E(n) Equivariant Models and E(n) Equivariant Steerable Models

Shih-Hsin Wang

Bwang-Team
Scientific Computing and Imaging Institute
University of Utah

Bi-weekly Meeting, Dec. 09, 2022

Outline

Outline

Recall

- The Euclidean group $E(n) := \{\text{isometries of the Euclidean space } \mathbb{R}^n\}$
- $E(n)$ is parametrized by $(\boldsymbol{Q}, \boldsymbol{t})$ where $\boldsymbol{Q} \in O(n), \boldsymbol{t} \in \mathbb{R}^n$ and the associated isometry on $\mathbb{R}^n$ is $\boldsymbol{x} \in \mathbb{R}^n \mapsto \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{t}$
- For any spatial embedding $\boldsymbol{X} \in \mathbb{R}^{n \times m}$, the isometry $(\boldsymbol{Q}, \boldsymbol{t})$ can transform the column vectors of $\boldsymbol{X}$ simultaneously by

$$\boldsymbol{X} \to \boldsymbol{Q}\boldsymbol{X} + \boldsymbol{t}\mathbb{1}_m^\top$$

  where $\mathbb{1}_m \in \mathbb{R}^{m \times 1}$ is a vector of ones.
- The Euclidean distance matrix $\boldsymbol{D}(\boldsymbol{X})$ of $\boldsymbol{X}$ is defined by

$$\boldsymbol{D}(\boldsymbol{X})_{ij} = d(\boldsymbol{x}_i, \boldsymbol{x}_j)^2 = ||\boldsymbol{x}_i - \boldsymbol{x}_j||^2$$

  where $\boldsymbol{x}_i$ is the $i-$th column vector of $\boldsymbol{X}$.

Recall: Euclidean distance matrix

- Let $\boldsymbol{S}$ be the operator sending $\boldsymbol{X}$ to $\boldsymbol{X}^\top \boldsymbol{X}$.
- Let $\mathbb{D}$ be the operator defined by

$$\mathbb{D}(\boldsymbol{S}) = \delta(\boldsymbol{S})\mathbb{1}^\top - 2\boldsymbol{S} + \mathbb{1}\delta(\boldsymbol{S})^\top$$

  where $\boldsymbol{S}$ is any symmetric matrix and is a column vector of the diagonal elements of $\boldsymbol{S}$. Then we see that $\boldsymbol{D}(\boldsymbol{X}) = \mathbb{D}(\boldsymbol{S}(\boldsymbol{X}))$.
- For any vector $\boldsymbol{b} \in \mathbb{R}^m$, we define $\mathcal{M}_{\boldsymbol{b}} := \{\boldsymbol{X} \in \mathbb{R}^{n\times m} | \boldsymbol{X}\boldsymbol{b} = 0\}$.

Proposition

The restriction $\mathbb{D}_{\boldsymbol{b}} : \boldsymbol{S}(\mathcal{M}_{\boldsymbol{b}}) \to \boldsymbol{D}(\mathcal{M}_{\boldsymbol{b}})$ of the operator $\mathbb{D}$ on $\boldsymbol{S}(\mathcal{M}_{\boldsymbol{b}})$ is bijective.

Group Representation

Definition

A representation of a group G on a vector space V is a map $\rho : G \to GL(V)$ such that

$$\rho(g \cdot h) = \rho(g)\rho(h) \text{ for any } g, h \in G$$

In particular, we say that ρ is a trivial representation if ρ sends all the elements of G to the identity mapping of V.

- For example, $\rho : E(n) \to GL(n+1)$ defined by

$$(Q, g = (g_1, g_2, \cdots, g_n)) \mapsto \begin{bmatrix} 1 & 0 & \cdots & 0 & g_1 \\ 0 & 1 & \cdots & 0 & g_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & g_n \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} Q & 0 \\ 0 & 1 \end{bmatrix}$$

is a representation of E(n) on $\mathbb{R}^n$

Equivariance and Invariance

Definition

Let $\rho_V : G \to GL(V)$ and $\rho_W : G \to GL(W)$ be the representations of G on vector spaces V and W, respectively.
A (nonlinear) function $\phi : V \to W$ is said to be **equivariant** if

$$\phi(\rho_V(g)(x)) = \rho_W(g)(\phi(x)) \text{ for any } g \in G, x \in X,$$

that is, we have the following commutative diagram for any $g \in G$

$$\begin{array}{ccc} V & \xrightarrow{\rho_V(g)} & V \\ \downarrow \phi & & \downarrow \phi \\ W & \xrightarrow{\rho_W(g)} & W \end{array}$$

In particular, we say that ϕ is **invariant** when ρ_W is a trivial group action.

