Point Group Equivariant Convolutional Graph Neural Networks

Alex Heilman

June 4, 2025

Overview

- Brief review of mathematics
- Group representation theory
  - Irreps
  - Basis Functions
  - Coupling Coefficients
- Equivariant networks
  - $SO(3)$ Equivariance
- Applications

Group Definition

A group G is a set of elements $\{g_1, ..., g_n\}$ with a binary operation $* : G \times G \rightarrow G$ between elements that satisfies the conditions of identity, associativity, invertability, and closure.

Example: General Linear Group
The general linear group $GL(V)$ formed over some vector space V is the set of non-singular $d_v \times d_v$ matrices acting on V with the group operation being matrix multiplication. The general linear group is itself a vector space.

Vector Space Definition

A vector space V over a field K is a group of vectors equipped with a distributive scalar multiplication. Vectors are often defined by way of a basis set that spans the space under scalar multiplication.

Example: $\mathbb{R}^3$, Real 3 Dimensional Space
Locations in physical space may be modeled with a three dimensional vector space $\mathbb{R}^3$ over the real numbers $\mathbb{R}$ with basis functions $\hat{x}, \hat{y}, \hat{z}$.

Example: Functions on Real 3 Dimensional Space
Scalar functions on physical space also form a vector space over the real numbers, albeit infinite-dimensional. In this case, the group operation between vectors (functions) is point-wise addition.

We may construct new vector spaces from sets of existing vector spaces by taking tensor products and direct sums.

Direct Sums

We may form direct sums $V \oplus W$ of vector spaces V, W by block-diagonal concatenation. This operation enjoys scalar distributivity to respective subspaces with the set of scalars $K_V \oplus K_W$.

Example: Direct Sum of Matrices

$$\begin{bmatrix} a_1 & b \\ c & a_2 \end{bmatrix} \oplus \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix} = \begin{bmatrix} a_1 & b & 0 & 0 \\ c & a_2 & 0 & 0 \\ 0 & 0 & d_1 & 0 \\ 0 & 0 & 0 & d_2 \end{bmatrix}$$

Direct sums of vector spaces are themselves vector spaces.

Tensor Products

Tensor products $V \otimes W$ of vector spaces V, W are uniquely bilinear so that $\lambda V \otimes W = V \otimes \lambda W = \lambda(V \otimes W)$.

Example: Tensor Product of Matrices

Also known as "Kronecker Product".

$$\begin{bmatrix} a_1 & b \\ c & a_2 \end{bmatrix} \otimes \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix} = \begin{bmatrix} a_1 d_1 & 0 & bd_1 & 0 \\ 0 & a_1 d_2 & 0 & bd_2 \\ cd_1 & 0 & a_2 d_1 & 0 \\ 0 & cd_2 & 0 & a_2 d_2 \end{bmatrix}$$

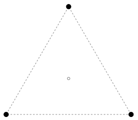
Tensor products of vector spaces are themselves vector spaces.

Group Representations

A representation ρ_G of a group G is a homomorphism from elements g to a set of linear operators (square matrices).

Example: 3D Representation of C_3

Consider three identical points in 3D space forming an equilateral triangle.



These clearly are symmetrical under three-fold rotations about the origin in the xy plane. These C_3 group actions act on this Cartesian basis with the representation ρ defined:

$$\rho(\mathbb{I}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \rho(C_3) = \begin{bmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \rho(C_3^2) = \begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} & 0 \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Irreducible Representations (IRs)

For atomic arrangements, 3D group representations ρ are often reducible in terms of a direct sum of 'smaller' group representations $\rho^{(\alpha)}$:

$$\rho = \bigoplus_{\alpha} c_\alpha \rho^{(\alpha)}$$

Maschke's theorem guarantees that any given representation is always decomposable as a direct sum of irreducible representations.

This set may always be taken to satisfy:

- Unitarity
- Orthogonality
- $\sum_\alpha d_\alpha^2 = N$ where d_α is the dimension of IR α and N is the order

IRs (cont.)

