Equivariant Prediction of Tensorial Properties and Transfer Learning

Alex Heilman

January 6, 2025

Overview

- J_z basis $\to Y_{\ell=1}^m$ unit vectors
- Clebsch-Gordon expansion for symmetric tensor spaces
- Constructing symmetric, $SO(3)$ invariant, tensor subspaces
- Equivariant networks and harmonics
- Test results: pretraining and prediction

J_z Basis

Recall $\ell = 1$ spherical harmonics (with Racah normalization):

$$
\begin{aligned}
Y_1^{+1} &= -\frac{1}{\sqrt{2}}(x + iy) = \frac{1}{\sqrt{2}} \sin\phi e^{i\theta} \\
Y_1^{0} &= z = \cos\phi \\
Y_1^{-1} &= -\frac{1}{\sqrt{2}}(x - iy) = \frac{1}{\sqrt{2}} \sin\phi e^{-i\theta}
\end{aligned}
$$

So, define J_z basis:

$$
\begin{bmatrix} a_+ \\ a_0 \\ a_- \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{2}} & -\frac{i}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ -\frac{1}{\sqrt{2}} & +\frac{i}{\sqrt{2}} & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}
$$

so that $\hat{n} = a_+ Y_1^1 + a_0 Y_1^0 + a_- Y_1^{-1}$

Clebsch-Gordon Expansion

Build larger spherical harmonic tensors with CG expansion:

$$Y_{\ell_1}^{m_1} \otimes Y_{\ell_2}^{m_2} = \sum_{L=-|\ell_1-\ell_2|}^{\ell_1+\ell_2} \sum_{M=-L}^{L} c_{\ell_1 0 \ell_2 0}^{L0} c_{\ell_1 m_1 \ell_2 m_2}^{LM} Y_L^M$$

where Y_L represents a $2L+1$ dimensional symmetric tensor space of rank L.
We use this as a relation between symmetric tensor's J_z basis components and higher order spherical harmonic tensors.

$$T^{(n)} = a_{\underbrace{\alpha\beta\ldots}_{n}} \left(Y_1^{\alpha} \otimes Y_1^{\beta} \otimes \ldots \right) \quad \Rightarrow \quad y_\ell^m Y_L^M$$

But, what about asymmetric tensors?

$SO(3)$ Invariant Tensor Subspaces

We can always reduce an arbitrary tensor T that transforms under a transformation as:

$$T_{x_1 x_2 \ldots x_n} \to T_{x_1' x_2' \ldots x_n'} = R^{x_1}_{x_1'} R^{x_2}_{x_2'} R^{x_3}_{x_3'} T_{x_1 x_2 \ldots x_n},$$

into a set of irreducible (but not necessarily unique) symmetric, $SO(3)$ invariant subtensors:

$$\{h^{(\ell)}\} \to \{h'^{(\ell)}\} = \{\mathcal{D}^{\ell}(R) h^{(\ell)}\}$$

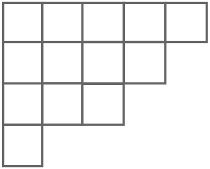
This decomposition can be constructed by consecutive decomposition with respect to GL and then O and SL

$$SO = SL \cap O \subset GL$$

GL Decomposition

- Decompositions under general linear group GL are simultaneous with decompositions under symmetric group S (*Schur-Weyl Duality*)

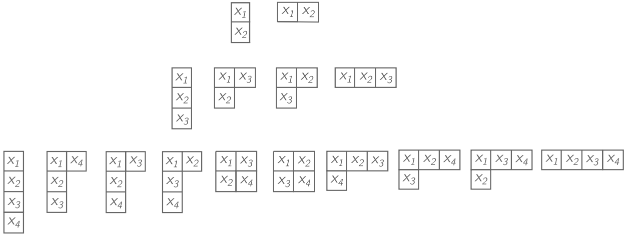
- Irreducible representations of symmetric group are diagramatically described by Young diagrams.



Young diagrams are said to be of some shape $\lambda : (\lambda_1, \lambda_2, ..., \lambda_k)$, where λ_i refers to the depth of row i and $\lambda_{i+1} \leq \lambda_i \leq \lambda_{i-1}$. Above: $(5, 4, 3, 1)$

GL Decomposition cont.