Definition: $E(n)$ Equivariant Functions

A function $f : \mathbb{R}^{n\times m} \to \mathbb{R}^{n\times m'}$ is said to be $E(n)$ equivariant, if for any orthogonal matrix $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ and any vector $\boldsymbol{t} \in \mathbb{R}^n$, we have

$$f(\boldsymbol{QX} + \boldsymbol{t}\mathbb{1}_m^\top) = \boldsymbol{Q}f(\boldsymbol{X}) + \boldsymbol{t}\mathbb{1}_{m'}^\top$$

where $\boldsymbol{X}$ is an arbitrary matrix in $\mathbb{R}^{n\times m}$

Definition: $E(n)$ Invariant Functions

A function $g : \mathbb{R}^{n\times m} \to \mathbb{R}^{n\times m'}$ is said to be $E(n)$ invariant, if for any orthogonal matrix $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ and any vector $\boldsymbol{t} \in \mathbb{R}^n$, we have

$$g(\boldsymbol{QX} + \boldsymbol{t}\mathbb{1}_m^\top) = g(\boldsymbol{X})$$

where $\boldsymbol{X}$ is an arbitrary matrix in $\mathbb{R}^{n\times m}$

Orthogonally Equivariant and Invariant Functions

Fundamental theorem of invariant theory for orthogonal groups

If $g(\boldsymbol{X})$ is an orthogonal invariant function, it can be written as a function of $\boldsymbol{X}^\top \boldsymbol{X}$.

Proposition [Villar et al., 2021] or [Ma and Ying, 2022]

For any orthogonal equivariant function f, i.e., $f(\boldsymbol{QX}) = \boldsymbol{Q}f(\boldsymbol{X})$, there is an orthogonal invariant function g s.t.

$$f(\boldsymbol{X}) = \boldsymbol{X}g(\boldsymbol{X})$$

In particular, each column vector of $f(\boldsymbol{X})$ is a linear combination of the column vectors of $\boldsymbol{X}$ where $g(\boldsymbol{X})$ gives the coefficient.

Orthogonally Equivariant and Invariant Functions

Proof.

Assume $f(\boldsymbol{X}) \in \mathbb{R}^{n\times 1}$ and let $V := \text{span}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m)$ where $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ are the column vectors of $\boldsymbol{X}$. Decompose $f(\boldsymbol{X})$ into the sum $\boldsymbol{v} + \boldsymbol{u}$ where $\boldsymbol{v} \in V$ and $\boldsymbol{u} \in V^{\perp}$ and consider the orthogonal matrix $\boldsymbol{Q_u} = \boldsymbol{I} - \frac{2}{||\boldsymbol{u}||^2}\boldsymbol{u}\boldsymbol{u}^{\top}$. Clearly, $\boldsymbol{Q_u w} = \boldsymbol{w}$ for any $\boldsymbol{w} \in V$ and $\boldsymbol{Q_u u} = -\boldsymbol{u}$. So the equivariance implies that

$$f(\boldsymbol{Q_u X}) = \boldsymbol{Q_u} f(\boldsymbol{X}) = \boldsymbol{Q_u}(\boldsymbol{v} + \boldsymbol{u}) = \boldsymbol{v} - \boldsymbol{u}$$

On the other hand, since $\boldsymbol{Q_u}$ preserves every vector in V, we see that $\boldsymbol{Q_u X} = \boldsymbol{X}$ and then $f(\boldsymbol{Q_u X}) = f(\boldsymbol{X}) = \boldsymbol{v} + \boldsymbol{u}$ which forces $\boldsymbol{u} = 0$. Thus, $f(\boldsymbol{X}) = \boldsymbol{v}$ lies in V, i.e. we can write $f(\boldsymbol{X}) = \boldsymbol{X} g(\boldsymbol{X})$.
Moreover, since $\boldsymbol{QX} g(\boldsymbol{QX}) = f(\boldsymbol{QX}) = \boldsymbol{Q} f(\boldsymbol{X}) = \boldsymbol{QX} g(\boldsymbol{X})$, we have $\boldsymbol{X} g(\boldsymbol{QX}) = \boldsymbol{X} g(\boldsymbol{X})$ for any orthogonal matrix $\boldsymbol{Q}$. So one can choose $g(\boldsymbol{X})$ to be orthogonal invariant. □

Theorem for $E(n)$ Equivariant and Invariant Functions

Theorem

If g is a Euclidean invariant function, then g can be written as a function of the Euclidean matrix $\boldsymbol{D}(\boldsymbol{X})$.

