Deep Ridgelet Transform: Harmonic Analysis for Deep Neural Networks

Sho Sonoda

Senior Research Scientist
Deep Learning Theory team
RIKEN AIP

CMAI: Interface of Math and Artificial Intelligence Seminar
at The Chinese University of Hong Kong (CUHK), online
July 3, 2024

Self-Introduction—Sho Sonoda

Brief Bio:

Apr 2018–Present: Postdoc Researcher → Senior Research Scientist, RIKEN AIP, Japan

Apr 2017–Mar 2018: Research Associate, Waseda Univ.

Apr 2015–Mar 2017: JSPS Research Fellowships (DC2)

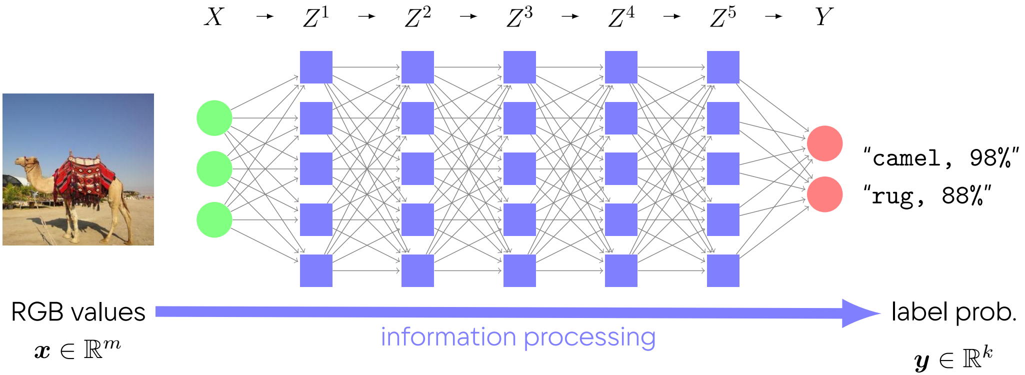
Sep 2013–Mar 2017: Doctor of Engineering, Waseda Univ., Japan

Apr 2012–Sep 2013: Software Engineer, Panasonic Corp.

Backgrounds:

- **Neural Network Theory** based on
  - harmonic analysis, differential equations
- Machine Learning Applications for...
  - theorem proving, quantum machine learning, autonomous driving, EEG analysis, steel making

Q. What is a typical solution obtained by deep learning?



- Want to identify what solution is typically acquired via deep learning
- Want to know why (and when) deep learning performs better (than shallow networks)

Problem Formulation

Problem (NN Learning Equation)

Given a data generating function $f : X \to \mathbb{R}$, *find a parameter* γ *of DNN satisfying*

$$\mathtt{DNN}[\gamma] = f.$$

We call a solution operator $R = \mathtt{DNN}^\dagger$ *the ridgelet transform, which satisfies*

$$\mathtt{DNN}[R[f]] = f.$$

Today's setting:

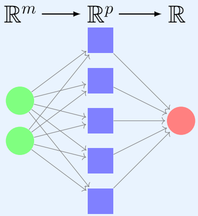
- Depth-2 fully-connected NN on Euclidean space $X = \mathbb{R}^m$

  and
- Depth-n fully-connected NN on Euclidean space $X = \mathbb{R}^m$

Shallow neural network (SNN)

Definition (single-hidden-layer fully-connected neural network)

For any function $\sigma : \mathbb{R} \to \mathbb{R}$,



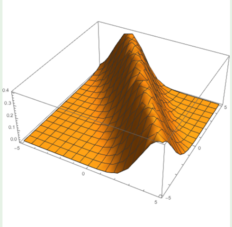
$$\mathtt{SNN}(\boldsymbol{x};\boldsymbol{\theta}) := \sum_{i=1}^{p} c_i \sigma(\boldsymbol{a}_i \cdot \boldsymbol{x} - b_i), \quad \boldsymbol{x} \in \mathbb{R}^m$$

Here $\boldsymbol{\theta} = \{(\boldsymbol{a}_i, b_i, c_i)\}_{i=1}^{p} \subset \mathbb{R}^m \times \mathbb{R} \times \mathbb{R}$ is the parameters

Visual Example: single Gaussian hidden unit

- input dimension $m = 2$
- hidden layer width $p = 1$
- activation function $\sigma(t) = \exp(-t^2/2)$

$$\mathtt{SNN}(\boldsymbol{x};\boldsymbol{\theta}) = c\sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b), \quad \boldsymbol{x} \in \mathbb{R}^2$$



Visual understanding of how SNN approximates functions

Theorem (Universal Approximation Property, or Universality)

A SNN can approximate any continuous function

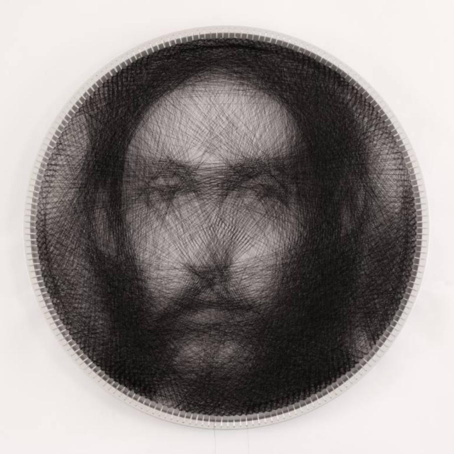


Figure: Petros Vrellis, Knit#1, 2016

Recap:

$$\mathtt{SNN}(\boldsymbol{x};\boldsymbol{\theta}) = \sum_{i=1}^{p} c_i \sigma(\boldsymbol{a}_i \cdot \boldsymbol{x} - b_i)$$

$\boldsymbol{a}$: width and direction of a ridge

b: location of a ridge

c: height of a ridge

drawing method by lines

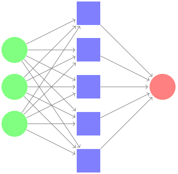
- Fourier series expansion, X-ray tomography

drawing method by points (pixels/dots)

- pixel art, computer screen, kernel regression

Integral Representation of Neural Network $S[\gamma]$

Finite-width (Discrete, or "Ordinary") NN



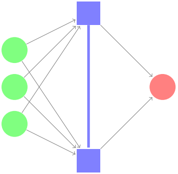
- $\mathtt{SNN}(\boldsymbol{x};\theta_p) = \sum_{i=1}^{p} c_i \sigma(\boldsymbol{a}_i \cdot \boldsymbol{x} - b_i)$
- *nonlinear* parameters: $\theta_p = \{(\boldsymbol{a}_i, b_i, c_i)\}_{i=1}^p \in \mathbb{R}^{(m+2)p}$

continuum limit

discretization

$\gamma_p = \sum_{i=1}^p c_i \delta_{(\boldsymbol{a}_i, b_i)}$

Infinite-width (Continuous, or Integral Representation of) NN



- $S[\gamma](\boldsymbol{x}) = \int_{\mathbb{R}^m \times \mathbb{R}} \gamma(\boldsymbol{a}, b) \sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b) \mathrm{d}\boldsymbol{a} \mathrm{d}b$
- *linear* parameter: $\gamma \in \mathrm{Map}(\mathbb{R}^m \times \mathbb{R} \to \mathbb{C})$

Ridgelet Transform $R[f;\rho]$

Definition (Ridgelet Transform)

For any function $f:\mathbb{R}^m \to \mathbb{C}$ and $\rho:\mathbb{R}\to\mathbb{C}$, put

$$R[f;\rho](\boldsymbol{a},b) = \int_{\mathbb{R}^m} f(\boldsymbol{x})\overline{\rho(\boldsymbol{a}\cdot\boldsymbol{x}-b)}\mathrm{d}\boldsymbol{x}, \quad (\boldsymbol{a},b)\in\mathbb{R}^m\times\mathbb{R}.$$

Theorem (Reconstruction Formula)

For any $\sigma\in\mathcal{S}'(\mathbb{R})$, $\rho\in\mathcal{S}(\mathbb{R})$ and $f\in L^2(\mathbb{R}^m)$, we have

$$S[R[f;\rho]](\boldsymbol{x}) = \int_{\mathbb{R}^m\times\mathbb{R}} R[f;\rho](\boldsymbol{a},b)\sigma(\boldsymbol{a}\cdot\boldsymbol{x}-b)\mathrm{d}\boldsymbol{a}\mathrm{d}b = (\!(\sigma,\rho)\!) f(\boldsymbol{x}), \quad \boldsymbol{x}\in\mathbb{R}^m$$

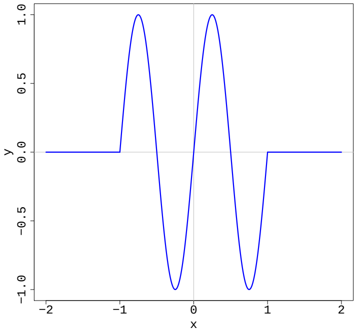
where $(\!(\sigma,\rho)\!) = (2\pi)^{m-1}\int_{\mathbb{R}} \sigma^\sharp(\omega)\overline{\rho^\sharp(\omega)}|\omega|^{-m}\mathrm{d}\omega$ and $\sharp$ denotes the Fourier transform

- Meaning 1: Continuous NN is a universal approximator
- Meaning 2: R and S play the same role as Fourier F and inverse Fourier F^{-1} transforms:

$$F^{-1}[F[f]](\boldsymbol{x}) = (2\pi)^{-m}\int_{\mathbb{R}^m} F[f](\boldsymbol{\xi})e^{i\boldsymbol{x}\cdot\boldsymbol{\xi}}\mathrm{d}\boldsymbol{\xi} = f(\boldsymbol{x})$$

