Stony Brook University

Equivariant GNNs

Wenhan Gao

Ph.D. student at Stony Brook, supervised by Prof. Yi Liu

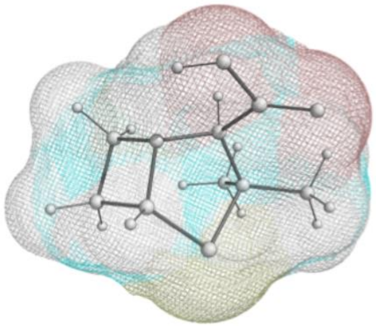
Department of Applied Mathematics and Statistics

Department of Computer Science

FAR BEYOND

Introduction

- Group Conv is an equivariant framework popular for **"functional data"** (images), where the data can be treated as "***functions***“; transformations **act on the domain of these functions, which is not the implicit input to the network**.
- For "***geometric data***", e.g., point clouds or molecular structures that can be viewed as a set of "points/nodes" with coordinate information; transformations **act on the input coordinates directly.**
- **Geometric GNNs are often employed to process geometric data.**



The pyridine molecule can be described in the XYZ format by the following:

```
11

C       -0.180226841      0.360945118     -1.120304970
C       -0.180226841      1.559292118     -0.407860970
C       -0.180226841      1.503191118      0.986935030
N       -0.180226841      0.360945118      1.29018350
C       -0.180226841     -0.781300882      0.986935030
C       -0.180226841     -0.837401882     -0.407860970
H       -0.180226841      0.360945118     -2.206546970
H       -0.180226841      2.517950118     -0.917077970
H       -0.180226841      2.421289118      1.572099030
H       -0.180226841     -1.699398882      1.572099030
H       -0.180226841     -1.796059882     -0.917077970
```

FAR BEYOND

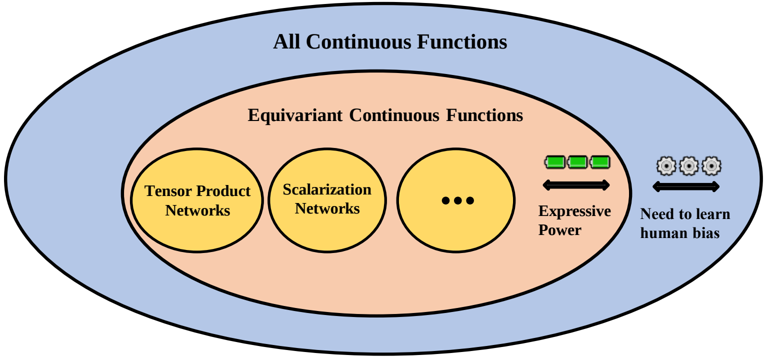
See previous notes on group conv and (constrained) geometric GNNs

Unconstrained Geometric GNNs

(Constrained) Geometric GNNs enforce symmetries directly into the architecture.

- **Restricting its set of possible operations or representations**.
- **Hindering network capacity to fully express the intricacies of the data**.
- **Computational inefficient**.
- For large pretrained models, such as GPT, **we cannot alter their network designs to ensure equivariance**!

Does enforcing equivariance/symmetries as an inductive bias truly offset a potential reduction in optimization diversity within the constrained learning spaces?



- Alternative Approach: Group/Frame Averaging

Caveat: GA/FA are not magic; they have their own issues, such as breaking the continuity of the underlying function.

Stony Brook University

Definition: Group Averaging

Consider an arbitrary $\Phi : X \to Y$.

- X, Y: Input and output (vector) spaces

The GA operator $\langle\Phi\rangle_G : X \to Y$ is defined as:

$$\langle\Phi\rangle_G(x) = \mathbb{E}_{g\sim\nu}\rho_2(g) \cdot \Phi\left(\rho_1(g)^{-1} \cdot x\right) = \int_G \rho_2(g) \cdot \Phi\left(\rho_1(g)^{-1} \cdot x\right) d\nu(x)$$

or in summation form for discrete groups:

$$\langle\Phi\rangle_G(x) = \frac{1}{|G|}\sum_{g\in G}\rho_2(g) \cdot \Phi\left(\rho_1(g)^{-1} \cdot x\right)$$

- $\rho_1(g), \rho_2(g)$: Group representations on X and Y respectively.
- ν: Harr measure over G ("uniform" over G)

Issue: When $|G|$ is large (e.g., combinatorial groups such as permutations) or infinite (e.g., continuous groups such as rotations), then exact averaging is intractable.