Proposition [Villar et al., 2021]

For any $E(n)$ equivariant function $f : \mathbb{R}^{n\times m} \to \mathbb{R}^{n\times m'}$, there is a Euclidean invariant function $g : \mathbb{R}^{n\times m} \to \mathbb{R}^{m\times m'}$ s.t. $\mathbb{1}_m^T g(\boldsymbol{X}) = \mathbb{1}_{m'}^T$ and

$$f(\boldsymbol{X}) = \boldsymbol{X} g(\boldsymbol{X})$$

Necessary and sufficient condition

Corollary

$f : \mathbb{R}^{n\times m} \to \mathbb{R}^{n\times m'}$ is a $E(n)$ equivariant function if and only if there is a function $g : \mathbb{R}^{m\times m} \to \mathbb{R}^{m\times m'}$ s.t.

$$\mathbb{1}_m^T g(\boldsymbol{D}(\boldsymbol{X})) = \mathbb{1}_{m'}^T \text{ and } f(\boldsymbol{X}) = \boldsymbol{X} g(\boldsymbol{D}(\boldsymbol{X}))$$

Proof.

For any orthogonal matrix $\boldsymbol{Q}$ and vector $\boldsymbol{t} \in \mathbb{R}^n$, we have

$$\begin{aligned} f(\boldsymbol{QX} + \boldsymbol{t}\mathbb{1}_m^T) &= (\boldsymbol{QX} + \boldsymbol{t}\mathbb{1}_m^T) g(\boldsymbol{D}(\boldsymbol{QX} + \boldsymbol{t}\mathbb{1}_m^T)) \\ &= (\boldsymbol{QX} + \boldsymbol{t}\mathbb{1}_m^T) g(\boldsymbol{D}(\boldsymbol{X})) \\ &= \boldsymbol{QX} g(\boldsymbol{D}(\boldsymbol{X})) + \boldsymbol{t}\mathbb{1}_{m'}^T \\ &= \boldsymbol{Q} f(\boldsymbol{X}) + \boldsymbol{t}\mathbb{1}_{m'}^T \end{aligned}$$

Central Idea: Direct sum decomposition

- Let $\mathcal{M}_{\boldsymbol{e}_1}$ be the subset $\{\boldsymbol{X}' \in \mathbb{R}^{n\times m} | \boldsymbol{X}'\boldsymbol{e}_1 = 0\}$ consisting of all the matrices with zero first column vector.
- Define a map $\phi : \mathbb{R}^{n\times m} \to \mathbb{R}^n \oplus \mathcal{M}_{\boldsymbol{e}_1}$ as follows:

$$\phi(\boldsymbol{X}) = (\boldsymbol{X}\boldsymbol{e}_1) \oplus (\boldsymbol{X} - \boldsymbol{X}\boldsymbol{e}_1 \mathbb{1}_m^\top)$$

  where $\boldsymbol{e}_1 = [1, 0, \ldots, 0]^\top$ and $\boldsymbol{X}\boldsymbol{e}_1$ is the first column vector of $\boldsymbol{X}$.
- Define a compatible $E(n)$ action on $\mathbb{R}^n \oplus \mathcal{M}_{\boldsymbol{e}_1}$:

$$(\boldsymbol{Q}, \boldsymbol{t}) \cdot (\boldsymbol{x}, \boldsymbol{X}') = (\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{t}, \boldsymbol{Q}\boldsymbol{X}')$$

- Remark that this direct sum decomposition absorbs the effect of translation into its first component, so it becomes easier to study the impact of translation on functions.

Central Idea: Direct sum decomposition

Lemma

ϕ is a $E(n)$ equivariant bijection.

Proof.