Numerical Example of Ridgelet Transform $R[f;\rho](a,b)$

- $f(x) = \sin(2\pi x)\,\mathbf{1}_{[-1,1]}(x)$
- $R[f;\rho](a,b) = \int_{\mathbb{R}} f(x)\rho(ax-b)\mathrm{d}x \approx \sum_i \sin(2\pi x_i)\rho(ax_i - b)\Delta x$
- $\rho_2(b) := H[\rho_0^{(2)}](b)$ with $\rho_0(b) := \exp(-b^2/2)$
  where H is Hilbert transform, $H[\rho_0]$ is the Dawson function



data $f(x)$

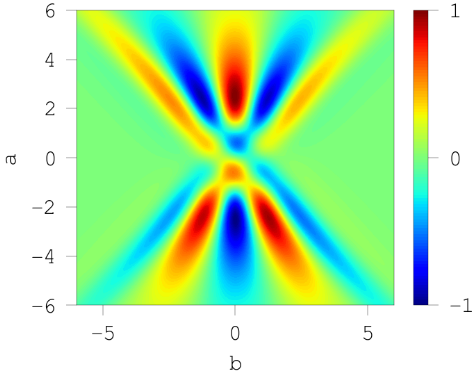
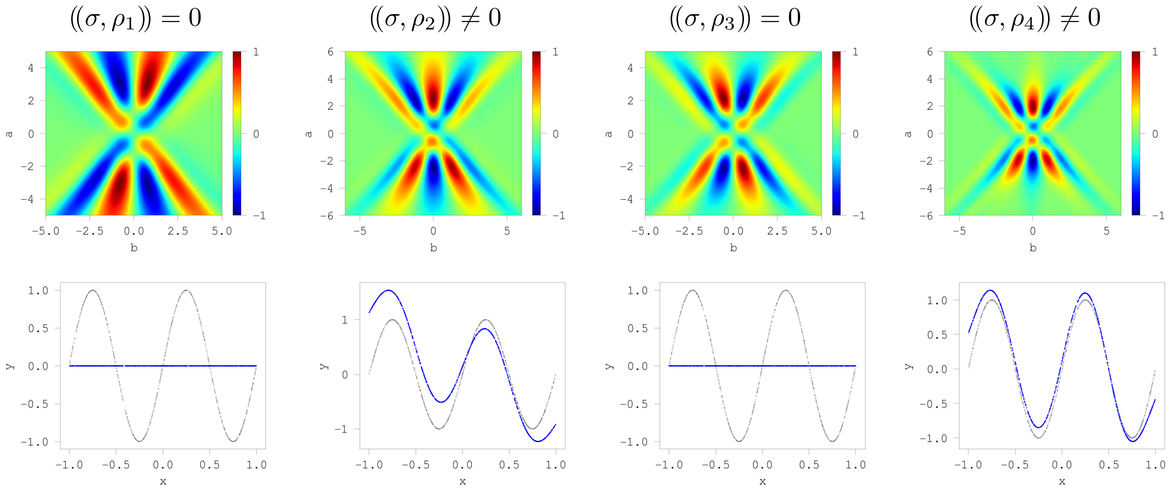


image $R[f;\rho_2](a,b)$

Visualization Results of Reconstruction Formula $S[R[f;\rho]] = ((\sigma,\rho))f$

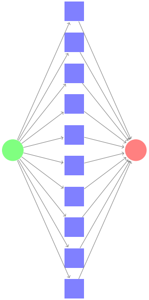
- $\rho_k := H\rho_0^{(k)}$, $\sigma(b) = \tanh(b)$
- $S[R[f;\rho]] = ((\sigma,\rho))f \equiv 0$ (degenerate) when $((\sigma,\rho_k)) = 0$

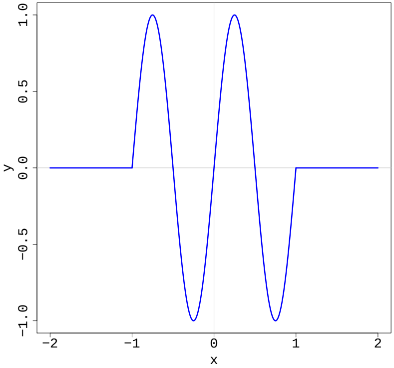


How the parameter distribution looks like?

We will train many ($n = 1,000$) neural networks $\mathtt{SNN}(x; \boldsymbol{\theta}) = \sum_{j=1}^{p} c_j \sigma(a_j \cdot x - b_j)$ with $p = 10$ hidden units, and see the distribution of trained parameters (a_j, b_j, c_j).

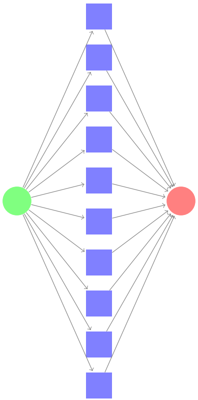
- $f(x) = \sin(2\pi x)\, \mathbf{1}_{[-1,1]}(x)$
- $\sigma(z) = \tanh(z)$
- $\widehat{L}(\boldsymbol{\theta})$ is square error loss
- SGD with weight decay

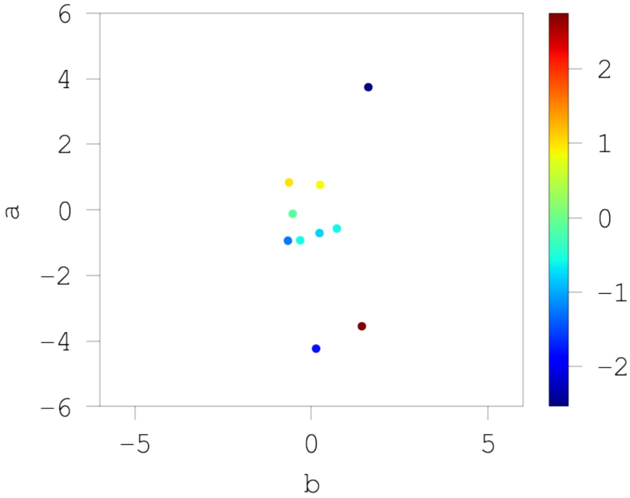




data $f(x)$

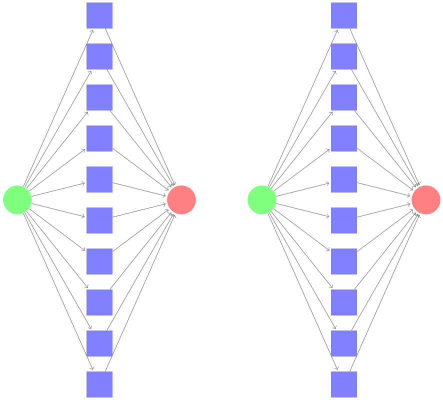
A scatter plot of $d \times n = 10$ hidden parameters (a_j, b_j, c_j) obtained from $n = 1$ neural network $\sum_{j=1}^{d} c_j \sigma(a_j \cdot x - b_j)$ with $p = 10$ hidden units.

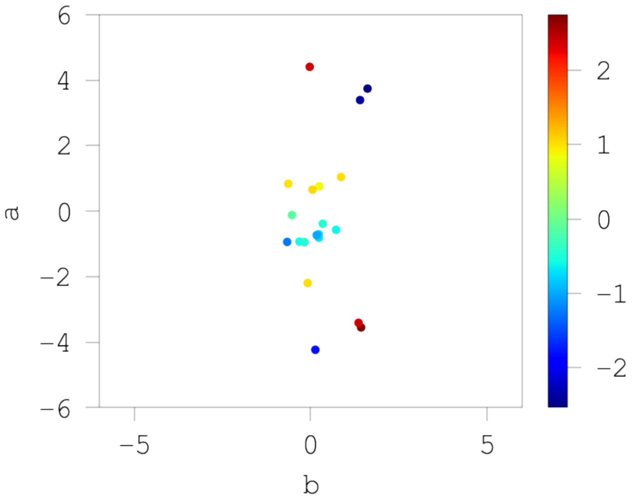




$\arg\min \widehat{L}(\boldsymbol{\theta})$

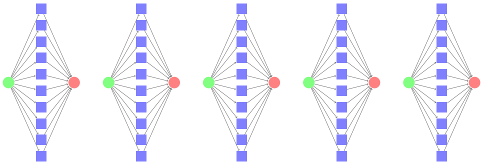
A scatter plot of $d \times n = 20$ hidden parameters (a_j, b_j, c_j) obtained from $n = 2$ neural networks with $p = 10$ hidden units.

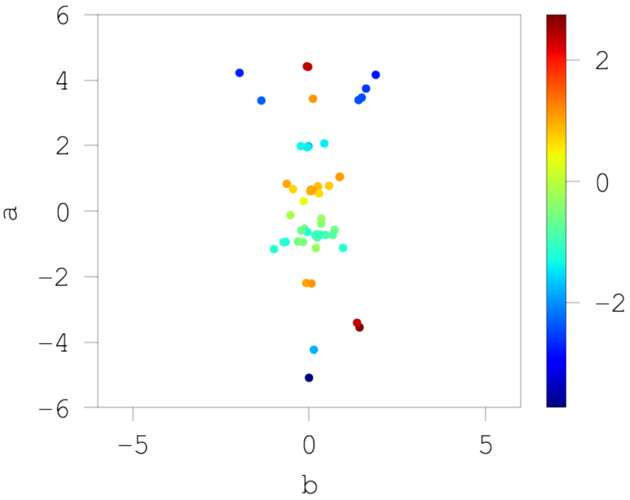




$\arg\min \widehat{L}(\boldsymbol{\theta})$

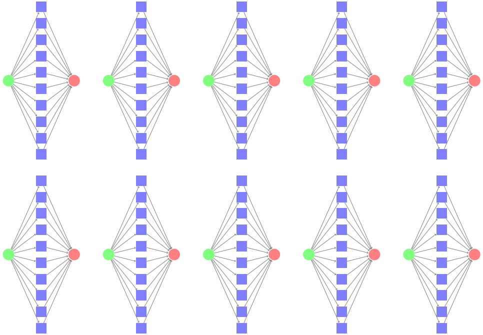
A scatter plot of $d \times n = 50$ hidden parameters (a_j, b_j, c_j) obtained from $n = 5$ neural networks with $p = 10$ hidden units.