FAR BEYOND

Intuition: Group Averaging

The GA operator is equivariant to G. Proof:

$$\begin{aligned}\langle\Phi\rangle_G(h\cdot x) &= \mathbb{E}_{g\sim\nu}\rho_2(g)\cdot\Phi\left(\rho_1(g)^{-1}\cdot(\rho_1(h)\cdot x)\right)\\ &= \mathbb{E}_{g\sim\nu}\rho_2(g)\cdot\Phi\left(\rho_1\left(h^{-1}g\right)^{-1}\cdot x\right)\\ &= \rho_2(h)\mathbb{E}_{g\sim\nu}\rho_2\left(h^{-1}g\right)\cdot\Phi\left(\rho_1\left(h^{-1}g\right)^{-1}\cdot x\right)\\ &= \rho_2(h)\langle\Phi\rangle_G(x)\end{aligned}$$

Intuition:

- Similar to group convolutions, we have already calculated all the transformed versions of the input ($\rho_1(g)^{-1}$).
- $\rho_2(g)$ "corrects" the output for equivariance.
  - $\{\Phi(\rho_1(g)^{-1}\cdot x), \forall g\}$ will result in the same set of outputs, but in a different order, for transformed inputs.
  - Why? Because the set of inputs $\{\rho_1(g)^{-1}\cdot x\}$ is the same but in a different order for a transformed x.
  - Thus, integrating/summing over these outputs will result in invariant outputs.
  - $\rho_2(g)$ "corrects" the output by applying the transformation back.

Expressive Power: Group Averaging

GA is as expressive as its backbone Φ when Φ is equivariant to G. To see this:

$$\begin{aligned}\langle\Phi\rangle_G(x) &= \mathbb{E}_{g\sim\nu} g \cdot \Phi\left(g^{-1}\cdot x\right)\\ &= \mathbb{E}_{g\sim\nu} g \cdot g^{-1} \cdot \Phi(x)\\ &= \Phi(x)\end{aligned}$$

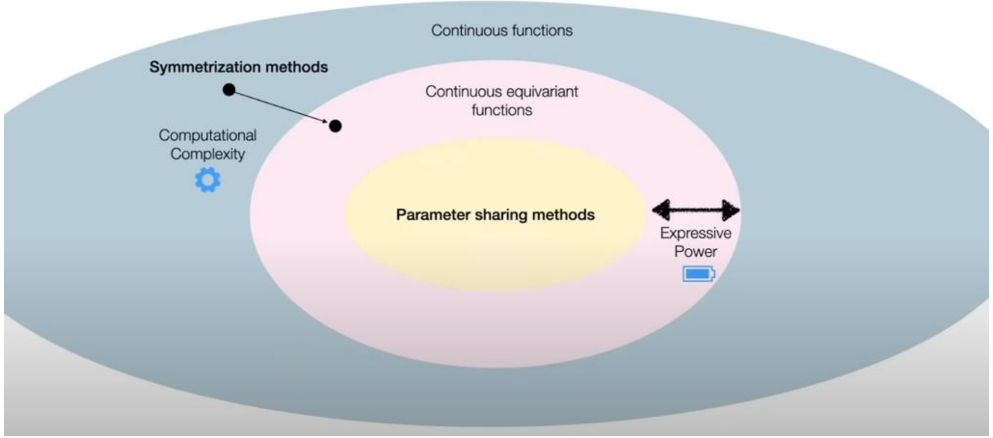
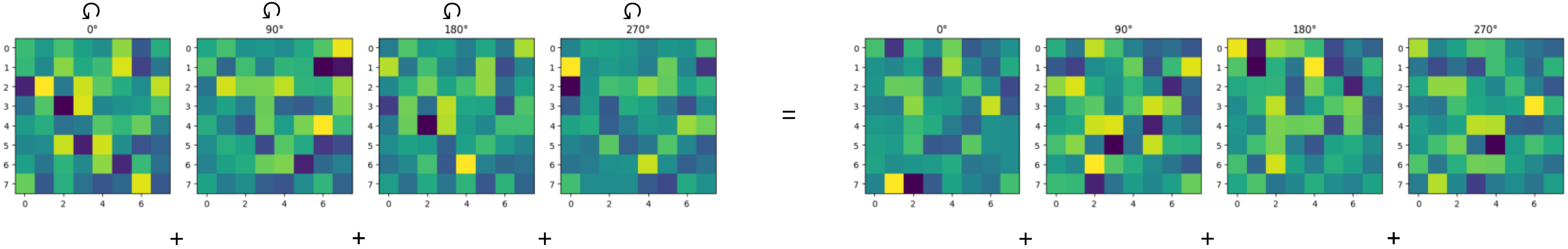


Image Credit: Frame Averaging for Invariant and Equivariant Network Design. Omri Puny et al.