Example: IRs of D_3 The previously shown representation of C_3 elements is reducible into a two-dimensional subspace and a one-dimensional subspace of D_3.

$$\rho(\mathbb{I}) = \rho^{(2)}(\mathbb{I}) \oplus \rho^{(1)}(\mathbb{I}) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \oplus \begin{bmatrix} 1 \end{bmatrix}$$

$$\rho(C_3) = \rho^{(2)}(C_3) \oplus \rho^{(1)}(C_3) = \begin{bmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix} \oplus \begin{bmatrix} 1 \end{bmatrix}$$

$$\rho(C_3^2) = \rho^{(2)}(C_3^2) \oplus \rho^{(1)}(C_3^2) = \begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix} \oplus \begin{bmatrix} 1 \end{bmatrix}$$

Equivalence Classes

Equivalence classes are subsets of group elements that mutually exchange under conjugation, where element g conjugated by element h means:

$$g \to hgh^{-1}$$

In the space of a representation, this is referred to as a similarity transformation, which is essentially a change of basis.

Note that the number of equivalence classes N_c is equal to the number of irreducible representations.

$$N_{\text{IR}} = N_c$$

Character Tables

Irreducible representations are only unique up to change of basis, but their traces are invariant.

$$\chi^{(\alpha)}(g) = \mathrm{Tr}\big(\rho^{(\alpha)}(g)\big)$$

The trace of a representation is known as it's character χ, which is unique for equivalence classes $\langle g \rangle$.

Example: C_4 Character Table

C_4	$\langle \mathbb{I} \rangle$	$\langle C_4 \rangle$	$\langle C_4^2 \rangle$	$\langle C_4^3 \rangle$
a_1	1	1	1	1
a_2	1	-1	1	-1
a_3	1	i	-1	$-i$
a_4	1	$-i$	-1	i

Characters are often displayed in 'character tables', with IRs on one axis and equivalence classes along the other.

Orthogonality Theorems

IRs are orthogonal in the following ways:

$$\frac{1}{N}\sum_{g}\chi^{(\alpha)*}(g)\chi^{(\beta)}(g) = \delta_{\alpha\beta}$$

$$\sum_{\alpha}\chi^{(\alpha)*}(c_k)\chi^{(\alpha)}(c_h) = \frac{N}{N_k}\delta_{kh}$$

where N is the number of elements in G and N_k is the number of elements in equivalence class k.

This allows us to decompose reducible representations by determining the coeffecients of expansion c_α as:

$$c_\alpha = \frac{1}{N}\sum_{g}\chi^{(\alpha)*}(g)\chi(g)$$

Orthogonality Theorems (cont.)

Example: d-shell Splitting in Octohedral Coordinations
Take the Hydrogen-like orbitals $\psi_{\ell m}$ as a basis for spherically symmetric states. The d-shell orbitals are the basis functions of the $\ell = 2$ representations.

The octohedral complex's symmetry group is O, with it's character table and the $\Gamma^{\ell=2}$ representation:

O	$1\langle \mathbb{I} \rangle$	$8\langle C_3 \rangle$	$3\langle C_2 \rangle$	$6\langle C_2' \rangle$	$6\langle C_4^3 \rangle$
$(d)\ \Gamma^{\ell=2}$	5	-1	1	1	-1
A_1	1	1	1	1	1
A_2	1	1	1	-1	-1
E	2	-1	2	0	0
T_1	3	0	-1	-1	1
T_2	3	0	-1	1	-1

Orthogonality then gives $\gamma^{\ell=2} = E \oplus T_2$. In practice, this results in a 5-fold degeneracy being lifted into a two- and three-fold degeneracy.

Basis Functions

Every representation of a group inherits a vector space on which it has a natural action. This vector space is spanned by a chosen set of basis functions $|\psi_i^\alpha\rangle$.

- An irrep Γ^α's components are determined by basis:

$$\Big(\Gamma^\alpha(g)\Big)_i^j = \langle\psi_j^\alpha|\Gamma^\alpha(g)|\psi_i^\alpha\rangle$$

- Basis functions are said to "transform as an irrep α" if:

$$\Gamma^\alpha(g)|\psi_i^\alpha\rangle = \sum_j \Big(\Gamma^\alpha(g)\Big)_i^j |\psi_j^\alpha\rangle$$

Basis Functions (cont.)