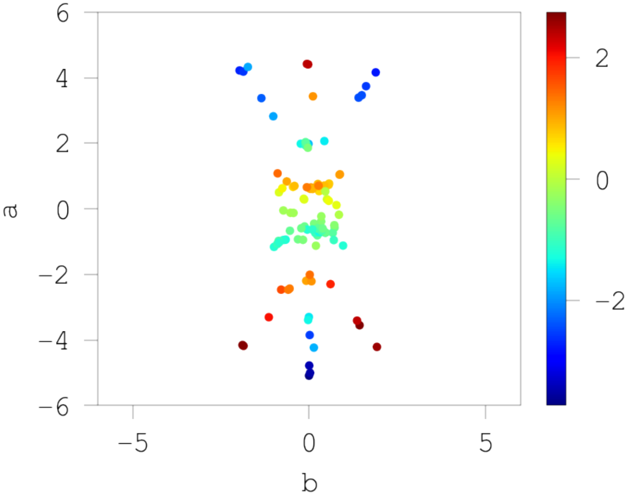




$\arg\min \widehat{L}(\boldsymbol{\theta})$

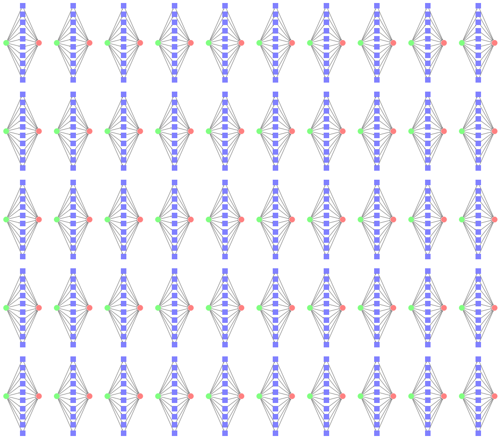
A scatter plot of $d \times n = 100$ hidden parameters (a_j, b_j, c_j) obtained from $n = 10$ neural networks with $p = 10$ hidden units.

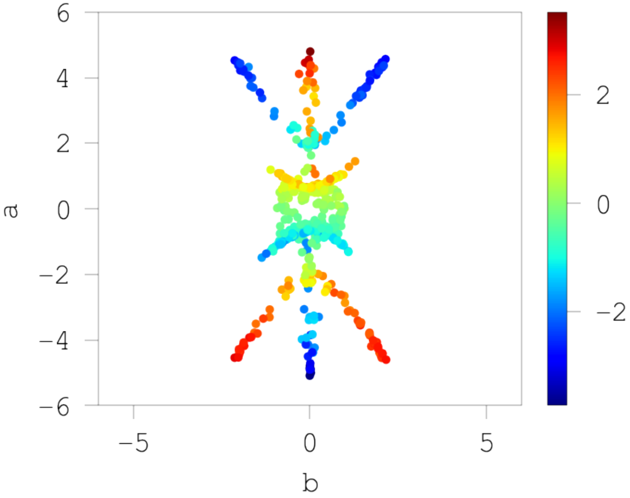




$\arg\min \widehat{L}(\boldsymbol{\theta})$

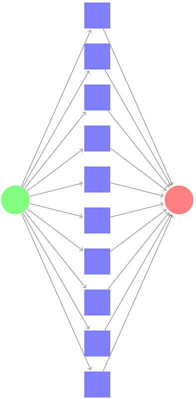
A scatter plot of $d \times n = 500$ hidden parameters (a_j, b_j, c_j) obtained from $n = 50$ neural networks with $p = 10$ hidden units.



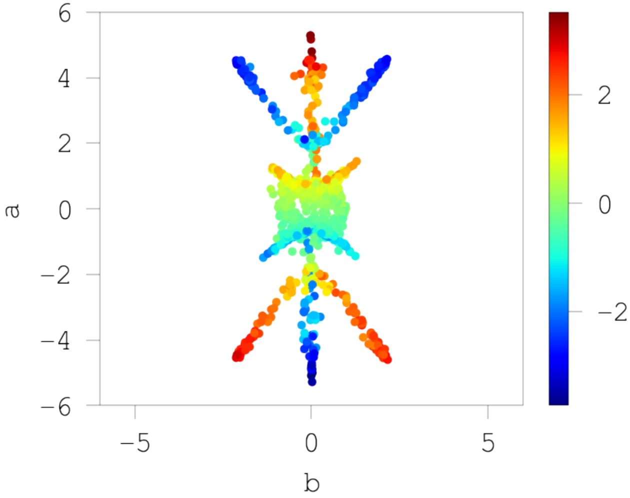


$\arg\min \widehat{L}(\boldsymbol{\theta})$

A scatter plot of $d \times n = 1,000$ hidden parameters (a_j, b_j, c_j) obtained from $n = 100$ neural networks with $p = 10$ hidden units.

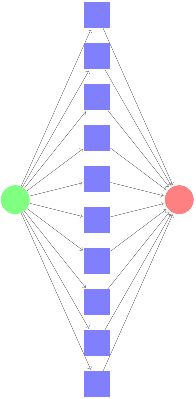


$\times 100$

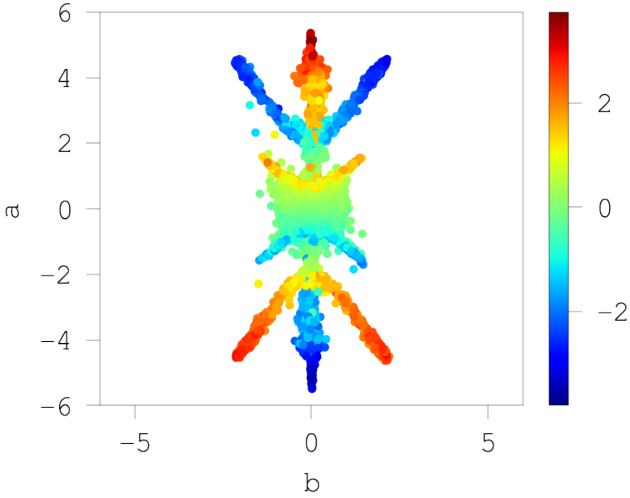


$\arg\min \widehat{L}(\boldsymbol{\theta})$

A scatter plot of $d \times n = 5,000$ hidden parameters (a_j, b_j, c_j) obtained from $n = 500$ neural networks with $p = 10$ hidden units.

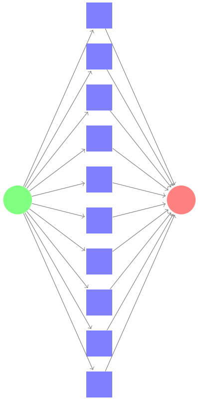


$\times 500$

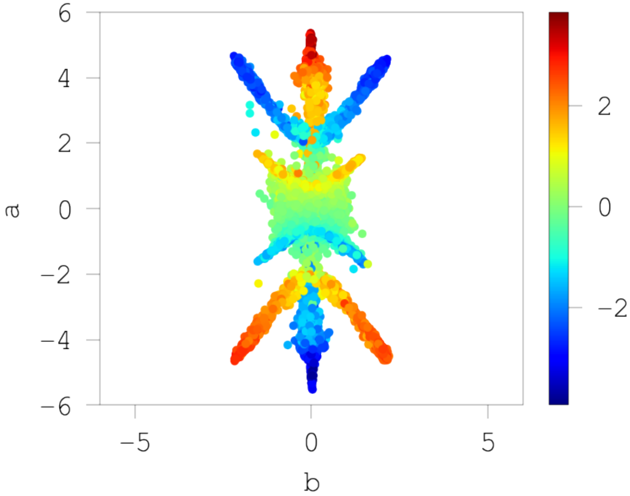


$\arg\min \widehat{L}(\boldsymbol{\theta})$

A scatter plot of $d \times n = 10,000$ hidden parameters (a_j, b_j, c_j) obtained from $n = 1,000$ neural networks with $p = 10$ hidden units.



$\times 1000$

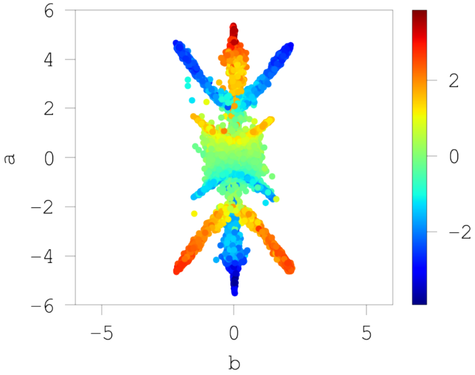


$\arg\min \widehat{L}(\boldsymbol{\theta})$

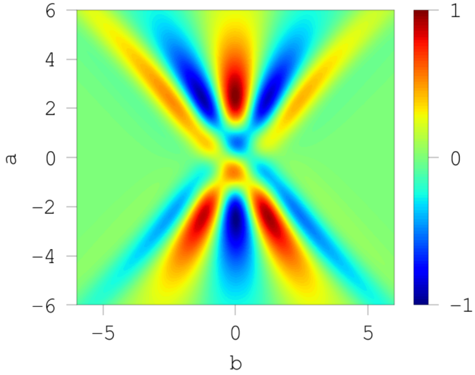
Ridgelet Transform Characterizes Gradient Descent Solutions

Theorem (simplified, S-Ishikawa-Ikeda, AISTATS2021)

$$\lim_{n\to\infty}\lim_{d\to\infty}\ \mathop{\arg\min}_{\gamma_d=\sum_{i=1}^{d} c_i\delta_{(\boldsymbol{a}_i,b_i)}} \left[\frac{1}{n}\sum_{i=1}^{n}|f(\boldsymbol{x}_i)-S[\gamma_d](\boldsymbol{x}_i)|^2+\beta\|(c_i)\|_{\ell^2}^2\right] = R\left[\frac{f}{\beta+1};\sigma_*\right]$$



scatter plot of GD trained parameters



ridgelet transform

How to find the ridgelet transform R?