One can check that $\phi^{-1}(\boldsymbol{x}, \boldsymbol{X}') = \boldsymbol{x}\mathbb{1}_m^\top + \boldsymbol{X}'$ is the inverse of ϕ. Moreover, for any orthogonal matrix $\boldsymbol{Q}$ and vector $\boldsymbol{t} \in \mathbb{R}^n$, we have

$$\begin{aligned}\phi(\boldsymbol{QX} + \boldsymbol{t}\mathbb{1}_m^\top) &= ((\boldsymbol{QX} + \boldsymbol{t}\mathbb{1}_m^\top)\boldsymbol{e}_1) \oplus (\boldsymbol{QX} + \boldsymbol{t}\mathbb{1}_m^\top - (\boldsymbol{QX} + \boldsymbol{t}\mathbb{1}_m^\top)\boldsymbol{e}_1\mathbb{1}_m^\top) \\ &= (\boldsymbol{QXe}_1 + \boldsymbol{t}) \oplus (\boldsymbol{Q}(\boldsymbol{X} - \boldsymbol{Xe}_1\mathbb{1}_m^\top)) \\ &= (\boldsymbol{Q}, \boldsymbol{t}) \cdot \phi(\boldsymbol{X})\end{aligned}$$

So we conclude that ϕ is a $E(n)$ equivariant bijection. □

Proof of our main theorem

We sketch our proof:

- ϕ allows us to consider $\boldsymbol{X} \cong \boldsymbol{x} \oplus \boldsymbol{X}'$ as the input of functions
- Note that $\boldsymbol{x}$ is translation equivariant and $\boldsymbol{X}'$ is translation invariant.
- This forces that a translation invariant function is independent of $\boldsymbol{x}$
- Then apply the orthogonal invariance (resp. equivariance) to deduce the desired form of invariant (resp. equivariant) functions.
- Also, the bijectivity of $\mathbb{D}_{\boldsymbol{b}} : \boldsymbol{S}(\mathcal{M}_{\boldsymbol{b}}) \to \boldsymbol{D}(\mathcal{M}_{\boldsymbol{b}})$ gives the correspondence between $\boldsymbol{X}'^{\top}\boldsymbol{X}'$ and $\boldsymbol{D}(\boldsymbol{X}')$
- Clearly, $\boldsymbol{D}(\boldsymbol{X}') = \boldsymbol{D}(\boldsymbol{X})$ since $\boldsymbol{X}' = \boldsymbol{X} - (\boldsymbol{X}\boldsymbol{e}_1)\mathbb{1}_m^{\top}$

Revisit: Equivariant Graph Neural Networks (EGNN)

Equivariant Graph Convolutional Layer [Satorras et al., 2021b]

$$
\begin{aligned}
\boldsymbol{m}_{ij} &= \phi_e(\boldsymbol{h}_i^l, \boldsymbol{h}_j^l, \|\boldsymbol{x}_i^l - \boldsymbol{x}_j^l\|^2, a_{ij}) \\
\boldsymbol{x}_i^{l+1} &= \boldsymbol{x}_i^l + C\sum_{j\neq i}(\boldsymbol{x}_i^l - \boldsymbol{x}_j^l)\phi_x(\boldsymbol{m}_{ij}) \\
\boldsymbol{m}_i &= \sum_{j\neq i}\boldsymbol{m}_{ij} \\
\boldsymbol{h}_i^{l+1} &= \phi_h(\boldsymbol{h}_i^l, \boldsymbol{m}_i)
\end{aligned}
\tag{1}
$$

- $\boldsymbol{x}_i^l \in \mathbb{R}^n$ is the coordinate embedding of node v_i at layer l
- C is chosen to be $1/(M-1)$ that computes the average of the sum
- $\phi_x : \mathbb{R}^{nf} \to \mathbb{R}$ is a learnable function (approximated by MLPs)

Discussion

- The updating scheme for $\boldsymbol{x}_i^l$ satisfies the condition given in our theorem.
- However, it will face an exploding problem (without a control of magnitude).
- A modification of EGNN has been done in [Satorras et al., 2021a] to make it stable

$$\boldsymbol{x}_i^{l+1} = \boldsymbol{x}_i^l + \sum_{j \neq i} \frac{(\boldsymbol{x}_i^l - \boldsymbol{x}_j^l)}{||\boldsymbol{x}_i^l - \boldsymbol{x}_j^l|| + C} \phi_x(\boldsymbol{m}_{ij})$$

  where they set C to be 1 (to ensure the differentiability)
- We should be able to design a better way to solve this problem.

