Stony Brook University

Equivariant GNNs

Wenhan Gao

Department of Applied Mathematics and Statistics

FAR BEYOND

Stony Brook University

TP-based Equivariant GNNs

A tensor is a multi-dimensional array with directional information.
We have two key operators for Cartesian tensors: tensor product and tensor contraction. They allow us to move up and down the rank of Cartesian tensors to produce tensors of higher ranks or contract them down to lower ranks.
Intuitively, tensor product is to create new tensors using the indices of old tensors and tensor tensor contraction is to collapse old tensors using summation and indices.

Today's Keywords:
1. Spherical tensors
2. Clebsch-Gordan coefficients
3. Representations and irreps
4. Wigner D-matrices
5. Spherical harmonics

Coming up:
1. TP based equivariant interactions
2. Equivariant GNNs with spherical tensors (Spherical EGNN)

New topic:
Unconstrained Geometric GNNs

Tensor Products

A vector v in 3D Euclidean space $\mathbb{R}^3$ can be expressed in the familiar Cartesian coordinate system in the standard basis e_x, e_y, e_z, and coordinates A_x, A_y, A_z, where $\mathbf{e}_x = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ $\mathbf{e}_y = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ $\mathbf{e}_z = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$.

When you perform the tensor (or outer) product of two vectors in $\mathbb{R}^3$, you obtain a matrix (or a rank2 tensor). If you have two vectors $\mathbf{u} = \begin{pmatrix} u_x \\ u_y \\ u_z \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix}$, their tensor product $\mathbf{u} \otimes \mathbf{v}$ is given by:

$$\mathbf{u} \otimes \mathbf{v} = \begin{pmatrix} u_x \\ u_y \\ u_z \end{pmatrix} \otimes \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} = \begin{pmatrix} u_x v_x & u_x v_y & u_x v_z \\ u_y v_x & u_y v_y & u_y v_z \\ u_z v_x & u_z v_y & u_z v_z \end{pmatrix}.$$

This is easier to understand or remember with the definition of outer product of two functions: $(f \otimes g)(x, y) = f(x)g(y)$.

Basis of Tensors

In terms of basis, if $\mathbf{u}$ and $\mathbf{v}$ are expressed in the standard basis $\{\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z\}$, the resulting tensor product $\mathbf{u} \otimes \mathbf{v}$ can be viewed as a linear combination of the outer products of the basis vectors:

$$\mathbf{u} \otimes \mathbf{v} = u_x v_x (\mathbf{e}_x \otimes \mathbf{e}_x) + u_x v_y (\mathbf{e}_x \otimes \mathbf{e}_y) + u_x v_z (\mathbf{e}_x \otimes \mathbf{e}_z) + u_y v_x (\mathbf{e}_y \otimes \mathbf{e}_x) + u_y v_y (\mathbf{e}_y \otimes \mathbf{e}_y) + u_y v_z (\mathbf{e}_y \otimes \mathbf{e}_z) + u_z v_x (\mathbf{e}_z \otimes \mathbf{e}_x) + u_z v_y (\mathbf{e}_z \otimes \mathbf{e}_y) + u_z v_z (\mathbf{e}_z \otimes \mathbf{e}_z)$$

The basis are given by:

$$\mathbf{e}_x \otimes \mathbf{e}_x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \mathbf{e}_x \otimes \mathbf{e}_y = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \mathbf{e}_x \otimes \mathbf{e}_z = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \mathbf{e}_y \otimes \mathbf{e}_x = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \ldots$$

Change of Basis

Let $\vec{v} \in V$ be a vector. Fix a basis $\{e_1, \ldots, e_n\}$, whence you have $\vec{v} = \sum_{i=1}^{n} e_i v^i = (e_1, \ldots e_n) \cdot (v^1, \ldots, v^n)^T$.