Problem: NN learning equation

Given a data generating function f and network $\mathtt{NN}[\gamma]$, find the unknown parameter γ satisfying

$$\mathtt{NN}[\gamma] = f$$

- By inspiration
  - Murata, An Integral Representation of Functions..., Neural Networks, 1996.
  - Candes, Ridgelets: Theory and Applications, dissertation, 1998.

- By Fourier expression
  - General solution for fully-connected SNN S-Ishikawa-Ikeda, arXiv:2106.04770
  - SNN on finite fields $\mathbb{Z}/p\mathbb{Z}$ Yamasaki-Subramanian-Hayakawa-S, ICML2023
  - Group convolution network S-Ishikawa-Ikeda, NeurIPS2022
  - SNN on manifold (noncompact symmetric space G/K) S-Ishikawa-Ikeda, ICML2022

- By group equivariant/invariant functions
  - joint-group-equivariant map S-Hashimoto-Ishikawa-Ikeda, arXiv2405.13682 $\leftarrow$ NEW!
  - joint-group-invariant map S-Ishi-Ishikawa-Ikeda, NeurReps2023
  - formal deep network S-Hashimoto-Ishikawa-Ikeda, NeurReps2023

Solve $S[\gamma] = f$

Appendix A.3, in Sonoda-Ishikawa-Ikeda, arXiv:2106.04770

Step 1. Turn the network into a *Fourier expression* as below.

$$S[\gamma](\boldsymbol{x}) = \int_{\mathbb{R}^m} \left[\int_{\mathbb{R}} \gamma(\boldsymbol{a}, b)\sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b)\mathrm{d}b \right] \mathrm{d}\boldsymbol{a}$$

By an identity $\frac{1}{2\pi}\int_{\mathbb{R}} \gamma^\sharp(\boldsymbol{a}, \omega)\sigma^\sharp(\omega)e^{i\omega b}\mathrm{d}\omega = (\gamma(\boldsymbol{a}, \bullet) * \sigma)(b)$,

$$= \int_{\mathbb{R}^m} \left[\frac{1}{2\pi} \int_{\mathbb{R}} \gamma^\sharp(\boldsymbol{a}, \omega)\sigma^\sharp(\omega)e^{i\omega \boldsymbol{a} \cdot \boldsymbol{x}}\mathrm{d}\omega \right] \mathrm{d}\boldsymbol{a}$$

By changing the variable $(\boldsymbol{a}, \omega) = (\boldsymbol{\xi}/\omega, \omega)$,

$$= \frac{1}{2\pi} \int_{\mathbb{R}} \left[\int_{\mathbb{R}^m} \gamma^\sharp(\boldsymbol{\xi}/\omega, \omega)e^{i\boldsymbol{\xi} \cdot \boldsymbol{x}}\mathrm{d}\boldsymbol{\xi} \right] |\omega|^{-m}\sigma^\sharp(\omega)\mathrm{d}\omega,$$

where $\cdot^\sharp$ denotes the Fourier transform in b

Recap:

$$S[\gamma](\boldsymbol{x}) = \frac{1}{2\pi}\int_{\mathbb{R}}\left[\int_{\mathbb{R}^m}\gamma^\sharp(\boldsymbol{\xi}/\omega,\omega)e^{i\boldsymbol{\xi}\cdot\boldsymbol{x}}\mathrm{d}\boldsymbol{\xi}\right]|\omega|^{-m}\sigma^\sharp(\omega)\mathrm{d}\omega$$

Step 2. Assume a *separation-of-variables* form

$$\gamma^\sharp_{f,\rho}(\boldsymbol{\xi}/\omega,\omega) := \widehat{f}(\boldsymbol{\xi})\overline{\rho^\sharp(\omega)}$$

Then, $\gamma_{f,\rho}$ is a particular solution

$$S[\gamma_{f,\rho}] = \frac{1}{2\pi}\left[\int \sigma^\sharp(\omega)\overline{\rho^\sharp(\omega)}|\omega|^{-m}\mathrm{d}\omega\right]\left[\int \widehat{f}(\boldsymbol{\xi})e^{i\boldsymbol{\xi}\cdot\boldsymbol{x}}\mathrm{d}\boldsymbol{\xi}\right] = (\!(\sigma,\rho)\!)f$$

Furthermore, $\gamma_{f,\rho}(\boldsymbol{a},b) = R[f;\rho](\boldsymbol{a},b)$.

$$\begin{aligned}
\gamma_{f,\rho}(\boldsymbol{a},b) &= \frac{1}{2\pi}\int_{\mathbb{R}}\widehat{f}(\omega\boldsymbol{a})\overline{\rho^\sharp(\omega)}e^{ib\omega}\mathrm{d}\omega \\
&= \frac{1}{2\pi}\int_{\mathbb{R}^m\times\mathbb{R}} f(\boldsymbol{x})\overline{\rho^\sharp(\omega)}e^{i\omega(b-\boldsymbol{a}\cdot\boldsymbol{x})}\mathrm{d}\boldsymbol{x}\mathrm{d}\omega \\
&= \int_{\mathbb{R}^m} f(\boldsymbol{x})\overline{\rho(\boldsymbol{a}\cdot\boldsymbol{x}-b)}\mathrm{d}\boldsymbol{x} \\
&=: R[f;\rho](\boldsymbol{a},b).
\end{aligned}$$

A *general solution* is given by

$$\gamma_f := R[f;\rho_0] + \sum_{ij} c_{ij} R[e_i;\rho_j].$$

Here,

- $\{c_{ij}\}_{i,j\in\mathbb{N}^2}$ is an ℓ^2-sequence
- $\{e_i\}_{i\in\mathbb{N}}$ is an o.n.b. of $L^2(\mathbb{R}^m)$,
- $\{\rho_j\}_{j\in\mathbb{N}}$ is a subsystem of o.n.s. of $L^2_m(\mathbb{R})$ satisfying $(\!(\sigma,\rho_j)\!) = 0$.

Necessity: γ_f is a solution.

$$S[\gamma_f] = (\!(\sigma,\rho_0)\!) f(\boldsymbol{x}) + \sum_{ij} c_{ij} (\!(\sigma,\rho_j)\!) e_i(\boldsymbol{x}) = (\!(\sigma,\rho_0)\!) f(\boldsymbol{x}) + 0.$$

Sufficiency: (Lemma) Any $\gamma \in L^2(\mathbb{R}^m \times \mathbb{R})$ can be expanded as

$$\gamma = \sum_{ij} \langle \gamma, R[e_i;\rho_j] \rangle R[e_i;\rho_j]$$

Deep Ridgelet Transform for Joint-Equivariant Feature Maps [1]

—the first ridgelet transform for deep networks—

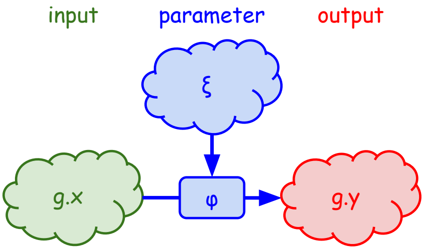
[1] S-Hashimoto-Ishikawa-Ikeda, arXiv:2405.13682, 2024

Joint-G-Equivariant Map

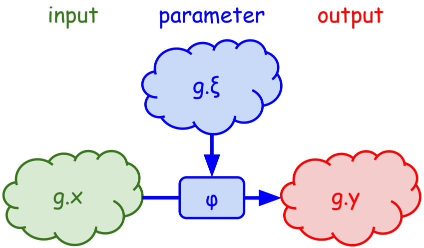
Definition (Joint-G-Equivariant Map)

- Let G be a group
- Let X, Y and Ξ be G-spaces
- A map $\phi : X \times \Xi \to Y$ is joint-G-equivariant when it satisfies

$$\phi(g \cdot x, g \cdot \xi) = g \cdot \phi(x, \xi), \quad \text{for all } g \in G \text{ and } (x, \xi) \in X \times \Xi$$



G-equivariance



joint-G-equivariance

Joint-Equivariant Maps are Ubiquitous

Lemma (1. Easy to synthesize)

- *Let G be a group, and X, Y be G-spaces*
- *Let $\phi_0 : X \to Y$ be any map*
- *Then, the following $\phi : X \times G \to Y$ with*

$$\phi(x, \xi) := \xi \cdot \phi_0(\xi^{-1} \cdot x), \quad (x, \xi) \in X \times G$$

is joint-equivariant

Lemma (2. Easy to be realized as a network)

- *Let $\phi : \mathbb{R}^m \times G \to \mathbb{R}^n$ be a joint-equivariant map*
- *Then, the following depth-2 fully-connected network*

$$\mathtt{NN}(\boldsymbol{x}, \xi) := \int_{\mathbb{R}^m \times \mathbb{R}} \mathrm{R}_2[\phi(\bullet, \xi)](\boldsymbol{a}, b)\sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b)\mathrm{d}\boldsymbol{a}\mathrm{d}b, \quad (\boldsymbol{x}, \xi) \in \mathbb{R}^m \times G$$

is joint-equivariant

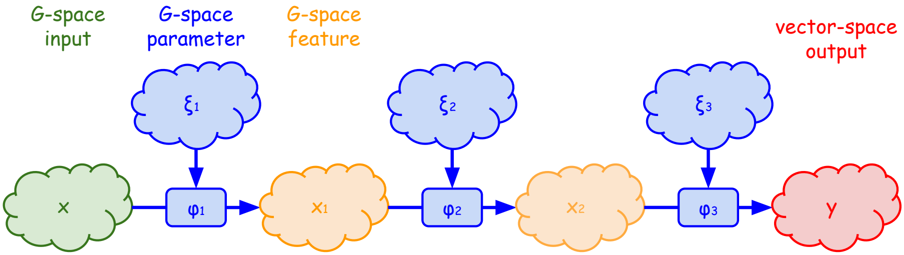
Composite of joint-equivariant maps is joint-equivariant

Lemma (3. Easy to compose)

- *Let* $\phi_i : X_{i-1} \times \Xi_i \to X_i$ *be joint-equivariant maps*
- *Then, the composite* $\phi : X_0 \times (\Xi_1 \times \cdots \times \Xi_n) \to X_n$

$$\phi(x, \xi) := \phi_n(\bullet, \xi_n) \circ \cdots \circ \phi_1(x, \xi_1)$$

is joint-equivariant as well.