Outline

Motivation

- THe message passing in EGNN

$$\boldsymbol{m}_{ij} = \phi_e(\boldsymbol{h}_i^l, \boldsymbol{h}_j^l, \|\boldsymbol{x}_i^l - \boldsymbol{x}_j^l\|^2, a_{ij})$$
$$\boldsymbol{m}_i = \sum_{j \neq i} \boldsymbol{m}_{ij}$$
$$\boldsymbol{h}_i^{l+1} = \phi_h(\boldsymbol{h}_i^l, \boldsymbol{m}_i)$$

- It sends invariant information $\boldsymbol{m}_{ij}$.

Question

Can we do message passing by sending **equivariant information**?

Answer

Design a message passing sending **steerable** messages.

Steerable Vectors

Definition

We say a vector space V is steerable for a group G if there is a representation $\rho : G \to GL(V)$ of G on V. Also, the vectors in V are called steerable vectors.

- We use Tilde $\tilde{}$ to denote steerable vectors.
- Steerability of a vector $\tilde{\boldsymbol{h}}$ means that for any $g \in G$, the vector is transformed by g via matrix multiplication $\rho(g)\tilde{\boldsymbol{h}}$.
- For example, a Euclidean vector in $\mathbb{R}^3$ is steerable for rotations $g = \boldsymbol{R} \in \boldsymbol{SO}(3)$ by multiplying the vector on right hand side.

Properties of Representations

Definition

Let $\rho : G \to GL(V)$ be a representation.

- ρ is said to be unitary if $\rho(g^{-1}) = \rho(g)^*$ for any $g \in G$
- A representation $\rho|_W$ of G on a vector subspace $W \subset V$ is said to be a subrepresentation if $\rho|_W(g) = \rho(g)|_W$.
- In particular, there is always a trivial subrepresentation given by the $G-$invariant subspace $V^G := \{v \in V | \rho(g)v = v \text{ for any } g \in G\}$.

Definition

A representation $\rho : G \to GL(V)$ is said to be irreducible if it has only trivial subrepresentations (given by V and 0). On the other hand, if V has a proper nontrivial $G-$invariant subspace, ρ is said to be reducible.

Completely Reducible Representations

Definition

A representation $\rho : G \to GL(V)$ is said to be completely reducible if it decomposes as a direct sum of irreducible subrepresentations.
In particular, there is an invertible matrix P s.t. for any $g \in G$, we have

$$P^{-1}\rho(g)P = \begin{pmatrix} \rho^{(1)}(g) & 0 & \cdots & 0 \\ 0 & \rho^{(2)}(g) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \rho^{(k)}(g) \end{pmatrix}$$
$$= \rho^{(1)}(g) \oplus \rho^{(2)}(g) \oplus \cdots \oplus \rho^{(k)}(g)$$

Theorem

Finite-dimensional unitary representations of any group are completely reducible.

Wigner-D matrices

- We have a well understanding of all the irreducible representations of $SO(3)$ (and hence $O(3) = SO(3) \times \{\pm \boldsymbol{I}\}$)
- Indeed, any irreducible representation of $SO(3)$ is characterized by l-th degree Wigner-D matrices $\boldsymbol{D}^{(l)}(g)$ for some $l \geq 0$.
- They define the only irreducible representation $\boldsymbol{D}^{(l)} : SO(3) \to GL(V)$ on a $(2l+1)$-dimensional vector space $V^{(l)}$.
- $V^{(l)}$ will be referred to as type-l steerable vector space.

Remark

- The trivial representation is given by the direct sum of copies of $\boldsymbol{D}^0$.
- The natural representation on $\mathbb{R}^3$ is given by $\boldsymbol{D}^{(1)}$

Spherical Harmonic

Question

How to convert a given vector into a steerable vector?

- Given a vector $\boldsymbol{x} \in \mathbb{R}^3$.
- Re-scale the vector to a vector $\frac{x}{\|x\|}$ on S^2
- Then convert it to a type-l steerable vector $\tilde{\boldsymbol{a}}^{(l)}$ through (real) spherical harmonics $Y_m^{(l)} : S^2 \to \mathbb{R}$ where $m = -l, \ldots, l$. That is,

$$\tilde{\boldsymbol{a}}^{(l)} = \left(Y_m^{(l)} \left(\frac{x}{\|x\|} \right) \right)_{m=-l,\ldots,l}^T$$

- This construction is equivariant since for any $g \in O(3)$, we have

$$D^{(l)}(g)\tilde{\boldsymbol{a}}^{(l)} = \left(Y_m^{(l)} \left(\frac{g \cdot x}{\|x\|} \right) \right)_{m=-l,\ldots,l}^T$$