Example: $3D$ Basis Functions of D_3
Consider previous $3D$ representation of C_3 rotations in D_3, which act on vectors $\vec{r}_\rho$:

$$\vec{r}_\rho = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \vec{r}_{\rho^{(2)}} \oplus \vec{r}_{\rho^{(1)}} = \begin{bmatrix} x \\ y \end{bmatrix} \oplus \begin{bmatrix} z \end{bmatrix}$$

Clearly, x, y is a basis transforming as Γ^2 and z is a basis for Γ^1.

Example: Tight Binding
The tight binding approximation often uses localized Hydrogen-like orbitals ψ^{ℓ}_{nm} as a basis for many-body systems

Projection Operators

If we have an explicit form for IR α, we may project an arbitrary function onto the k-th basis function f_k^α of IR α with $\hat{P}_\alpha^{kk}$:

$$\hat{P}_\alpha^{kk} = \frac{d_\alpha}{N} \sum_g \left[\Gamma_\alpha^{kk}(g)\right]^* O(g)$$

where d_α is the dimensional of IR α, and then we have:

$$f_\alpha^k(\vec{r}) = \hat{P}_\alpha^{kk} f(\vec{r})$$

From the characters alone, we may project a function onto it's total α subspace with $\hat{P}_\alpha$:

$$\hat{P}_\alpha = \sum_k \hat{P}_\alpha^{kk} = \frac{d_\alpha}{N} \sum_g \chi^{(\alpha)*}(g) \hat{O}(g)$$

Coupling Coefficients

Consider a direct product decomposition of some irreps Γ:

$$\Gamma^\alpha \otimes \Gamma^\beta = \bigoplus_\gamma c_{\alpha\beta\gamma}\Gamma^\gamma$$

Products of basis functions $u_i^\alpha v_j^\beta$ then decompose similarly into a direct sum of irreps with basis functions ψ_n^γ via the coupling coefficients $U_{\alpha i\beta j}^{\gamma n}$ as:

$$\psi_n^\gamma = \sum_{i,j} U_{\alpha i\beta j}^{\gamma n} u_i^\alpha v_j^\beta$$

Example: Clebsch-Gordan Coefficients

The Clebsch-Gordan coefficients $C_{\ell_1 m_1 \ell_2 m_2}^{\ell_f m_f}$ are the coupling coefficients of $SO(3)$, which relate tensor product spaces of spherical harmonics Y_m^ℓ to direct sums of spherical harmonics.

$$Y_{\ell_1}^{m_1}(\Omega)Y_{\ell_2}^{m_2}(\Omega) = \sum_{\ell_3, m_3} \sqrt{\frac{(2\ell_1+1)(2\ell_2+1)}{4\pi(2\ell_3+1)}}\, C_{\ell_1 m_1 \ell_2 m_2}^{\ell_3 m_3} C_{\ell_1 0 \ell_2 0}^{\ell_3 0} Y_{\ell_3 m_3}(\Omega)$$

Coupling Coefficients (cont.)

Dirl's Formula [**dirl1979**]:

$$(U^{\gamma n}_{\alpha i \beta j})^m = \sqrt{\frac{d_\gamma}{N_G}} \left[\sum_{g \in G} \Gamma^\alpha_{qq}(g) \Gamma^\beta_{ss}(g) \Gamma^{\gamma \dagger}_{aa}(g) \right]^{-\frac{1}{2}} \cdot \sum_{g \in G} \Gamma^\alpha_{iq}(g) \Gamma^\beta_{js}(g) \Gamma^{\gamma \dagger}_{na}(g)$$

Equivariant Functions

An equivariant function $f : X \to Y$, where X, Y are vector spaces, is one that 'commutes' with a group's actions, satisfying:

$$f\left(\mathcal{D}^X(g)x\right) = \mathcal{D}^Y(g)f(x)$$

Tensor products are uniquely equivariant with respect to their argument vector spaces.

Why Equivariant Functions?

Physics assumes there exists a map between configurations and properties of physical systems:

$$\{\vec{r}_i\} \xrightarrow[\text{Nature}]{} \{\vec{r}_i'\}$$

Neural network techniques assume maps between abstract input and output spaces exist:

$$X \xrightarrow[\text{Neural Network}]{} Y$$

and that these may be expressed in a basis of adequately large neural networks.

