Deep Ridgelet Transform and Unified Universality Theorem for Deep and Shallow Joint-Group-Equivariant Machines

Sho Sonoda [1,2] Yuka Hashimoto [3,1] Isao Ishikawa [4,1] Masahiro Ikeda [5,1]

[1]RIKEN AIP [2]CyberAgent, Inc. [3]NTT Corporation [4]Kyoto University [5]The University of Osaka

ICML International Conference On Machine Learning

Motivation

- Universal approximation theorems are fundamental, but proofs are given case-by-case
- We introduce joint-group equivariance + ridgelet transform to unify the proofs
- The ridgelet transform provides closed-form parameter distributions $\gamma = \mathtt{R}[f]$
  $\Rightarrow$ solves the learning-equation

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

- Covers deep fully-connected networks, deep G-ConvNets, new quadratic nets, and more.

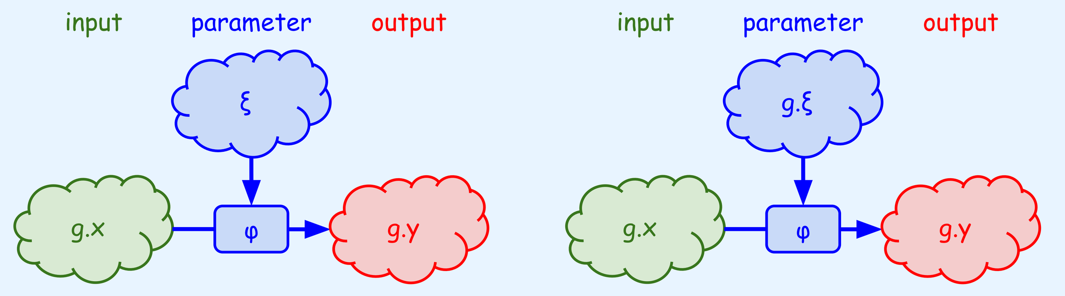
Main Results

Definition (Joint-G-Equivariant Map)

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

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

- In other words, ϕ is a G-map in $\hom_G(X \times \Xi, Y)$



G-equivariance joint-G-equivariance

Example (fully-connected network)

$$\phi(\boldsymbol{x}, (\boldsymbol{a}, b)) := \sigma(\boldsymbol{a} \cdot \boldsymbol{x} - b) \text{ is not } G\text{-equivariant but joint-}G\text{-equivariant}$$

- 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,

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

Definition (Joint-G-Equivariant Machine)

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

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

Example (integral representation of depth-2 fully-connected network)

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

Definition (Ridgelet Transform)

- Let $\psi : X \times \Xi \to Y$ be another joint-G-equivariant map
- For any map $f : X \to Y$, put

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

Main Theorem (Reconstruction Formula)

- Suppose that the representation $\pi : G \to \mathcal{U}(L^2(X; Y)), \pi_g[f](x) = g \cdot f(g^{-1} \cdot x)$ is irreducible
- Then, there exists a constant $c_{\phi,\psi}$ such that for any $f \in L^2(X; Y)$,