We can then form a set of Young tableaux from diagrams by filling in the boxes from a set of ordered indices $\{x_1, x_2, ..., x_k\}$ corresponding to tensor components $T^{x_1 x_2 \dots x_k}$.



A standard tableu is one filled with indices x_i (without repeats) with entries increasing in index i down each column and across (to the right) rows.

GL Decomposition cont.

Each of these standard tableaux correspond to an invariant subspace under S_k.

These invariant subspaces may be constructed by correspond products of symmetrizers s and antisymmetrizers a:

$$s_\lambda = \prod_{\mathcal{I} \in \text{Cols}(\lambda)} \mathcal{S}(\mathcal{I})$$

$$a_\lambda = \prod_{\mathcal{I} \in \text{Rows}(\lambda)} \mathcal{A}(\mathcal{I})$$

where $\mathcal{S}$ and $\mathcal{A}$ act component-wise as:

$$\left[\mathcal{S}(\mathcal{I})T\right]_{ijk\ldots} = \sum_{\sigma_\mathcal{I}} T_{\sigma_\mathcal{I}(ijk\ldots)}$$

$$\left[\mathcal{A}(\mathcal{I})T\right]_{ijk\ldots} = \sum_{\sigma_\mathcal{I}} \text{sgn}(\sigma_\mathcal{I}) T_{\sigma_\mathcal{I}(ijk\ldots)}$$

O Decomposition

Orthogonal group O, preserves inner product between vectors $< \cdot, \cdot >$. defined by means of a metric tensor g_{ij} , which transforms as a second rank tensor under some transformation R as:

$$g_{x_1 x_2} \to g_{x_1' x_2'} = R^{x_1}_{x_1'} R^{x_2}_{x_2'} g_{x_1 x_2}$$

Contractions of arbitrary tensors T with g_{ij} are invariant under O.

For example, consider the rank-3 covariant T^{ijk}:

$$g_{i'j'} T^{i'j'k'} = S^{k'} = R^{i}_{i'} R^{j}_{j'} g_{ij} R^{i'}_{i} R^{j'}_{j} R^{k'}_{k} T^{ijk} = R^{k'}_{k} S^{k}$$

so that S acts like an invariant vector subspace under O.

SL Decomposition

Special linear group SL of transformations is defined as invertible linear transformations with determinant equal to positive one.

Under SL, orientation and volume are preserved, where volume is defined as the contraction of a tensor with the fully antisymmetric tensor ϵ_{ijk}, which transforms under $R \in GL(3)$ as:

$$\epsilon_{x_1 x_2 x_3} \to \epsilon_{x_1' x_2' x_3'} = \det(R) R^{x_1}_{x_1'} R^{x_2}_{x_2'} R^{x_3}_{x_3'} \epsilon_{x_1 x_2 x_3}$$

Similar to the case of g_{ij} in O, contractions with ϵ_{ijk} of arbitrary tensors yield SL invariant subspaces.

$SO(3)$ Decompositions

In $SO(3)$, we may use all of the above (Young, g_{ij}, ϵ_{ijk}):

- Young symmetrizers return a set of tensors of known symmetries (under index permutation).
- Contractions with g_{ij} along antisymmetric pairs of indices vanish.
- Contractions with ϵ_{ij} along symmetric sets of indices vanish.

These may all be used together to decompose an arbitrary tensor into a set of symmetric, $SO(3)$ invariant, symmetric tensor subspaces. These then may be related to harmonic coefficients by way of the CG expansion from the J_z basis given before.

Example: Piezoelectric Tensors

The piezoelectric strain components d_{ijk} are symmetric under i, j so that:

$$d_{ijk} = d_{jik}$$

according to this symmetry, we see all Young tableaux but the following must disappear:

$$\begin{array}{|c|}\hline x_1 \\ \hline x_2 \\ \hline x_3 \\ \hline \end{array} \qquad\qquad \begin{array}{|c|c|}\hline x_1 & x_3 \\ \hline x_2 & \multicolumn{1}{c}{} \\ \cline{1-1} \end{array}$$

defined component-wise (using the defining symmetry):

$$S_{ijk} = \frac{1}{3}\left(d_{ijk} + d_{jki} + d_{ikj}\right)$$
$$A_{ijk} = \frac{1}{3}\left(2d_{ijk} - d_{jki} - d_{ikj}\right)$$

Example: Piezoelectric Tensors

This may be derived as (adopting the convention of symmetrization before antisymmetrization):

$$\begin{array}{|c|}\hline x_1 \\ \hline x_2 \\ \hline x_3 \\ \hline \end{array} \qquad\qquad \begin{array}{|c|c|}\hline x_1 & x_3 \\ \hline x_2 \\ \cline{1-1} \end{array}$$

$$\mathcal{S}(ijk) \qquad \mathcal{A}(ij)\mathcal{S}(ik)$$

$$\begin{aligned} \mathcal{S}(ijk)d_{ijk} &= d_{ijk} + d_{ikj} + d_{kji} + d_{jki} + d_{kij} + d_{kji} \\ &= 2(d_{ijk} + d_{ikj} + d_{kji}) \end{aligned}$$

$$\begin{aligned} \mathcal{A}(ik)\mathcal{S}(ij)d_{ijk} &= \mathcal{A}(ij)(d_{ijk} + d_{jik}) \\ &= d_{ijk} - d_{kji} + d_{jik} - d_{jki} \\ &= 2d_{ijk} - d_{kji} - d_{jki} \end{aligned}$$

Where the respective normalization coefficients (neglected here) may be derived from the diagram's shape via the hook-length formula.

Example: Piezoelectric Tensors

Alex Heilman

Further decompose A into the trace vector v_i:

$$v^i = g_{jk} A^{ijk}$$

and the traceless, symmetric tensor b_{ij}:

$$b_{ij} = \frac{1}{2}\left(\epsilon_i^{mk} A_{mkj} + \epsilon_j^{mk} A_{mki}\right)$$

$SO(3)$ Equivariant Networks

An $SO(3)$ equivariant network is a neural network satisfying:

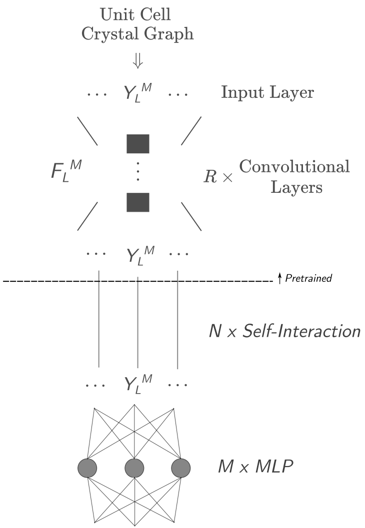
$$f(\{\mathcal{D}^{\ell_1}(R)h_{(\ell_1)}\}) = \{\mathcal{D}^{\ell_2}(R)h'_{(\ell_2)}\} \qquad \forall R \in SO(3)$$

This can be accomplished by associating hidden features V and filters F with spherical harmonics defining convolution as their tensor product:

$$\mathcal{L}^{\ell_o}_{acm_o}\left(\vec{r}_a, V^{\ell_i}_{acm_i}\right) = \sum_{m_f, m_i} c^{\ell_o m_o}_{\ell_i m_i \ell_f m_f} \sum_b F^{\ell_f \ell_i}_{cm_f}(r_{ab}) V^{\ell_i}_{bcm_i}$$

The important thing to note here is that these networks naturally yield coefficients of spherical harmonic tensors.

General Model Architecture



Training Progressions

Pretraining consisted of a funnel-down approach where the first graph layers were first trained on the largest set, then trained on the second largest, etc.

Here, this corresponds to the progression:

$$\text{Band Gap} \rightarrow \text{Elasticity} \rightarrow \text{Dielectric} \rightarrow \text{Piezoelectric}$$

This progression is compared to blind training on each dataset individually

Results

Model	Elasticity (10,829) MAE (log(GPA))	Dielectric MAE	Piezoelectric MAE (GPA^{-1})
Blind	7.387	4.818	0.170
Pretrain	7.274	4.525	0.170

The low impact of pretraining may be due to several factors:

- Lack of shared domain-relevance in graph layers
- Lack of overlap in datasets (unlikely)
- Lack of overlap in filters for different ℓ order targets