Connections between GA and G-Conv

Recall that in Group Conv, a rotation to the input will result in planar rotation + periodic shifting for the output feature map. **If we average over the feature maps to get a single feature map, the result is still equivariant to rotations, i.e., a rotation in the input will result in a rotation in the output feature map.**



Recall the definition of lifting layer: $[f \star k](g) = \int_{\mathbb{R}^d} f(y) k\left(g^{-1} y\right) dy = (L_g \cdot k) \star f.$

Make cross-correlation to be the backbone for group averaging:

Sum over all feature maps

$$\langle \Phi \rangle_G(f) = \mathbb{E}_{g \sim \nu} L_g \cdot \left[k \star \left(L_g^{-1} \cdot f\right)\right] = \frac{1}{|G|} \sum_{g \in G} L_g \cdot \left[k \star \left(L_g^{-1} \cdot f\right)\right] = \frac{1}{|G|} \sum_{g \in G} (L_g \cdot k) \star f.$$

A feature map after lifting

Here, we have the "correction term", but in group convolutions, we do not have the "correction term". This is the result of moving the rotation from the image to the kernel. We'll demonstrate this in the next slide.

Connections between GA and G-Conv

- Rotation to kernel instead of image: $k \star L_g f = L_g \left(L_g^{-1} k \star f\right)$
- Intuition: Just distribute L_g inside (However, intuition is wrong here, mathematically this is incorrect, and for many representations, this is wrong, we can do this here because rotations are unitary matrices).
- Proof:

$$\begin{aligned}
&[(L_g k) \star (L_g f)](x) \\
&= \int_{\mathbb{R}^2} \mathrm{Lg}\, k(x-y)\, \mathrm{Lg}\, f(y) dy \\
&= \int_{\mathbb{R}^2} k(gx - gy) f(gy) dy \\
&= \int_{\mathbb{R}^2} k\left(gx - y'\right) f\left(y'\right) dy' \\
&= \mathrm{Lg}(k \star f)(x)
\end{aligned}$$

,

- First equality: Just definition of conv/cross-correlation
- Second equality: Again just distribute L_g
- Third equality: Change of variable $gy = y'$ since the determinant of g is 1.
- Fourth equality: Definition

FAR BEYOND

Motivating Example: Pre-processing

Consider image segmentation, in which we want translation equivariance. Assuming we are using a model other than CNN, so that we do not have inherent translation equivariance. What can we do if we want translation equivariance? We can use group averaging, but it will be computationally intractable for the translation group (large in the discrete case or even infinite if we view it in a continuous manner). We should think of another way of achieving equivariance.

One way is geometric pre-processig:

Given an image,

we can achieve equivariance by preprocessing the image.

For example, if we have the location of the left eye of a cat, we can preprocess the image such that all cats will have their left eyes in the same location.

Definition: Frame Averaging

A frame is defined as a set valued function $\mathscr{F}: X \to 2^G$. (Taking an input $x \in X$ and mapping to a subset of G)

A frame is equivariant to G if $\mathscr{F}(g \cdot x) = g\mathscr{F}(x), \forall g \in G$. (equality of sets)

Frame Averaging operator is defined as:

$$\langle \Phi \rangle_{\mathcal{F}}(x) = \frac{1}{|\mathcal{F}(x)|} \sum_{g \in \mathcal{F}(x)} \rho_2(g) \Phi\left(\rho_1(g)^{-1} x\right)$$

Similar to group averaging, we can prove that frame averaging operator is equivariant to G if $\mathcal{F}$ is equivariant to G, and it is as expressive as its backbone if its backbone is equivariant.

Example:

Consider $X = \mathbb{R}^n, Y = \mathbb{R}^{\mathrm{n}}$, and $G = \mathbb{R}$ with addition as the group action. We choose the group actions in this case to be $\rho_1(t)\boldsymbol{x} = \boldsymbol{x} + t\mathbf{1}$, and $\rho_2(a)y = y + t$, where $t \in G, \boldsymbol{x} \in X, y \in Y$, and $\mathbf{1} \in \mathbb{R}^n$ is the vector of all ones.

We can define the frame in this case using the averaging operator $\mathcal{F}(\boldsymbol{x}) = \left\{\frac{1}{n}\mathbf{1}^T\boldsymbol{x}\right\} \subset G = \mathbb{R}$.

Note that in this case the frame contains only one element from the group, in other cases finding such a small frame is hard or even impossible.

One can check that this frame is equivariant. The FA: $\langle \Phi \rangle_{\mathcal{F}}(\boldsymbol{x}) = \Phi\left(\boldsymbol{x} - \frac{1}{n}\left(\mathbf{1}^T x\right)\mathbf{1}\right) + \frac{1}{n}\mathbf{1}^T x$ in the equivariant case.

- Intuition: Geometric pre-processing, we subtract the average and then add the average back to obtain equivariance.

