be a hyperplane. Then the distance $\bar{d}(x, \mathcal{H}_{\xi,p})$ between a point $x \in \mathbb{B}_m$ and $\mathcal{H}_{\xi,p}$ is computed by*

$$\bar{d}(x, \mathcal{H}_{\xi,p}) = -\frac{d_{\mathbb{B}}(x, p)}{\| -p \oplus_M x\|} \log \frac{1 - \| -p \oplus_M x\|^2}{\| -p \oplus_M x - \xi\|^2}.$$

Proposition. *Let $\phi : \mathrm{Sym}^+_m \to \mathrm{Sym}_m$ be a diffeomorphism. Let $\oplus$ and $\ominus$ be the binary and inverse operations defined by*

$$x \oplus y = \phi^{-1}(\phi(x) + \phi(y)),$$
$$\ominus x = \phi^{-1}(-\phi(x)),$$

where $x, y \in \mathrm{Sym}^+_m$. Let $\|\cdot\|_{\mathbb{S}}$ be the norm induced by the inner product $\langle\cdot,\cdot\rangle_{\mathbb{S}}$ given as

$$\langle x, y\rangle_{\mathbb{S}} = \langle \phi(x), \phi(y)\rangle.$$

Let $\delta(t) = \phi^{-1}(ta)$ be a geodesic line in Sym^+_m, where $a \in \mathrm{Sym}_m$ and $\|a\| = 1$. Let $\xi = \delta(\infty)$, $p \in \mathrm{Sym}^+_m$, and let $\mathcal{H}_{\xi,p}$ be a hyperplane. Then the distance $\bar{d}(x, \mathcal{H}_{\xi,p})$ between a point $x \in \mathrm{Sym}^+_m$ and $\mathcal{H}_{\xi,p}$ is computed as

$$\bar{d}(x, \mathcal{H}_{\xi,p}) = \langle a, \phi(p) - \phi(x)\rangle.$$

Corollary. *Let $\delta(t) = k \exp(ta)K$ be a geodesic ray, where $k \in K$, $a \in \mathfrak{a}$, $\|a\| = 1$, and let $\xi = \delta(\infty)$. Let $p = hK \in X$, $h \in G$, and let $\mathcal{H}_{\xi,p}$ be a hyperplane. Then the distance $\bar{d}(x, \mathcal{H}_{\xi,p})$ between a point $x = gK \in X$, $g \in G$ and $\mathcal{H}_{\xi,p}$ is computed as*

$$\bar{d}(x, \mathcal{H}_{\xi,p}) = \langle a, H(g^{-1}hk)\rangle.$$

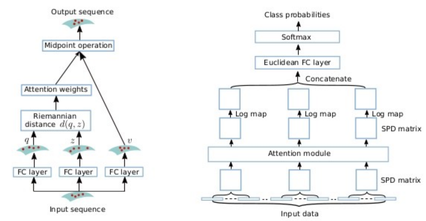
FC Layer

Proposition. *Let $\delta_j(t) = k_j \exp(ta_j)K, j = 1, \ldots, m$ be geodesic rays, where $k_j \in K$, $a_j \in \mathfrak{a}$, $\|a_j\| = 1$. Let $v_j(x) = B_{\xi_j}(\ominus p_j \oplus x), j = 1, \ldots, m$, where $\xi_j = \delta_j(\infty)$, $p_j \in X$, and $x \in X$ is the input of an FC layer. Then the output y of the FC layer can be expressed as*

$$y = n \exp([-v_1(x) \ldots - v_m(x)])K,$$

where $n \in N$.

Attention Mechanism



Our proposed attention block (a) and the network architecture for EEG classification (b).

Experimental Results

Image Classification

Mean accuracies and standard deviations of Hybrid ResNet-18 models for image classification.

Method	CIFAR-10	CIFAR-100
Hybrid Poincaré (Guo et al., 2022)	95.04±0.13	77.19±0.50
Poincaré ResNet (van Spengler et al., 2023)	94.51±0.15	76.60±0.32
Euclidean-Poincaré-H (Fan et al., 2023)	81.72±7.84	44.35±2.93
Euclidean-Poincaré-G (Ganea et al., 2018b)	95.14±0.11	**77.78**±0.09
Euclidean-Poincaré-B (Ours)	**95.23**±0.08	**77.78**±0.15

Image Generation

Reconstruction and generation FID of hyperbolic VAEs (lower is better).

Method	CIFAR-10		CIFAR-100	
	Rec. FID	Gen. FID	Rec. FID	Gen. FID
Lorentz-Poincaré-H (Fan et al., 2023)	125.53±5.94	69.11±1.67	110.36±11.50	62.32±6.34
Lorentz-Poincaré-G (Ganea et al., 2018b)	39.68±1.45	49.91±2.06	42.82±2.48	60.24±4.01
Lorentz-Poincaré-B (Ours)	**38.32**±2.11	**48.45**±1.31	**42.05**±2.58	**59.76**±1.81

Natural Language Inference

Mean accuracies and standard deviations of HypGRU models for natural language inference.

Method	SNLI	PREFIX-10%	PREFIX-30%	PREFIX-50%
HypGRU (Ganea et al., 2018b)	80.89±0.17	96.75±0.40	87.59±0.46	**76.45**±0.61
HypGRU-H (Fan et al., 2023)	80.66±0.46	92.20±8.33	83.29±6.34	74.67±3.12
HypGRU-B (Ours)	**81.01**±0.35	**97.03**±0.11	**87.69**±0.04	76.25±0.07