Main Results

Definition

- Let G be a locally compact group
- Let X (input domain) and Ξ (parameter domain) be G-spaces with Haar measures $\mathrm{d}x$ and $\mathrm{d}\xi$,
- Let Y (output domain) be a separable Hilbert space with unitary G-action
- Let $\phi, \psi : X \times \Xi \to Y$ be joint-equivariant maps
- For any maps $f : X \to Y$ and $\gamma : \Xi \to \mathbb{C}$, put

$$\mathtt{DNN}[\gamma;\phi](x) := \int_\Xi \gamma(\xi)\phi(x,\xi)\mathrm{d}\xi, \qquad \mathtt{R}[f;\psi](\xi) := \int_X \langle f(x), \psi(x,\xi)\rangle_Y \mathrm{d}x.$$

Theorem (Reconstruction Formula)

- *Suppose that the induced representation $\pi : G \to \mathcal{U}(L^2(X;Y))$ is irreducible*
- *Then, there exists a constant $c_{\phi,\psi}$ such that for any $f \in L^2(X;Y)$,*

$$\mathtt{DNN} \circ \mathtt{R}[f] = c_{\phi,\psi} f.$$

Proof

Theorem (Schur's Lemma)

A unitary representation π of G on $\mathcal{H}$ is irreducible

$\iff$ *If a bounded operator $T : \mathcal{H} \to \mathcal{H}$ commutes with π, then $T = c\,\mathrm{Id}_{\mathcal{H}}$ for some $c \in \mathbb{C}$*

We can check $T := \mathtt{DNN} \circ \mathtt{R} : L^2(X;Y) \to L^2(X;Y)$ commutes with π as below: For each $g \in G$,

$$\mathtt{R}[\pi_g[f]](\xi) = \int_X \langle g \cdot f(g^{-1} \cdot x), \psi(x,\xi)\rangle_Y \mathrm{d}x = \int_X \langle f(x), \psi(x, g^{-1} \cdot \xi)\rangle_Y \mathrm{d}x = \widehat{\pi}_g[\mathtt{R}_\psi[f]](\xi),$$

$$\mathtt{NN}[\widehat{\pi}_g[\gamma]](x) = \int_\Xi \gamma(g^{-1} \cdot \xi)\phi(x,\xi)\mathrm{d}\xi = \int_\Xi \gamma(\xi)\left(g \cdot \phi(g^{-1} \cdot x, \xi)\right)\mathrm{d}\xi = \pi_g[\mathtt{NN}_\phi[\gamma]](x).$$

Therefore,

$$\mathtt{NN} \circ \mathtt{R} \circ \pi_g = \mathtt{NN} \circ \widehat{\pi}_g \circ \mathtt{R} = \pi_g \circ \mathtt{NN} \circ \mathtt{R}.$$

Hence Schur's lemma yields that there exist a constant $c_{\phi,\psi} \in \mathbb{C}$ such that $\mathtt{NN}_\phi \circ \mathtt{R}_\psi = c_{\phi,\psi}\,\mathrm{Id}_{L^2(X;Y)}$.

Example: Depth-2 Fully-Connected Network

- G is the affine group $\mathrm{Aff}(m) = GL(m) \ltimes \mathbb{R}^m$
- $X = \mathbb{R}^m$ (data domain) with G-action

$$g \cdot \boldsymbol{x} := L\boldsymbol{x} + \boldsymbol{t}, \quad g = (L, \boldsymbol{t}) \in G,$$

- $\Xi = \mathbb{R}^m \times \mathbb{R}$ (parameter domain) with dual G-action

$$g \cdot (\boldsymbol{a}, b) = (L^{-\top}\boldsymbol{a}, b + \boldsymbol{t}^\top L^{-\top}\boldsymbol{a}), \quad g = (L, \boldsymbol{t}) \in G$$

- Then, $\phi(\boldsymbol{x}, (\boldsymbol{a}, b)) := \sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b)$ and $\psi(\boldsymbol{x}, (\boldsymbol{a}, b)) := \rho(\boldsymbol{a} \cdot \boldsymbol{x} - b)$ are joint-G-invariant. Indeed,

$$\phi(g \cdot \boldsymbol{x}, g \cdot (\boldsymbol{a}, b)) = \sigma\left(L^{-\top}\boldsymbol{a} \cdot (L\boldsymbol{x} + \boldsymbol{t}) - (b + \boldsymbol{t}^\top L^{-\top}\boldsymbol{a})\right) = \sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b) = \phi(\boldsymbol{x}, (\boldsymbol{a}, b))$$

- Put

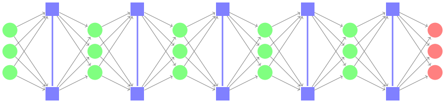
$$\mathtt{NN}_2[\gamma](\boldsymbol{x}) = \int_{\mathbb{R}^m \times \mathbb{R}} \gamma(\boldsymbol{a}, b)\sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b)\mathrm{d}\boldsymbol{a}\mathrm{d}b, \quad \text{and} \quad \mathtt{R}_2[f](\boldsymbol{a}, b) = \int_{\mathbb{R}^m} f(\boldsymbol{x})\overline{\rho(\boldsymbol{a} \cdot \boldsymbol{x} - b)}\mathrm{d}\boldsymbol{x}$$

Lemma

The regular representation $\pi_g : \mathrm{Aff}(m) \to \mathcal{U}(L^2(\mathbb{R}^m; \mathbb{R}))$ *is irreducible*

- So, $\mathtt{NN}_2 \circ \mathtt{R}_2 = (\!(\sigma, \rho)\!)\, \mathrm{Id}_{L^2(\mathbb{R}^m)}$

Example: Depth-n Fully-Connected Network



- For any $\boldsymbol{f}_i \in L^2(\mathbb{R}^m;\mathbb{R}^m)$ and $\xi = G := (Q, L, \boldsymbol{t}) \in O(m) \times (GL(m) \ltimes \mathbb{R}^m)$, put

$$\mathtt{NN}_i(\boldsymbol{x};\xi) := \int_{\mathbb{R}^m \times \mathbb{R}} Q\mathtt{R}_2[\boldsymbol{f}_i](\boldsymbol{a}, b)\sigma(\boldsymbol{a} \cdot L^{-1}(\boldsymbol{x} - \boldsymbol{t}) - b)\mathrm{d}\boldsymbol{a}\mathrm{d}b$$

- For any functions $\gamma : \mathbb{R}^m \to \mathbb{C}$ and $\boldsymbol{f} : \mathbb{R}^m \to \mathbb{R}^m$, put

$$\mathtt{DNN}[\gamma](\boldsymbol{x}) := \int_{G^n} \gamma(\boldsymbol{\xi})\mathtt{NN}_n(\bullet, \xi_n) \circ \cdots \circ \mathtt{NN}_1(\boldsymbol{x}, \xi_1)\mathrm{d}\boldsymbol{\xi},$$

$$\mathtt{R}_n[\boldsymbol{f}](\boldsymbol{\xi}) := \int_X \langle \boldsymbol{f}(\boldsymbol{x}), \mathtt{NN}_n(\bullet, \xi_n) \circ \cdots \circ \mathtt{NN}_1(\boldsymbol{x}, \xi_1)\rangle \mathrm{d}\boldsymbol{x}$$

Lemma

The unitary representation $\pi_g : G \to \mathcal{U}(L^2(\mathbb{R}^m;\mathbb{R}^m))$ is irreducible

Thus, $\mathtt{DNN} \circ \mathtt{R}_n[\boldsymbol{f}] = c\boldsymbol{f}$

Conclusion

- Ultimate goal:
  - Characterize deep solutions

- integral representation $S[\gamma]$ is a linearization trick of NN parameters
- ridgelet transform $R[f]$ is its right inverse operator: $S[R[f]] = ((\sigma, \rho))f$
  - can prove the universality in a constructive manner
  - can visualize/analyze the distribution γ of parameters

- How to find the ridgelet transform?
  - (By inspiration)
  - By Fourier expression
  - By group equivariant functions

Supplementary Slides

NNs on Finite Field

Yamasaki et al. Quantum Ridgelet Transform: Winning Lottery Ticket of Neural Networks with Quantum Computation, ICML2023

Let $\mathbb{F} = \mathbb{Z}/p\mathbb{Z}$.

$$S[\gamma](\boldsymbol{x}) := \sum_{(\boldsymbol{a},b)\in\mathbb{F}^m\times\mathbb{F}} \gamma(\boldsymbol{a},b)\sigma(\boldsymbol{a}\cdot\boldsymbol{x}-b), \quad \boldsymbol{x}\in\mathbb{F}^m$$

$$R[f;\rho](\boldsymbol{a},b) = \sum_{\boldsymbol{x}\in\mathbb{F}^m} f(\boldsymbol{x})\overline{\rho(\boldsymbol{a}\cdot\boldsymbol{x}-b)}, \quad (\boldsymbol{a},b)\in\mathbb{F}^m\times\mathbb{F}$$

$$(\!(\sigma,\rho)\!) = \frac{1}{|\mathbb{F}|^{m-1}} \sum_{\omega\in\mathbb{F}} \sigma^\sharp(\omega)\overline{\rho^\sharp(\omega)}$$

$$S[R[f;\rho]](\boldsymbol{x}) = (\!(\sigma,\rho)\!) f(\boldsymbol{x})$$

Sketch Proof

First, turn to the Fourier expression. Then, assume the separation-of-variables form.