Spherical Harmonic & Fourier Transform on the Sphere

- Spherical harmonics $\{\boldsymbol{Y}_m^{(l)} : S^2 \to \mathbb{R} | l \geq 0, -l \leq m \leq l\}$ form an orthonormal basis of $L^2_{\mathbb{R}}(S^2)$, the Hilbert space of square-integrable functions on the sphere.
- Indeed, for any function $f : S^2 \to \mathbb{R}$ in $L^2_{\mathbb{R}}(S^2)$, we have the expansion

$$f = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} a_m^{(l)} Y_m^{(l)}$$

where the coefficients are given by

$$a_m^{(l)} = \int_{S^2} f \; Y_m^{(l)} dS^2$$

- Here $\tilde{\boldsymbol{a}} = (a_m^{(l)})_{l \geq 0, -l \leq m \leq l}$ is a steerable vector in $V_L = \bigoplus_{l=0} V^{(l)}$

Spherical Harmonic & Fourier Transform on the Sphere

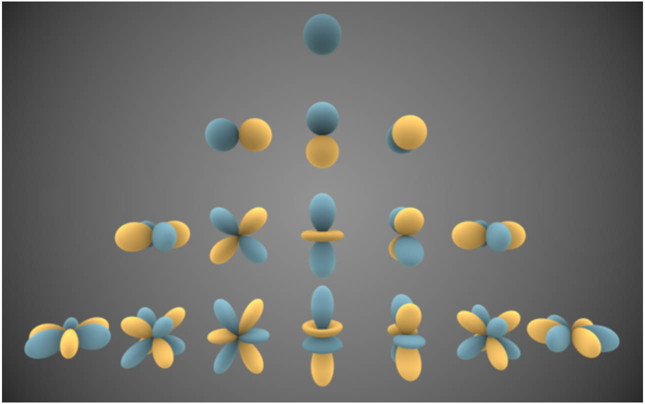

Figure: Visual representations of the first few real spherical harmonics. Blue portions represent regions where the function is positive, and yellow portions represent where it is negative. The distance of the surface from the origin indicates the absolute value of $Y_m^{(l)}(\theta, \varphi)$ in angular direction (θ, φ).

Clebsh-Gordan (CG) tensor product

The Clebsh-Gordan (CG) tensor product

$$\otimes_{cg}^{w} : V^{(l_1)} \times V^{(l_2)} \to V^{(l)}$$

is a bilinear operator between steerable vector spaces defined by

$$(\tilde{\boldsymbol{h}}^{(l_1)} \otimes_{cg}^{w} \tilde{\boldsymbol{h}}^{(l_2)})_m^{(l)} = w \sum_{m_1=-l_1}^{l_1} \sum_{m_2=-l_2}^{l_2} C_{(l_1,m_1)(l_2,m_2)}^{(l,m)} h_{m_1}^{(l_1)} h_{m_2}^{(l_2)},$$

- $\tilde{\boldsymbol{h}}^{(l)} \in V^{(l)} = \mathbb{R}^{2l+1}$ denote a steerable vector of type l and $h_m^{(l)}$ its components for $m = -l, \dots, l$
- w is a learnable parameter
- $C_{(l_1,m_1)(l_2,m_2)}^{(l,m)}$ are the Clebsh-Gordan coefficients (ensure the resulting vector is type-l steerable)
- This product is commonly sparse, as many coefficients are zero.

Clebsh-Gordan (CG) tensor product

- Note that it combines two steerable input vectors of types l_1 and l_2 and then returns a steerable vector of type l.
- There is no limitation on the choices of l_1, l_2, and l
- For a fix steerable vector $\tilde{\boldsymbol{a}}$ and a collection of learnable parameters $\boldsymbol{W}$, we define the **steerable linear layer conditioned on $\tilde{\boldsymbol{a}}$** to be

$$\boldsymbol{W}_{\tilde{\boldsymbol{a}}} \tilde{\boldsymbol{h}} = \tilde{\boldsymbol{h}} \otimes_{cg}^{\boldsymbol{W}} \tilde{\boldsymbol{a}}$$

  where $\tilde{\boldsymbol{h}}$ is a steerable vector.
- This steerable linear layer is equivariant, i.e.

$$(\boldsymbol{D}^{(l_1)}(g)\tilde{\boldsymbol{h}}) \otimes_{cg}^{\boldsymbol{W}} (\boldsymbol{D}^{(l_2)}(g)\tilde{\boldsymbol{a}}) = \boldsymbol{D}^{(l)}(g)(\tilde{\boldsymbol{h}} \otimes_{cg}^{\boldsymbol{W}} \tilde{\boldsymbol{a}})$$

- When l is chosen to be 0, $\boldsymbol{W}_{\tilde{\boldsymbol{a}}}$ defines a steerable linear layer mapping into $V^{(0)}$ and hence is an invariant function.