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

Proof

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

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

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

Therefore,

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

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

Lemmas

Construction of Joint-Equivariant Maps

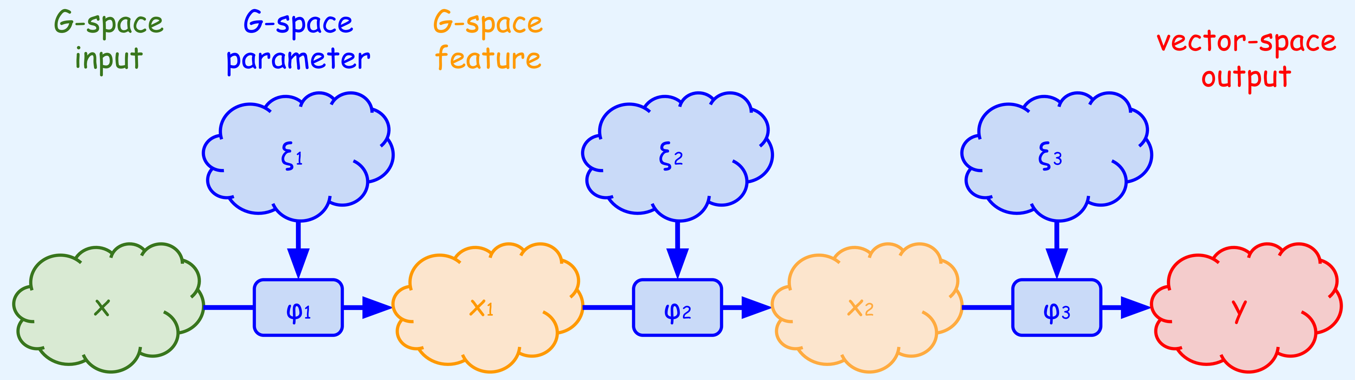
The following maps are joint-equivariant.

- Given any map $\phi_0 : X \to Y$, put $\phi : X \times G \to Y$ as

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

- Given any joint-equivariant maps $\phi_i : X_{i-1} \times \Xi_i \to X_i$, put $\phi_{1:n} : X_0 \times (\Xi_1 \times \cdots \times \Xi_n) \to X_n$ as

$$\phi_{1:n}(x, \xi_{1:n}) := \phi_n(\bullet, \xi_n) \circ \cdots \circ \phi_1(x, \xi_1)$$



- Let $\widehat{\pi}_g$ be the dual representation of G on parameter distributions $\gamma : \Xi \to \mathbb{C}$ defined by

$$\widehat{\pi}_g[\gamma](\xi) := \gamma(g^{-1} \cdot \xi).$$

Then, $\mathtt{M}$ is joint-equivariant because

$$\mathtt{M}[\widehat{\pi}_g[\gamma]; \phi] = \int_\Xi \gamma(g^{-1} \cdot \xi)\phi_c(\xi)\mathrm{d}\xi = \int_\Xi \gamma(\xi)\pi_g[\phi(\xi)]\mathrm{d}\xi = \pi_g[\mathtt{M}[\gamma; \phi]].$$

Examples

Example (Depth-n Fully-Connected Network)

- $G = O(m) \times \mathrm{Aff}(m)$ where $\mathrm{Aff}(m) = GL(m) \ltimes \mathbb{R}^m$
- For each $i \in \{1, \ldots, n\}$, put $X_i = \mathbb{R}^{d_i}$, and $X_1 = X_{n+1} = \mathbb{R}^m$ with G-action

$$g \cdot \boldsymbol{x} = L\boldsymbol{x} + \boldsymbol{t}, \quad g = (Q, L, \boldsymbol{t}) \in O(m) \times (GL(m) \ltimes \mathbb{R}^m)$$

- $\Xi_i = \mathbb{R}^{p_i \times d_i} \times \mathbb{R}^{p_i} \times \mathbb{R}^{d_{i+1} \times q_i}$ with dual G-action (not unique)

$$g \cdot (A_1, \boldsymbol{b}_1, C_n) := (A_1 L^{-1}, \boldsymbol{b}_1 + A_1 L^{-1}\boldsymbol{t}, QC_n), \quad g = (Q, L, \boldsymbol{t}) \in O(m) \times (GL(m) \ltimes \mathbb{R}^m)$$

- Then, the following $\phi : X_1 \times \Xi_{1:n} \to X_{n+1}$ is joint-equivariant:

$$\phi_{1:n}(\boldsymbol{x}, \boldsymbol{\xi}) := \phi_n(\bullet, \xi_n) \circ \cdots \circ \phi_1(\boldsymbol{x}, \xi_1), \quad \phi_i(\bullet, \xi_i) := C_i\sigma_i(A_i \bullet - \boldsymbol{b}_i),$$