$$
\begin{aligned}
S[\gamma](\boldsymbol{x}) &:= \sum_{(\boldsymbol{a},b)\in\mathbb{F}^m\times\mathbb{F}} \gamma(\boldsymbol{a},b)\sigma(\boldsymbol{a}\cdot\boldsymbol{x}-b) \\
&= \frac{1}{|\mathbb{F}|} \sum_{(\boldsymbol{a},\omega)\in\mathbb{F}^m\times\mathbb{F}} \gamma^\sharp(\boldsymbol{a},\omega)\sigma(\omega)e^{2\pi i\omega\boldsymbol{a}\cdot\boldsymbol{x}}
\end{aligned}
$$

Put $\boldsymbol{\xi} = \omega\boldsymbol{a}$

$$
= \frac{1}{|\mathbb{F}|} \sum_{(\boldsymbol{\xi},\omega)\in\mathbb{F}^m\times\mathbb{F}} \gamma^\sharp(\boldsymbol{\xi}/\omega,\omega)\sigma(\omega)e^{2\pi i\boldsymbol{\xi}\cdot\boldsymbol{x}}
$$

Assume $\gamma^\sharp(\boldsymbol{\xi}/\omega,\omega) = \widehat{f}(\boldsymbol{\xi})\overline{\rho^\sharp(\omega)}$

$$
\begin{aligned}
&= \left(|\mathbb{F}|^{m-1} \sum_{\omega\in\mathbb{F}} \sigma^\sharp(\omega)\overline{\rho^\sharp(\omega)} \right) \left(\frac{1}{|\mathbb{F}|^m} \sum_{\boldsymbol{\xi}\in\mathbb{F}^m} \widehat{f}(\boldsymbol{\xi})e^{2\pi i\boldsymbol{\xi}\cdot\boldsymbol{x}} \right) \\
&= (\!(\sigma,\rho)\!) f(\boldsymbol{x})
\end{aligned}
$$

Group Convolutional NNs on Hilbert Space $\mathcal{H}$[2]

[2] S-Ishikawa-Ikeda, NeurIPS2022

Fourier Transform on a Hilbert space $\mathcal{H}_m \subset \mathcal{H}$

Definition

Let $\mathcal{H}$ be a Hilbert space, $\mathcal{H}_m \subset \mathcal{H}$ be an m-dimensional subspace, and λ be the Lebesgue measure induced from $\mathbb{R}^m$. Put

$$\widehat{f}(\xi) := \int_{\mathcal{H}_m} f(x) e^{-i\langle x, \xi \rangle} \mathrm{d}\lambda(x), \quad x \in \mathcal{H}_m$$

Theorem

For any $f \in L^2(\mathcal{H}_m)$,

$$\frac{1}{(2\pi)^m} \int_{\mathcal{H}_m} \widehat{f}(\xi) e^{i\langle x, \xi \rangle} \mathrm{d}\lambda(\xi) = f(x), \quad x \in \mathcal{H}_m$$

Definition (Generalized group convolution)

Let G be a group, $\mathcal{H}$ be a Hilbert space, and $T : G \to GL(\mathcal{H})$ be a group representation. The (G, T)-convolution is given by

$$(a * x)(g) := \langle T_{g^{-1}}[x], a \rangle_{\mathcal{H}}, \quad a, x \in \mathcal{H}.$$

Definition (Group CNN)

Let $\mathcal{H}_m \subset \mathcal{H}$ be an m-dimensional subspace equipped with the Lebesgue measure λ. Put

$$S[\gamma](x)(g) := \int_{\mathcal{H}_m \times \mathbb{R}} \gamma(a, b) \sigma((a * x)(g) - b) \mathrm{d}\lambda(a) \mathrm{d}b, \quad x \in \mathcal{H}, g \in G$$

Example (Cyclic CNN for n-channel $m \times m$-image)

$$\mathtt{CNN}(\boldsymbol{x})(p, q) = \sum_{\ell=1}^{n'} c^\ell \sigma \left(\sum_{k=1}^{n} \sum_{i,j=1}^{m} a_{ij}^{k\ell} x_{i+p,j+q}^k - b^\ell \right), \quad \boldsymbol{x} = (x_{ij}^k) \in \mathbb{R}^{m^2 \times n}, \ (p, q) \in (\mathbb{Z}/m\mathbb{Z})^2$$

i.e., $G = (\mathbb{Z}/m\mathbb{Z})^2, \mathcal{H} = \mathbb{R}^{m^2 \times n}, \ T_{p,q}(\boldsymbol{x}) := (x^{\bullet}_{\bullet - p, \bullet - q})$

In the following, $e \in G$ denotes the identity element.

Definition (Ridgelet Transform)

For any function $f : \mathcal{H}_m \to \mathbb{C}^G$ and $\rho : \mathbb{R} \to \mathbb{C}$, put

$$R[f;\rho](a,b) := \int_{\mathcal{H}_m} f(x)(e)\overline{\rho(\langle a, x\rangle_{\mathcal{H}} - b)}\mathrm{d}\lambda(x).$$

Definition ((G,T)-Equivariance)

A (nonlinear) map $f : \mathcal{H} \to \mathbb{C}^G$ is (G,T)-equivariant when

$$f(T_g[x])(h) = f(x)(g^{-1}h), \quad \forall x \in \mathcal{H}_m, g, h \in G$$

Theorem (Reconstruction Formula)

Suppose that f is (G,T)-equivariant and $f(\bullet)(e) \in L^2(\mathcal{H}_m)$, then $S[R[f;\rho]] = ((\sigma,\rho))f$.

- Meaning: Universality of continuous GCNN
- Corollary: cc-universality of finite GCNNs

Sketch Proof

Step 1. Turn to Fourier expression:

$$
\begin{aligned}
S[\gamma](x)(g) &= \int_{\mathcal{H}_m \times \mathbb{R}} \gamma(a,b)\sigma(\langle T_{g^{-1}}[x], a\rangle_{\mathcal{H}} - b)\mathrm{d}a\mathrm{d}b \\
&= \frac{1}{2\pi}\int_{\mathcal{H}_m \times \mathbb{R}} \gamma^\sharp(a,\omega)\sigma^\sharp(\omega)e^{i\omega\langle T_{g^{-1}}[x], a\rangle_{\mathcal{H}}}\mathrm{d}a\mathrm{d}\omega \\
&= \frac{1}{2\pi}\int_{\mathcal{H}_m \times \mathbb{R}} \gamma^\sharp(\xi/\omega,\omega)\sigma^\sharp(\omega)e^{i\langle T_{g^{-1}}[x], \xi\rangle_{\mathcal{H}}}|\omega|^{-m}\mathrm{d}\xi\mathrm{d}\omega.
\end{aligned}
$$

Step 2. Put separation-of-variables form:

$$
\gamma^\sharp_{f,\rho}(\xi/\omega,\omega) := \widehat{f}(\xi)(e)\overline{\rho^\sharp(\omega)}.
$$

By the construction it is a particular solution:

$$
\begin{aligned}
S[\gamma_{f,\rho}](x)(g) &= \frac{1}{2\pi}\int_{\mathcal{H}_m} \widehat{f}(\xi)(e)e^{i\langle T_{g^{-1}}[x], \xi\rangle_{\mathcal{H}}}\mathrm{d}\lambda(\xi)\int_{\mathbb{R}} \sigma^\sharp(\omega)\overline{\rho^\sharp(\omega)}|\omega|^{-m}\mathrm{d}\omega \\
&= ((\sigma,\rho))f(x)(g).
\end{aligned}
$$

and $\gamma_{f,\rho} = R[f;\rho]$.

Fully-Connected NNs on Noncompact Symmetric Space[3]

[3] S-Ishikawa-Ikeda, ICML2022

Past Attempts to Design Neural Networks on Manifolds "G/K"

Difficulty: But, how to define an "affine map $\boldsymbol{a} \cdot \boldsymbol{x} - b$ on a manifold"?

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 $x = u^\top \lambda u$

Ideas

1. Use the Fourier transform on manifolds, and
2. geometrically rewrite $\boldsymbol{a} \cdot \boldsymbol{x} - b$ as the *distance between point $\boldsymbol{x}$ and hyperplane ξ*, namely,

$$\boldsymbol{a} \cdot \boldsymbol{x} - b = r d(\boldsymbol{x}, \xi).$$

where

- $\boldsymbol{a} = r\boldsymbol{u}$ (polar coordinates $(r, \boldsymbol{u}) \in \mathbb{R} \times \mathbb{S}^{m-1}$)
- $\xi := \{\boldsymbol{y} \in \mathbb{R}^m \mid \boldsymbol{a} \cdot \boldsymbol{y} = b\}$ (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

Geometric Reparametrization of Euclidean NN

For scales $r_i > 0$, hyperplanes $\xi_i \subset \mathbb{R}^m$, and weights $c_i \in \mathbb{R}$,

$$\mathtt{SNN}(\boldsymbol{x}) = \sum_{i=1}^{p} c_i \sigma(r_i d(\boldsymbol{x}, \xi_i)), \quad \boldsymbol{x} \in \mathbb{R}^m$$

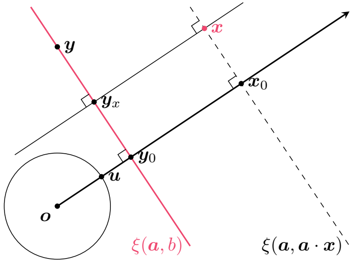


Figure: Points $\boldsymbol{x}$ and hyperplane ξ in the Euclidean space $\mathbb{R}^m$

3. we employed *horospheres* as the G/K-counter of hyperplanes

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 (K compact, A abelian, N nilpotent).
- A noncompact symmetric space is given by the quotient (the set of all left cosets)

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

Example (Hyperbolic Space $\mathbb{H}^m = O(1,m)/O(1) \times O(m)$)

for embedding words, and tree-structured dataset

Example (SPD Manifold $\mathbb{P}_m = GL(m)/O(m)$)

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

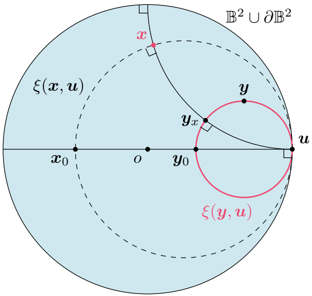


Figure: Poincare🐮 Disk $\mathbb{B}^2$ is a noncompact symmetric space $SU(1,1)/SO(2)$

Basic Example: Hyperbolic Space $\mathbb{H}^m$

Poincare🐯 ball model $\mathbb{B}^m$—One of a few models of $\mathbb{H}^m$

- Unit ball $\mathbb{B}^m := \{\boldsymbol{x} \in \mathbb{R}^m \mid |\boldsymbol{x}|_E < 1\}$ equipped with metric $g_{\boldsymbol{x}} = \frac{4}{(1-|\boldsymbol{x}|_E^2)^2} \sum_{i=1}^m \mathrm{d}x_i \otimes \mathrm{d}x_i$,
- $d_P(\boldsymbol{x}, \boldsymbol{y}) = \cosh^{-1}\left(1 + \frac{2|\boldsymbol{x}-\boldsymbol{y}|_E^2}{(1-|\boldsymbol{x}|_E^2)(1-|\boldsymbol{y}|_E^2)}\right)$; $\quad \mathrm{d}\,\mathrm{vol}_g(\boldsymbol{x}) = \left(\frac{2}{1-|\boldsymbol{x}|_E^2}\right)^m \mathrm{d}\boldsymbol{x}$

Basic objects in $\mathbb{B}^m$

- Boundary/Ideal sphere $\partial\mathbb{B}^m$:= the set of points at infinity
  - Unit sphere $\partial\mathbb{B}^m = \{\boldsymbol{x} \in \mathbb{R}^m \mid |\boldsymbol{x}|_E = 1\}$
- Geodesic
  - Euclidean arcs/lines orthogonal to $\partial\mathbb{B}^m$
- (Hyperbolic) sphere $S(\boldsymbol{x}, r) := \{\boldsymbol{y} \in \mathbb{B}^m \mid d_P(\boldsymbol{x}, \boldsymbol{y}) = r\}$
  - Euclidean sphere (but the center is biased outward)
- Horosphere ξ := a sphere with infinite radius $S(\boldsymbol{x}, \infty)$
  - Euclidean sphere tangent to $\partial\mathbb{B}^m$
  - A hyperbolic counterpart of the *Euclidean hyperplane*, since it is a *Euclidean sphere with infinite radius*

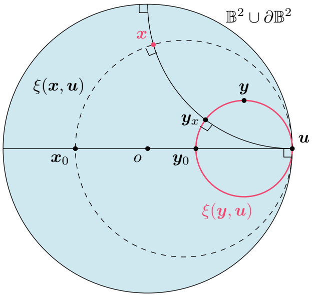
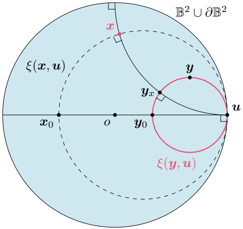


Figure: Poincare🐯 Disk $\mathbb{B}^2$

Horospheres in G/K and Vector-Valued Composite Distance $\langle x, u\rangle$

Known Facts on the Horospheres in G/K

- On every G/K, the horosphere ξ can be defined.
- A horosphere $\xi(x, u)$ is parametrized by points $x \in X$ and $u \in \partial X$
  like "a horosphere ξ passing through point x with normal u"



Vector-Valued Composite Distance $\langle x, u\rangle$

- denotes the distance between origin o and a horosphere $\xi(x, u)$ as

$$\langle x, u\rangle := d(o, \xi(x, u)) \quad \Big(= d(o, x_o)\Big)$$

- In general, it is *vector-valued*, which means that the absolute value $|\langle x, u\rangle|$ coincides with the Riemannian distance
- To be precise, it is $\mathfrak{a}$*-valued*, where $\mathfrak{a}$ is the Lie algebra of A in $G = KAN$

Fourier Transform on $X = G/K$

Helgason, GGA (1984, Introduction); GASS (2008, Chapter III)

Definition (Helgason-Fourier Transform)

For any function $f : X \to \mathbb{C}$,

$$\widehat{f}(\lambda, u) := \int_X f(x) e^{(-i\lambda + \varrho)\langle x, u \rangle} \mathrm{d}x, \quad (\lambda, u) \in \mathfrak{a}^* \times \partial X$$

with a certain constant vector $\varrho \in \mathfrak{a}^*$. Here, $\langle x, u \rangle$ denotes the *vector-valued distance* from the origin o to the *horosphere* through point $x \in G/K$ with normal $u \in \partial X$.

Theorem (Inversion Formula)

For any $f \in L^2(X)$ (or $f \in C_c^\infty(X)$),

$$f(x) = |W|^{-1} \int_{\mathfrak{a}^* \times \partial X} \widehat{f}(\lambda, u) e^{(i\lambda + \varrho)\langle x, u \rangle} |\boldsymbol{c}(\lambda)|^{-2} \mathrm{d}\lambda \mathrm{d}u, \quad x \in X$$

where $\boldsymbol{c}$ is the Harish-Chandra $\boldsymbol{c}$-function, and $|W|$ is a constant.

This is a “Fourier transform” because $e^{(-i\lambda + \varrho)\langle x, u \rangle}$ is the eigenfunction of Laplace-Beltrami operator Δ_X on X

Definition (Fully-Connected NNs on Noncompact Symmetric Space G/K)

Let G be a connected semisimple real Lie group, let $G = KAN$ be the Iwasawa decomposition, and let $X := G/K$ be the noncompact symmetric space. Put

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

where $\mathfrak{a}^*$ is the dual of Lie algebra of A, ∂X is the boundary, and $\langle x, u\rangle$ is an X-counter of the Euclidean inner product $\boldsymbol{x} \cdot \boldsymbol{u}$ for $(\boldsymbol{x}, \boldsymbol{u}) \in \mathbb{R}^m \times \mathbb{S}^{m-1}$.

Example (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 $\mathfrak{g} = 4(1 - |\boldsymbol{x}|)^{-2} \sum_{i=1}^m \mathrm{d}x_i \otimes \mathrm{d}x_i$,

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

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

Definition (Ridgelet Transform)

For any function $f : X \to \mathbb{C}$ and an auxiliary function $\rho : \mathbb{R} \to \mathbb{C}$, put

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

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

Theorem (Reconstruction Formula)

For any $\sigma \in \mathcal{S}'(\mathbb{R})$, $\rho \in \mathcal{S}(\mathbb{R})$, and $f \in L^2(X)$, we have

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

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

- Meaning: Universality of continuous Fully-Connected NN on X
- Corollary: cc-universality of finite Fully-Connected NNs on X

Sketch Proof

- Given a function $f : G/K \to \mathbb{C}$, consider solving an integral equation $S[\gamma] = f$ of unknown γ.
- Step 1: Change the frame of $S[\gamma]$ from neurons to a *Fourier expression*:

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

  where $\sharp$ denotes the Euclidean-Fourier transform in b.
- 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 can verify that $\gamma_{f,\rho}$ is the ridgelet transform $R[f; \rho]$.

Generalized NN and Ridgelet Transform Induced from Invariant Functions [4]

[4] S-Ishi-Ishikawa-Ikeda, NeurReps2023

Joint G-Invariant Function

- Let G be a locally compact group equipped with invariant Haar measure $\mathrm{d}g$
  - e.g. a finite group, compact group, Lie group, ...
- Let X and Ξ be homogeneous G-spaces equipped with invariant Haar measures $\mathrm{d}x$ and $\mathrm{d}\xi$
  - we call X the "data domain" and Ξ the "parameter domain"
- Let π_g and $\widehat{\pi}_g$ be regular G-acitions on $L^2(X)$ and $C_c(\Xi)$, i.e.

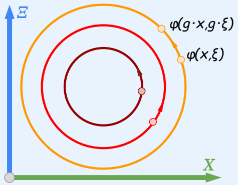
$$\pi_g[f](x) := f(g^{-1} \cdot x), \quad \text{and} \quad \widehat{\pi}_g[\gamma](\xi) := \gamma(g^{-1} \cdot \xi)$$

Definition (Joint G-Invariant Functions)

- We say a function ϕ on $X \times \Xi$ is joint G-invariant when

$$\phi(g \cdot x, g \cdot \xi) = \phi(x, \xi),$$

  for all $g \in G$ and $(x, \xi) \in X \times \Xi$.
- $\mathcal{A}$ denote the algebra of all joint G-invariant functions.