Steerable MLPs

- A steerable $n-$layers MLP is constructed as follows:

$$\widetilde{MLP}(\tilde{\boldsymbol{h}}) = \sigma(\boldsymbol{W}_{\tilde{\boldsymbol{a}}}^{(n)}(\ldots(\sigma(\boldsymbol{W}_{\tilde{\boldsymbol{a}}}^{(1)}\tilde{\boldsymbol{h}}))))$$

where σ is a steerable activation function to ensure the equivariance, i.e. for any $g \in O(3)$,

$$\boldsymbol{D}'(g)\widetilde{MLP}(\tilde{\boldsymbol{h}}) = \widetilde{MLP}(\boldsymbol{D}(g)\tilde{\boldsymbol{h}})$$

- Several classes of steerable activation functions are Fourier-based [Cohen et al., 2018], norm-altering [Thomas et al., 2018] and gated-non-linearities [Weiler et al., 2018].

Steerable Equivariant GNNs (SEGNNs)

The message passing equations are extended by considering an update to the steerable node features $\tilde{\boldsymbol{f}} \in V_L$ at node v_i via the following steps:

$$\tilde{\boldsymbol{m}} = \phi_m\left(\tilde{\boldsymbol{f}}_i, \tilde{\boldsymbol{f}}_j, \|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2, \tilde{\boldsymbol{a}}_{ij}\right),$$

$$\tilde{\boldsymbol{f}}_i' = \phi_f\left(\tilde{\boldsymbol{f}}_i, \sum_{j\in\mathcal{N}(i)} \tilde{\boldsymbol{m}}_{ij}, \tilde{\boldsymbol{a}}_i\right).$$

- $\|\boldsymbol{x}_i - \boldsymbol{x}_j\|$ is the relative distance between $i-$th node and $j-$th node
- ϕ_m and ϕ_f are $\boldsymbol{O}(3)$ steerable MLPs
- The steerable edge attribute $\tilde{\boldsymbol{a}}_{ij} \in V_L$ is obtained by converting $\boldsymbol{x}_i - \boldsymbol{x}_j$ into steerable vectors of different types and then concatenating them together
- $\tilde{\boldsymbol{a}}_i = \sum_{j\in\mathcal{N}(i)} \tilde{\boldsymbol{a}}_{ij}$ is the node attribute in V_L.

Discussion

- We see that steerable models propagate steerable messages
- However, the data is acted by the trivial representation or the natural representation; we should be able to analyze steerable models from the perspective of general forms.
- For example, a steerable model predicting an invariant output should be able to write as a function of relative distances.

Reference

Cohen, T. S., Geiger, M., Köhler, J., and Welling, M. (2018).
Spherical CNNs.
In *International Conference on Learning Representations*.

Ma, C. and Ying, L. (2022).
Why self-attention is natural for sequence-to-sequence problems? a perspective from symmetries.
arXiv preprint arXiv:2210.06741.

Satorras, V. G., Hoogeboom, E., Fuchs, F. B., Posner, I., and Welling, M. (2021a).
E (n) equivariant normalizing flows.
arXiv preprint arXiv:2105.09016.

Satorras, V. G., Hoogeboom, E., and Welling, M. (2021b).
E (n) equivariant graph neural networks.
In *International conference on machine learning*, pages 9323–9332. PMLR.

Reference

Thomas, N., Smidt, T., Kearnes, S., Yang, L., Li, L., Kohlhoff, K., and Riley, P. (2018).
Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds.

Villar, S., Hogg, D. W., Storey-Fisher, K., Yao, W., and Blum-Smith, B. (2021).
Scalars are universal: Equivariant machine learning, structured like classical physics.
Advances in Neural Information Processing Systems, 34:28848–28863.

Weiler, M., Geiger, M., Welling, M., Boomsma, W., and Cohen, T. S. (2018).
3d steerable cnns: Learning rotationally equivariant features in volumetric data.
In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc.