Fully-Connected Network on Noncompact Symmetric Space and Ridgelet Transform based on Helgason-Fourier Analysis

Sho Sonoda [1] Isao Ishikawa [1,2] Masahiro Ikeda [1]

[1]RIKEN Center for Advanced Intelligence Project (RIKEN AIP), Tokyo, Japan

[2]Ehime University, Ehime, Japan

The 39th International Conference on Machine Learning (ICML2022),
Baltimore, Maryland USA
July 17-23, 2022

Main Contributions

In this study, we have

- devised a *fully-connected network* $S[\gamma]$ *on noncompact symmetric spaces* $X = G/K$, which covers Hyperbolic NNs and SPD Nets as special cases;
- derived the *ridgelet transform* $R[f;\rho]$, a closed-form analysis operator satisfying the reconstruction formula $S[R[f;\rho]] = ((\sigma,\rho))f$, based on the *Helgason-Fourier transform on* G/K, and
- presented a constructive proof of the cc*-universality of finite networks on* G/K by discretizing the reconstruction formula in a *coordinate-free and unified manner*

A Noncompact Symmetric Space G/K

- is a homogeneous space G/K with *nonpositive sectional curvature* on which *G acts transitively*

Definition (Noncompact Symmetric Space G/K)

- Let G be a connected semisimple real Lie group, and
- let $G = KAN$ be the Iwasawa decomposition.
- The quotient (the set of all left cosets)

$$X := G/K = \{gK \mid g \in G\},$$

is called a noncompact symmetric space.

Example (Hyperbolic Space $\mathbb{H}^m$)

for embedding words, and tree-structured dataset

Example (SPD Manifold $\mathbb{P}_m$)

or a manifold of postive definite matrices, e.g., covariance matrices

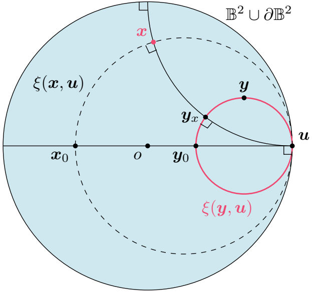


Figure: The Poincaré Disk $\mathbb{B}^2$ is a 2-dim. Hyperbolic space

Neural Networks on G/K

- are developing, but lack expressive power analysis
- Difficulty: There are no canonical ways to define an "affine map $\boldsymbol{a} \cdot \boldsymbol{x} - b$ on manifold"

(Reference) Euclidean NN

For any nonlinear function $\sigma : \mathbb{R} \to \mathbb{R}$, parameters $(\boldsymbol{a}_i, b_i, c_i) \in \mathbb{R}^m \times \mathbb{R} \times \mathbb{R}$, a Euclidean NN (a depth-2 fully-connected neural network on $\mathbb{R}^m$) is given by

$$f(\boldsymbol{x}) = \sum_{i=1}^{p} c_i \sigma(\boldsymbol{a}_i \cdot \boldsymbol{x} - b_i), \quad \boldsymbol{x} \in \mathbb{R}^m$$

Hyperbolic NNs (Ganea+18, Gulcehre+19, Shimizu+21)

- For each point $x \in \mathbb{H}^m$,
- the affine map is re-defined by Gyrovector calculus,
- the elementwise activation is defined on a tangent space: $\exp_0 \circ \sigma \circ \log_0(x)$

SPDNets (Huang-Gool17, Dong+17, Gao+19)

- For an SPD matrix $x \in \mathbb{P}_m$,
- BiMap layer: $w^\top x w$
- ReEig layer: $u^\top \max(0, \lambda - b) u$ where $x = u^\top \lambda u$

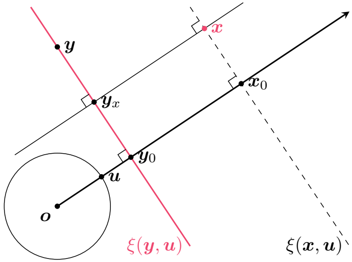
Geometric Reformulation of Euclidean Fully-Connected Neuron $\sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b)$