Main Results

Definition (Generalized Network)

For any joint G-invariant function $\phi \in \mathcal{A}$ and $\gamma \in L^2(\Xi)$, put

$$\mathtt{NN}[\gamma;\phi](x) := \int_\Xi \gamma(\xi)\phi(x,\xi)\mathrm{d}\xi, \quad x \in X$$

Definition (Generalized Ridgelet Transform)

For any joint G-invariant function $\psi \in \mathcal{A}$ and $f \in L^2(X)$, put

$$\mathtt{R}[f;\psi](\xi) := \int_X f(x)\overline{\psi(x,\xi)}\mathrm{d}x, \quad \xi \in \Xi$$

Theorem (Reconstruction Formula)

Supopse that the regular action π_g is an irreducible unitary representation of G on $L^2(X)$. Then, there exists a bilinear form $(\!(\phi,\psi)\!)$ such that for any $f \in L^2(X)$

$$\mathtt{NN}_\phi[\mathtt{R}_\psi[f]] = \int_\Xi \mathtt{R}[f;\psi](\xi)\phi(\bullet,\xi)\mathrm{d}\xi = (\!(\phi,\psi)\!) f.$$

Proof

Theorem (Schur's Lemma)

A unitary representation $(\pi, \mathcal{H})$ of G is irreducible

$\iff$ Any bounded operator T that commutes with π is a scalar multiple of the identity

$$\iff \qquad \forall g \in G \ [T \circ \pi_g = \pi_g \circ T] \implies \exists c \in \mathbb{C} \ [T = c \operatorname{Id}_{\mathcal{H}}]$$

By the left-invariances of $\mathrm{d}x$ and ψ (resp. $\mathrm{d}\xi$ and ϕ), we have

$$\mathtt{R}_\psi[\pi_g[f]](\xi) = \int_X f(g^{-1} \cdot x)\overline{\psi(x,\xi)}\mathrm{d}x = \int_X f(x)\overline{\psi(g \cdot x, \xi)}\mathrm{d}x = \int_X f(x)\overline{\psi(x, g^{-1} \cdot \xi)}\mathrm{d}x = \widehat{\pi}_g[\mathtt{R}_\psi[f]](\xi)$$

$$\mathtt{NN}_\phi[\widehat{\pi}_g[\gamma]](x) = \int_\Xi \gamma(g^{-1} \cdot \xi)\phi(x,\xi)\mathrm{d}\xi = \int_\Xi \gamma(\xi)\phi(x, g \cdot \xi)\mathrm{d}\xi = \int_\Xi \gamma(\xi)\phi(g^{-1} \cdot x, \xi)\mathrm{d}\xi = \pi_g[\mathtt{NN}_\phi[\gamma]](x).$$

As a consequence, $\mathtt{NN}_\phi \circ \mathtt{R}_\psi : L^2(X) \to L^2(X)$ commutes with π, namely

$$\mathtt{NN}_\phi \circ \mathtt{R}_\psi \circ \pi_g = \mathtt{NN}_\phi \circ \widehat{\pi}_g \circ \mathtt{R}_\psi = \pi_g \circ \mathtt{NN}_\phi \circ \mathtt{R}_\psi$$

for all $g \in G$. Hence Schur's lemma yields that there exist a constant $C_{\phi,\psi} \in \mathbb{C}$ such that $\mathtt{NN}_\phi \circ \mathtt{R}_\psi = C_{\phi,\psi} \operatorname{Id}_{L^2(X)}$. By the construction of left-hand side, $C_{\phi,\psi}$ is bilinear in ϕ and ψ. □

Example

Original Ridgelet Transform

- G is the affine group $\mathrm{Aff}(m) = GL(m) \ltimes \mathbb{R}^m$
- $X = \mathbb{R}^m$ (data domain) with G-action

$$g \cdot \boldsymbol{x} := L\boldsymbol{x} + \boldsymbol{t}, \quad g = (L, \boldsymbol{t}) \in G,$$

- $\Xi = \mathbb{R}^m \times \mathbb{R}$ (parameter domain) with dual G-action

$$g \cdot (\boldsymbol{a}, b) = (L^{-\top}\boldsymbol{a}, b + \boldsymbol{t}^\top L^{-\top}\boldsymbol{a}), \quad g = (L, \boldsymbol{t}) \in G$$

- Then, $\phi(\boldsymbol{x}, (\boldsymbol{a}, b)) := \sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b)$ and $\psi(\boldsymbol{x}, (\boldsymbol{a}, b)) := \rho(\boldsymbol{a} \cdot \boldsymbol{x} - b)$ are joint G-invariant. Indeed,

$$\phi(g \cdot \boldsymbol{x}, g \cdot (\boldsymbol{a}, b)) = \sigma\left(L^{-\top}\boldsymbol{a} \cdot (L\boldsymbol{x} + \boldsymbol{t}) - (b + \boldsymbol{t}^\top L^{-\top}\boldsymbol{a})\right) = \sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b) = \phi(\boldsymbol{x}, (\boldsymbol{a}, b))$$

- Thus, we may put

$$\mathtt{NN}[\gamma](\boldsymbol{x}) = \int_{\mathbb{R}^m \times \mathbb{R}} \gamma(\boldsymbol{a}, b)\sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b)\mathrm{d}\boldsymbol{a}\mathrm{d}b, \quad \text{and} \quad \mathtt{R}[f](\boldsymbol{a}, b) = \int_{\mathbb{R}^m} f(\boldsymbol{x})\overline{\rho(\boldsymbol{a} \cdot \boldsymbol{x} - b)}\mathrm{d}\boldsymbol{x}$$

- Since π_g is known to be irreducible, $\mathtt{NN} \circ \mathtt{R} = (\!(\sigma, \rho)\!)\, \mathrm{Id}_{L^2(\mathbb{R}^m)}$

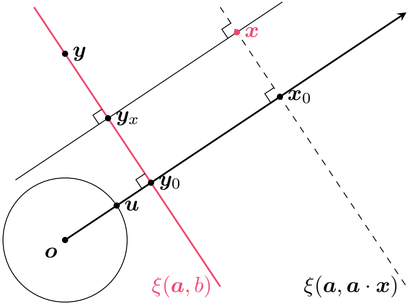
Geometric Interpretation

- Let
$$\xi(\boldsymbol{a}, b) := \{\boldsymbol{y} \in \mathbb{R}^m \mid \boldsymbol{a} \cdot \boldsymbol{x} - b = 0\}$$
be the hyperplane determined by $\boldsymbol{a} \cdot \boldsymbol{x} - b = 0$

- The invariant $\phi(\boldsymbol{x}, (\boldsymbol{a}, b)) = \sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b)$ is a signed distance between point $\boldsymbol{x}$ and hyperplane $\xi(\boldsymbol{a}, b)$ followed by scale $|\boldsymbol{a}|$ and nonlinearity σ:
$$\sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b) = \sigma(|\boldsymbol{a}| d(\boldsymbol{x}, \xi(\boldsymbol{a}, b)))$$

- While G-action $g \cdot \boldsymbol{x}$ is cast as the transportation of point $\boldsymbol{x}$,

- the dual G-action $g \cdot (\boldsymbol{a}, b)$ is cast as the parallel transport of hyperplane $\xi(\boldsymbol{a}, b)$ so that
$$d(g \cdot \boldsymbol{x}, g \cdot \xi(\boldsymbol{a}, b)) = d(\boldsymbol{x}, \xi(\boldsymbol{a}, b))$$



Deep Ridgelet Transform for Formal Deep Network[5]

[5] S-Hashimoto-Ishikawa-Ikeda, NeurReps2023

Formal DNN

Idea 1. Identify each hidden map $g : X \to X$ with a group action (of group G on X)
$\implies g$ is invertible (without specific implemenetation)

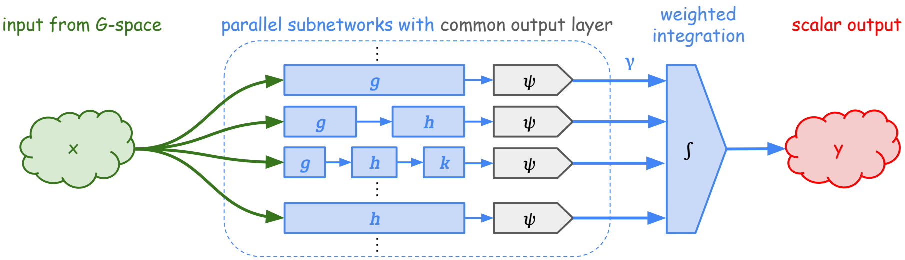
Idea 2. Integrate all the candidate subnetworks $\{\psi \circ g \mid g \in G\}$ (linearization trick!)

Idea 3. Then, $\theta(x, g) := \psi \circ g(x)$ is joint G-invariant!

Definition (Formal DNN)

For any function $\psi \in L^2(X)$ and measure γ on G, put

$$\mathtt{DNN}[\gamma; \psi](x) := \int_G \gamma(g)\, \psi \circ g(x) \mathrm{d}g, \quad x \in X.$$



Main Theorem

Definition (Deep Ridgelet Transform)

For any functions f and ϕ on X, put

$$R[f;\phi](g) := \int_X f(x)\, \overline{\phi \circ g(x)} \mathrm{d}x.$$

Theorem (Reconstruction Formula)

Provided $\pi_g[f] := f \circ g$ is an irreducible unitary representation of G acting from right on $L^2(X)$, there exists a bilinear form $((\psi, \phi))$ such that

$$\mathtt{DNN}_\psi[R_\phi[f]] = \int_G R_\phi[f](g)\, \psi \circ g(x) \mathrm{d}g = ((\psi, \phi)) f, \quad \text{for all } f \in L^2(X).$$

Proof

- Put the dual G-action on the parameter domain G as $g \cdot \xi := \xi \circ g^{-1}$
- Then, $\theta(x, \xi) := \psi \circ \xi(x)$ is joint G-invariant. Indeed,

$$\theta(g \cdot x, g \cdot \xi) = \psi \circ (g \cdot \xi)(g \cdot x) = \psi \circ (\xi \circ g^{-1})(g(x)) = \psi \circ \xi(x) = \theta(x, \xi)$$

- Thus, formal DNN and deep ridgelet transform are generalized DNN and generalized ridgelet transform, resp.