Then a change of basis is equivalent to the choice of an invertible $n \times n$ matrix M via

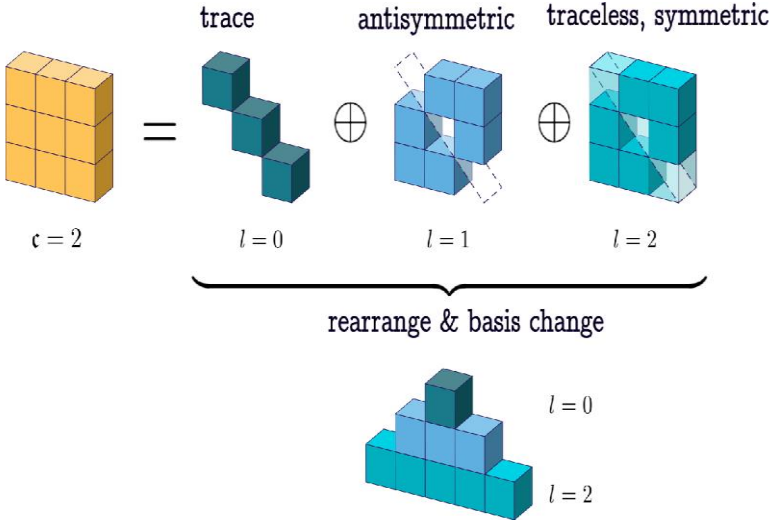
$\vec{v} = (e_1, \ldots, e_n)\, M M^{-1} (v^1, \ldots, v^n)^T = (\epsilon_1, \ldots, \epsilon_n) \cdot (\nu^1, \ldots, \nu^n)^T$, where $\{\epsilon_1, \ldots, \epsilon_n\}$ is the new basis and $\nu^1, \ldots, \nu^n$ are the new coefficients.

Example:

To change the basis of the vector $\mathbf{v} = [1, 2, 3]$ from the standard basis to the new basis $(1, 1, 0)$, $(0, 1, 1)$, $(0, 0, 1)$, we need to express $\mathbf{v}$ in terms of the new basis vectors.

1. Construct M: For new basis vectors $(1, 1, 0)$, $(0, 1, 1)$, and $(0, 0, 1)$, $M = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$.
2. $M^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix}$, $\mathbf{v}_{\text{new}} = B^{-1}\mathbf{v} = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$.

Change of Basis



Decompose the Catersian Tensor into a direct sum of some subspaces.

Image Credit: A Hitchhiker's Guide to Geometric GNNs for 3D Atomic Systems, Duvel et al.. and their presentation slides

$$
\begin{aligned}
\vec{T}^{[2]} &= \begin{bmatrix} T_{xx} & T_{xy} & T_{xz} \\ T_{yx} & T_{yy} & T_{yz} \\ T_{zx} & T_{zy} & T_{zz} \end{bmatrix} \\
&= \frac{T_{xx}+T_{yy}+T_{zz}}{3} \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} && (l=0) \\
&+ \begin{bmatrix} & \frac{T_{xy}-T_{yx}}{2} & \frac{T_{xz}-T_{zx}}{2} \\ -\frac{T_{xy}-T_{yx}}{2} & & \frac{T_{yz}-T_{zy}}{2} \\ -\frac{T_{xz}-T_{zx}}{2} & -\frac{T_{yz}-T_{zy}}{2} & \end{bmatrix} && (l=1) \\
&+ \begin{bmatrix} T_{xx} & \frac{T_{xy}+T_{yx}}{2} & \frac{T_{xz}+T_{zx}}{2} \\ \frac{T_{xy}+T_{yx}}{2} & T_{yy} & \frac{T_{yz}+T_{zy}}{2} \\ \frac{T_{xz}+T_{zx}}{2} & \frac{T_{yz}+T_{zy}}{2} & T_{zz} \end{bmatrix} - \frac{T_{xx}+T_{yy}+T_{zz}}{3} \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} && (l=2) \\
&= \overbrace{\frac{T_{xx}+T_{yy}+T_{zz}}{3}}^{\lambda_1} \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} && (l=0) \\
&+ \overbrace{\frac{T_{yz}-T_{zy}}{2}}^{\lambda_2} \begin{bmatrix} & & \\ & & 1 \\ & -1 & \end{bmatrix} + \overbrace{\frac{T_{zx}-T_{xz}}{2}}^{\lambda_3} \begin{bmatrix} & & -1 \\ & & \\ 1 & & \end{bmatrix} + \overbrace{\frac{T_{xy}-T_{yx}}{2}}^{\lambda_4} \begin{bmatrix} & 1 & \\ -1 & & \\ & & \end{bmatrix} && (l=1) \\
&+ \overbrace{\frac{2T_{xx}-T_{yy}-T_{zz}}{3}}^{\lambda_5} \begin{bmatrix} 1 & & \\ & -1 & \\ & & \end{bmatrix} + \overbrace{\frac{T_{xy}+T_{yx}}{2}}^{\lambda_6} \begin{bmatrix} & 1 & \\ 1 & & \\ & & \end{bmatrix} + \overbrace{\frac{T_{xz}+T_{zx}}{2}}^{\lambda_7} \begin{bmatrix} & & 1 \\ & & \\ 1 & & \end{bmatrix} \\
&+ \overbrace{\frac{T_{yz}+T_{zy}}{2}}^{\lambda_8} \begin{bmatrix} & & \\ & & 1 \\ & 1 & \end{bmatrix} + \overbrace{\frac{2T_{zz}-T_{xx}-T_{yy}}{3}}^{\lambda_9} \begin{bmatrix} & & \\ & -1 & \\ & & 1 \end{bmatrix} && (l=2)
\end{aligned}
$$