- Put

$$\mathtt{DNN}[\gamma](\boldsymbol{x}) := \int_{\Xi_{1:n}} \gamma(\boldsymbol{\xi})\phi_{1:n}(\boldsymbol{x}, \boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}, \quad \mathtt{R}[\boldsymbol{f}](\xi) := \int_X \langle \boldsymbol{f}(\boldsymbol{x}), \psi_{1:n}(\boldsymbol{x}, \boldsymbol{\xi})\rangle \mathrm{d}\boldsymbol{x}$$

- (Lemma) *The unitary representation $\pi_g[\boldsymbol{f}](\boldsymbol{x}) := |\det L|^{-1/2} Q\boldsymbol{f}(L^{-1}(\boldsymbol{x} - \boldsymbol{t}))$ is irreducible*
- Therefore, we have $\mathtt{DNN} \circ \mathtt{R}[\boldsymbol{f}] = c_{\phi,\psi}\boldsymbol{f}$ for any $\boldsymbol{f} \in L^2(\mathbb{R}^m; \mathbb{R}^m)$

Example (Depth-n Group-Convolution Network)

- Let G be the primary group (for G-convolution), and let T be G-actions on X_i's
- Turn FC layers ϕ into (generalized) G-convolution layers ϕ^τ by:

$$\phi^\tau(x, \xi)(g) = T_g\left[C\sigma(AT_{g^{-1}}[x] - b)\right].$$

- The term $AT_g[x]$ covers the standard G-convolution, DeepSets, $E(n)$-nets, etc.
- For any $\gamma \in L^2(\Xi_{1:n})$ and G-equivariant map $f : X \to Y^G$ that is square-integrable at $1_G \in G$, put

$$\mathtt{GCNN}[\gamma](x)(g) := \int_{\Xi_{1:n}} \phi^\tau_{1:n}(x, \xi_{1:n})(g)\mathrm{d}\xi_{1:n}, \quad \mathtt{R}[f](\xi_{1:n}) := \int_{\mathbb{R}^m} \langle f(x)(1_G), \psi_{1:n}(x, \xi_{1:n})\rangle_Y \mathrm{d}x.$$

Then

$$\mathtt{GCNN} \circ \mathtt{R}[f] = (\!(\phi_{1:n}, \psi_{1:n})\!) f.$$

Example (An Unknown Network with Quadratic-Form)

- A new network whose universality was not known.
- $G = \mathrm{Aff}(m)$
- $X = \mathbb{R}^m$ (data domain) with G-action

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

- M = the class of m-dim. symmetric matrices with Lebesgue measure $\mathrm{d}A = \bigwedge_{i \geq j} \mathrm{d}a_{ij}$.
- $\Xi = M \times \mathbb{R}^m \times \mathbb{R}$ with dual G-action

$$g \cdot (A, \boldsymbol{b}, c) := (L^{-\top}AL^{-1}, L^{-\top}\boldsymbol{b} - 2L^{-\top}AL^{-1}\boldsymbol{t}, c + \boldsymbol{t}^\top L^{-\top}AL^{-1}\boldsymbol{t} - \boldsymbol{t}^\top L^{-\top}\boldsymbol{b}).$$

- Then, for any function $\sigma : \mathbb{R} \to \mathbb{R}$, the following map ϕ is joint-equivariant:

$$\phi(\boldsymbol{x}, \xi) := \sigma(\boldsymbol{x}^\top A\boldsymbol{x} + \boldsymbol{x}^\top\boldsymbol{b} + c)$$

- Hence the following network is $L^2(\mathbb{R}^m)$-universal.

$$\mathtt{QNN}[\gamma](\boldsymbol{x}) := \int_{M \times \mathbb{R}^m \times \mathbb{R}} \gamma(A, \boldsymbol{b}, c)\sigma(\boldsymbol{x}^\top A\boldsymbol{x} + \boldsymbol{x}^\top\boldsymbol{b} + c)\mathrm{d}A\mathrm{d}\boldsymbol{b}\mathrm{d}c.$$

Summary

- Joint-group equivariance unifies shallow & deep universality
- Schur's lemma + irreducibility give a *one-step* proof
- Ridgelet transform produces closed-form parameter distributions
- Applicability: FCNs, G-CNNs, quadratic nets, and beyond