As a "*wavelet function* on the *signed distance* between a *point* $\boldsymbol{x}$ and a *hyperplane* ξ":

$$\sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b) = \sigma(r d(\boldsymbol{x}, \xi)), \quad \text{(geometric, or coordinate-free)}$$

where

- $\boldsymbol{a} = r\boldsymbol{u}$ (polar coordinates $(r, \boldsymbol{u}) \in \mathbb{R} \times \mathbb{S}^{m-1}$)
- $\xi := \{\boldsymbol{y} \in \mathbb{R}^m \mid r\boldsymbol{u} \cdot \boldsymbol{y} = b\}$ (a hyperplane passing through point $(b/r)\boldsymbol{u}$ with normal $\boldsymbol{u}$)
- $d(\boldsymbol{x}, \xi)$ signed distance from point $\boldsymbol{x}$ to hyperplane ξ
- $d \mapsto \sigma(rd)$ wavelet function



Main Results 1/3: A Continuous Neural Network on G/K

Definition (Continuous Horospherical X-NN)

$$S[\gamma](x) := \int_{\mathfrak{a}^* \times \partial X \times \mathbb{R}} \gamma(a,u,b)\sigma(a \cdot \langle x,u\rangle - b)e^{\varrho\langle x,u\rangle} \mathrm{d}a\mathrm{d}u\mathrm{d}b, \;\; x \in X$$

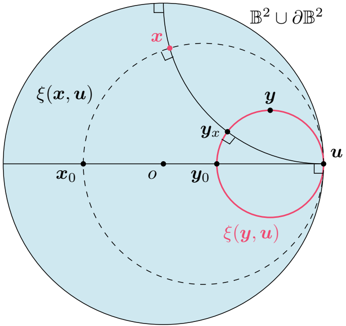
- i.e., a continuous frame defined by weighted fully-connected neurons $x \mapsto \sigma(\cdots)$
- $\mathfrak{a} = \mathfrak{a}^* = \mathbb{R}^r$ for $r := \operatorname{rank} X$ (Lie algebra and its dual of A of $G = KAN$)
- $\varrho \in \mathfrak{a}^*$ constant vector

$\mathfrak{a}$-valued composite distance $\langle x,u\rangle$

is a vector distance from the origin o to a horosphere $\xi(x,u)$ (i.e., $|\langle x,u\rangle|$ becomes the Riemannian distance)

horosphere $\xi(x,u)$ passing through point $x \in X$ normal $u \in \partial X$

is a sphere in G/K with infinite radius (Recall that a Euclidean hyperplane is a *Euclidean sphere with infinite radius*)



Main Results 2/3: Ridgelet Transform

- is an analysis operator (or a pseudo-inverse operator) of integral representation operator S

Definition (Ridgelet Transform)

For any function $f : G/K \to \mathbb{C}$ and an auxiliary function $\rho : \mathbb{R} \to \mathbb{R}$,

$$R[f](a,u,b) := \int_X \boldsymbol{c}[f](x)\overline{\rho(a \cdot \langle x,u\rangle - b)}e^{\varrho\langle x,u\rangle}\mathrm{d}x$$

where $\boldsymbol{c}[f]$ is a Helgason-Fourier multiplier.

Theorem (Reconstruction Formula)

For any $f \in L^2(X)$ *(or* $f \in C_c^\infty(X)$*)*, $\sigma \in \mathcal{S}'(\mathbb{R}), \rho \in \mathcal{S}(\mathbb{R})$,

$$S[R[f;\rho]] = ((\sigma,\rho))f.$$

where $((\sigma,\rho))$ *is a scalar product.*

- i.e., a *constructive* universal approximation theorem for *continuous* neural networks on G/K

Main Results 3/3: cc-Universality

- is a *constructive* universal approximation theorem for *finite* neural networks on G/K

Theorem (cc-Universality of X-NNs)

For any $\varepsilon > 0$, *compact set* $Z \subset X$, *functions* $f \in C(Z), \sigma \in \mathcal{S}'(\mathbb{R}), \rho \in \mathcal{S}(\mathbb{R})$, *there exists a finite neural network of the form*

$$f(x) := \sum_{i=1}^{n} c_i \sigma(a_i \cdot \langle x, u_i \rangle - b_i) e^{\varrho \langle x, u_i \rangle}, \quad x \in X = G/K$$

satisfying

$$\sup_{x \in Z} |f_n(x) - f(x)| < \varepsilon.$$

- *i.e., the density of finite NNs in continuous functions w.r.t. the compact convergence*

Examples: Horospherical HNN and SPD Net

Continuous Horospherical Hyperbolic NN

On the *Poincaré ball model* $\mathbb{B}^m := \{\boldsymbol{x} \in \mathbb{R}^m \mid |\boldsymbol{x}| < 1\}$ equipped with the Riemannian metric $\mathfrak{g} = 4(1 - |\boldsymbol{x}|)^{-2} \sum_{i=1}^{m} \mathrm{d}x_i \otimes \mathrm{d}x_i$,

$$S[\gamma](\boldsymbol{x}) := \int_{\mathbb{R} \times \partial\mathbb{B}^m \times \mathbb{R}} \gamma(a, \boldsymbol{u}, b)\sigma(a\langle \boldsymbol{x}, \boldsymbol{u}\rangle - b)e^{\varrho\langle \boldsymbol{x}, \boldsymbol{u}\rangle} \mathrm{d}a\mathrm{d}\boldsymbol{u}\mathrm{d}b, \quad \boldsymbol{x} \in \mathbb{B}^m$$

$$\varrho = (m-1)/2,\ \langle \boldsymbol{x}, \boldsymbol{u}\rangle = \log\left(\frac{1 - |\boldsymbol{x}|_E^2}{|\boldsymbol{x} - \boldsymbol{u}|_E^2}\right), \quad (\boldsymbol{x}, \boldsymbol{u}) \in \mathbb{B}^m \times \partial\mathbb{B}^m$$

Continuous Horospherical SPD Net

$$S[\gamma](x) := \int_{\mathbb{R}^m \times \partial\mathbb{P}_m \times \mathbb{R}} \gamma(\boldsymbol{a}, u, b)\sigma(\boldsymbol{a} \cdot \langle x, u\rangle - b)e^{\boldsymbol{\varrho}\cdot\langle x, u\rangle} \mathrm{d}\boldsymbol{a}\mathrm{d}u\mathrm{d}b, \quad \boldsymbol{x} \in \mathbb{P}_m$$

$\boldsymbol{\varrho} = (-\frac{1}{2}, \cdots, -\frac{1}{2}, \frac{m-1}{4})$, $\langle \boldsymbol{x}, \boldsymbol{u}\rangle = \dfrac{1}{2}\log\lambda\left(uxu^\top\right), \quad (x, u) \in \mathbb{P}_m \times \partial\mathbb{P}_m$ where λ is the diagonal in the *Cholesky decomposition*

Sketch Proof of Reconstruction Formula

- is based on *Helgason-Fourier analysis* on G/K
- Given a function $f : G/K \to \mathbb{C}$, solve an integral equation $S[\gamma] = f$ of unknown γ.
- Step 1: Turn $S[\gamma]$ to a *Helgason-Fourier expression*:

$$S[\gamma](x) := \int_{\mathfrak{a}^* \times \partial X \times \mathbb{R}} \gamma(a,u,b)\sigma(a \cdot \langle x,u \rangle - b) e^{\varrho \langle x,u \rangle} \mathrm{d}a\mathrm{d}u\mathrm{d}b$$

$$= \frac{1}{2\pi} \int_{\mathbb{R}} \left[\int_{\mathfrak{a}^* \times \partial X} \gamma^\sharp(\lambda/\omega, u, \omega) |\boldsymbol{c}(\lambda)|^2 e^{(i\lambda + \varrho)\langle x,u \rangle} \frac{\mathrm{d}\lambda \mathrm{d}u}{|\boldsymbol{c}(\lambda)|^2} \right] |\omega|^{-r} \sigma^\sharp(\omega) \mathrm{d}\omega.$$

- Step 2: Since inside $[\cdots]$ is the *inverse Helgason-Fourier transform*, put a separation-of-variables form:

$$\gamma^\sharp_{f,\rho}(\lambda/\omega, \boldsymbol{u}, \omega) = \widehat{f}(\lambda, \boldsymbol{u}) \overline{\rho^\sharp(\omega)} |\boldsymbol{c}(\lambda)|^{-2}.$$

  Then, by the construction, it is a particular solution:

$$S[\gamma_{f,\rho}] = ((\sigma, \rho)) f,$$

  where $((\sigma,\rho)) := \frac{|W|}{2\pi} \int_{\mathbb{R}} \sigma^\sharp(\omega) \overline{\rho^\sharp(\omega)} |\omega|^{-m} \mathrm{d}\omega$.
- In the end, we employ $\gamma_{f,\rho}$ as the definition of $R[f;\rho]$