Frame Averaging: Practical Usage in Geometric GNNs

Let $\Phi : V \to W$ be an arbitrary function, where V, W are some vector spaces. The group averaging operator Ψ can be made equivariant by symmetrization, that is averaging over the group:

$$\Psi(X) = \frac{1}{|G|} \sum_{g \in G} g \cdot \Phi\left(g^{-1} \cdot X\right).$$

- Φ: Node update MLPs in GNN
- $X \in V$: Input features (coordinates in particular)
- W: Space of output node embeddings
- $G, |G|$: Croup and cardinatlity of the group, respectively

It can be shown that $\Psi : V \to W$ is equivariant w.r.t. G.

Or alternatively, if we can find a frame for the group, we can use frame averaging instead.

Frame Averaging: Practical Instantiation for E(3)

Goal:

- We would like to incorporate Euclidean symmetry to existing permutation invariant/equivaraint point cloud networks.

Settings:

- Input Space: $V = \mathbb{R}^{n\times d}$ (n nodes, each holding a d-dimensional vector as its location)
- Group: $G = E(d) = O(d) \times T(d)$, namely the group of Euclidean motions in $\mathbb{R}^d$ defined by rotations and reflections $O(d)$, and translations $T(d)$.
- Representation acting on $\boldsymbol{X} \in V$: $\rho_1(g)\boldsymbol{X} = \boldsymbol{X}\boldsymbol{R}^T + \mathbf{1}\boldsymbol{t}^T$, where $\boldsymbol{R} \in O(d)$, and $\boldsymbol{t} \in \mathbb{R}^d$. (Apply rotation and translation to every node)
- W, ρ_2 are defined similarly, unless we want invariance.

Frame Averaging: Practical Instantiation for E(3)

Frame $\mathcal{F}(\boldsymbol{X})$ is defined based on Principle Component Analysis (PCA), as follows:

- Let $\boldsymbol{t} = \frac{1}{n}\boldsymbol{X}^T\mathbf{1} \in \mathbb{R}^d$ be the centroid of $\boldsymbol{X}$
- $\boldsymbol{C} = \left(\boldsymbol{X} - \mathbf{1}\boldsymbol{t}^T\right)^T\left(\boldsymbol{X} - \mathbf{1}\boldsymbol{t}^T\right) \in \mathbb{R}^{d\times d}$ the covariance matrix computed after removing the centroid from $\boldsymbol{X}$. In the generic case the eigenvalues of $\boldsymbol{C}$ satisfy $\lambda_1 < \lambda_2 < \cdots < \lambda_d$.
- Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_d$ be the unit length corresponding eigenvectors.
- Then we define $\mathcal{F}(\boldsymbol{X}) = \{([\alpha_1\boldsymbol{v}_1, \ldots, \alpha_d\boldsymbol{v}_d], t) \mid \alpha_i \in \{-1, 1\}\} \subset E(d)$.
- $[\boldsymbol{v}_1, \ldots, \boldsymbol{v}_d]$ is a set of orthonormal vectors in $\mathbb{R}^d$, i.e., a basis of $\mathbb{R}^d$. Moreover, these vectors will "rotate" in the same way as the input.
- $\mathcal{F}(\boldsymbol{X})$ based on the covariance and centroid are $E(d)$ equivariant.

Example in 3D:

Image Credit: Frame Averaging for Invariant and Equivariant Network Design. Omri Puny et al.

Intuition: Frame Averaging

Overall, frame averaging achieves equivariance by:

1. Create a "coordinate system" (or a set of orthonormal basis in the general case) that will change accroding the group actions acting on the input.
   - Meaning that if one input is a transformed version of another, this coordinate system will be transformed in a predictable way!
2. Representing the input in terms of this "coordinate system".
   - Inputs will be the same after this step (if one is a transformed version of another), the group action is reflected on the coordinate system now!
3. Reconstruction by the "coordinate system".
   - Now the group action is reflected on the output!

$$\langle \Phi \rangle_{\mathcal{F}}(x) = \frac{1}{|\mathcal{F}(x)|} \sum_{g \in \mathcal{F}(x)} \rho_2(g) \Phi \left(\rho_1(g)^{-1} x \right)$$

Mapping the input to a set of "coordinate systems" that respects transformations on the input.

Just synchronize over all possible such "coordinate systems".

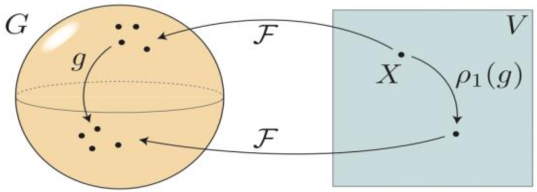


Image Credit: Frame Averaging for Invariant and Equivariant Network Design. Omri Puny et al.

Stony Brook University

FAR BEYOND