Why Equivariant Functions? (cont.)

Physical processes respect coordinate system rotations R:

$$\{R\vec{r}_i\} \xrightarrow[\text{Nature}]{} \{R\vec{r}_i'\}$$

In general, neural networks do not:

$$RX \xrightarrow[\text{Neural Network}]{} ?$$

So what do we want?

Equivariant Networks

We want equivariant networks for physics!

$$RX \xrightarrow[\text{Neural Network}]{\text{Equivariant}} RY$$

In equivariant networks, we consider feature vectors in basis of representation space as:

$$v_i^\alpha$$

with indexes α denoting the irrep it transforms as, and i denoting the dimension of irrep space α.

Equivariant Networks (cont.)

Compositions of equivariant functions are equivariant [1].
We consider building blocks:

- Self-interaction:
$$v'_{nc} = W_c^i v_{ni}$$
- Non-linear functions along channels (for trivial irrep $\alpha = 1$):
$$v'_{nc} = f(v_{nc} + b_{nc})$$
- Group correlation with filters:
$$[F \star v](g) = \sum_{h \in G} \sum_c F_c(h) v_c(g^{-1}h)$$
- Tensor products with filters:
$$(v'_{nc})^\gamma_n = \sum_{ij} U^{\gamma n}_{\alpha i \beta j} (F_j^\beta \otimes V_i^\alpha)$$
- Pooling over nodes or channels:
$$v_c = \sum_n v_{nc}$$

$SO(3)$ Properties

Most modern equivariant networks are specifically $SO(3)$-Equivariant.

- Inifinite order $\rightarrow$ infinite irreps
- Irreps are Wigner $\mathcal{D}^\ell$ matrices
- Natural basis set Y_m^ℓ of dimension $d_\ell = 2\ell + 1$

Many models are further referred to as $E(3)$-equivariant for the Euclidean group (add parity and translation invariance).

Tensor Field Networks

Tensor field networks [2] use SO(3) equivariant convolution:

$$\left(v_{nc}^{L+1}\right)_m^{\ell} = \left(v_{nc}^{L}\right)_m^{\ell} + \sum_{b\in\mathcal{N}(n)} \sum_{\ell_f m_f, \ell_i m_i}^{\ell_{\max} m_{\max}} c_{\ell_f m_f \ell_i m_i}^{\ell m} \left(F_c^L(r_{nb})\right)_{m_f}^{\ell_f} \left(v_{bc}^{L}\right)_{m_i}^{\ell_i}$$

- v_{nc}^L: Node feature of node n, channel c, layer L
- $c_{\ell_f m_f \ell_i m_i}^{\ell_o m_o}$: Clebsch-Gordan coefficients
- F_c^L: Filter function (trainable)
- r_{nb}: Radius between nodes n and neighbor $b \in \mathcal{N}(n)$

F is generally a neural network; r is often expanded with some radial basis function (RBF): Gaussian, Bessel, etc.

Point Group Properties

Groups of operations that leave at least one point fixed and are compatible with a Bravais lattice.

Point groups (in 3D) have following properties:

- 32 crystallographic point groups in 3D
- Finite number of irreps for each
- All irreps of point groups are of dimension $d_\alpha = 1, 2, 3$
- All allow a (often reducible) 3D representation
- Decompositions with $c_{\alpha\beta\gamma} = 0, 1, 2$

(Point group representations may be used to generate induced representations of space groups)

Site-symmetry Equivariant Networks

Index vectors by α, i from site-symmetry group instead!

$$\left(v_{nc}^{L+1}\right)_m^{\ell} = \left(v_{nc}^{L}\right)_n^{\gamma} + \sum_{b\in\mathcal{N}(n)} \sum_{\alpha i,\beta j} U_{\alpha i\beta j}^{\gamma n}\left(F_c^L(r_{nb})\right)_j^{\beta}\left(v_{bc}^{L}\right)_i^{\alpha}$$

- Natural cutoff (finite summation over α, β)
- Natural readout for symmetric properties ($\gamma = 1$)
- Accounts for all physical symmetry but not more

Space Groups

Semidirect product of point group R and lattice translation group T:

$$G = T \ltimes R$$

- Describes all symmetries of crystals
- Infinite order $\rightarrow$ infinite irreps
- Can learn features associated with point group and induce for space group

Induced Representations

"Opposite of a reduced representation"

Use subgroup G representation $\tilde{\rho}$ to form representation ρ of parent H:

$$\rho_{\alpha i,\beta j}(h) = \begin{cases} \tilde{\rho}(g)_{ij} & \text{if } h_\alpha^{-1} h h_\beta = g \in G \\ 0 & \text{else} \end{cases}$$

where α indexes a coset decomposition of H into $G \subset H$.

Example: Induced Representation of S_2

Take trivial group E with representation $\tilde{\rho}(E) = 1$. Induced rep. of S_2 then is:

$$\rho_{S_2}(\mathbb{I}) = \begin{bmatrix} \tilde{\rho}(\mathbb{I}) & \tilde{\rho}([12]) \\ \tilde{\rho}([12]) & \tilde{\rho}([12][12]) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\rho_{S_2}([12]) = \begin{bmatrix} \tilde{\rho}([12]) & \tilde{\rho}([12][12]) \\ \tilde{\rho}([12][12]) & \tilde{\rho}([12][12][12]) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Hamiltonian Group

We define the operator $\hat{O}_G$ to be a representation of a group G that acts on H's input space $\vec{r}$ as:

$$\hat{O}_G(g)\hat{H}(\vec{r}) = \hat{H}(g^{-1}\vec{r})$$

The "group of the Hamiltonian" is the largest group of the form above that commutes with the Hamiltonian.

$$[\hat{H}, \hat{O}(g)] = 0 \qquad \forall g \in G$$

In such a case, the operators must have a simultaneous set of eigenvectors ψ_α^k that span the space of functions:

$$f(\vec{r}) = \sum_{k,\alpha} c_k^\alpha \psi_\alpha^k = \sum_{k,\alpha} f_\alpha^k(\vec{r})$$

We have some freedom in choice of this set of basis functions on which the linear operators of the representations act.

Hydrogen Orbitals

The Hydrogen Hamiltonian $\hat{H}_H(\vec{r})$ for a non-relativistic electron has separable eigenfunctions:

$$\psi_i(\vec{r}) = R_i(r)\Omega_i(\theta, \phi)$$

Furthermore, $\hat{H}_H$ commutes with $SO(3)$ so these must be simultaneous with $SO(3)$:

$$[\hat{H}, \hat{O}(g)] = 0 \quad \forall g \in SO(3) \quad \Rightarrow \quad \psi^{\ell}_{nm}(\vec{r}) = R^{\ell}_n(r) Y^{\ell}_m(\theta, \phi)$$

(where R is also indexed by ℓ due to an 'accidental' symmetry)

- θ, ϕ dependence determined by group theory
- r dependence determined by specific form of $\hat{H}_H$

Tight-Binding Approximations

In simplest form, consider H-like orbitals $|\psi^{\ell}_{nmp}\rangle$ localized at points p as basis for many particle system defined by $\hat{H}$ so:

$$\hat{H}_{nmp\ell,jkhl} = \langle\psi^{l}_{jkh}|\hat{H}|\psi^{\ell}_{nmp}\rangle$$

DeepH

- Take hydrogen-like orbitals and treat them as $SO(3)$ features
- Allows for the learning of all rotational symmetries but doesn't enforce them from physical considerations.
- Predicts DFT Hamiltonian from first large set of data

DeepSymmetry

- Incorporate point group symmetry of crystal and molecular sites by directly learning features associated with basis functions that transform as irreducible representations.
- Predict symmetry of properties *ab initio* but use machine learning for specific values
- Target data less clear, Wannier functions?

Maybe variational approach, minimize basic Hamiltonian?

References

[1] Taco S. Cohen and Max Welling. *Group Equivariant Convolutional Networks*. 2016. arXiv: 1602.07576 [cs.LG]. URL: https://arxiv.org/abs/1602.07576.

[2] Nathaniel Thomas et al. "Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds". In: *arXiv preprint arXiv:1802.08219* (2018). arXiv: 1802.08219 [cs.LG]. URL: https://arxiv.org/abs/1802.08219.