Inner product

$$\vec{v}\cdot\vec{w} = v_x w_x + v_y w_y + v_z w_z = 3\lambda_1$$

Cross product

$$\vec{v}\times\vec{w} = \begin{bmatrix} v_y w_z - v_z w_y \\ v_z w_x - v_x w_z \\ v_x w_y - v_y w_x \end{bmatrix} = \begin{bmatrix} \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix}$$

Image Credit: A Hitchhiker's Guide to Geometric GNNs for 3D Atomic Systems, Duvel et al.. and their presentation slides

Preliminary: Representations

A representation $\rho : G \to GL(V)$ is a group homomorphism from G to the general linear group $GL(V)$. That is, $\rho(g)$ is a linear transformation parameterized by group elements $g \in G$ that transforms some vector $\mathbf{v} \in V$ (e.g. an image or a tensor) such that

$$\rho(g') \circ \rho(g)[\mathbf{v}] = \rho(g' \cdot g)[\mathbf{v}].$$

Example: The representation of $SO(3)$ acting on a geometric 3D vector is a 3×3 orthogonal matrices with determinant 1.

Irreducible Representation: A representation $\rho : G \to GL(V)$ is said to be irreducible if there are no proper non-zero subspaces W of V that are invariant under all group actions, i.e., $\rho(g)W \subseteq W$ for all $g \in G$. In other words, V cannot be split into smaller subspaces that are individually invariant under the group action.

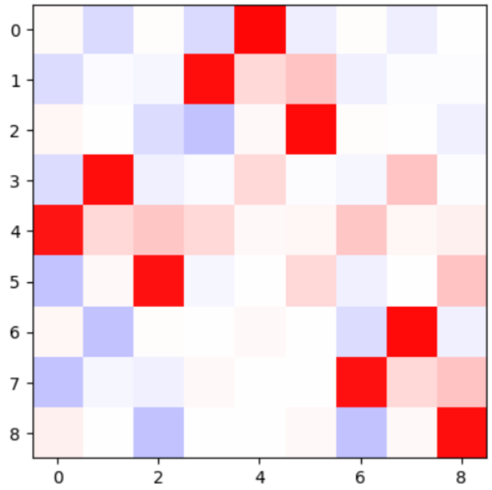
Decomposition: If a representation is reducible, it can be decomposed into a direct sum of irreducible representations (irreps). A block diagonal matrix can be written as the direct sum of the matrices that lie along the diagonal. In contrast, an irreducible representation cannot be decomposed further in this way.

Block-Diagonal Form: If a representation is reducible (i.e., can be decomposed into simpler parts), it can be rewritten (decomposed) into block-diagonal form (each of the blocks are the simpler parts, i.e., irreps).

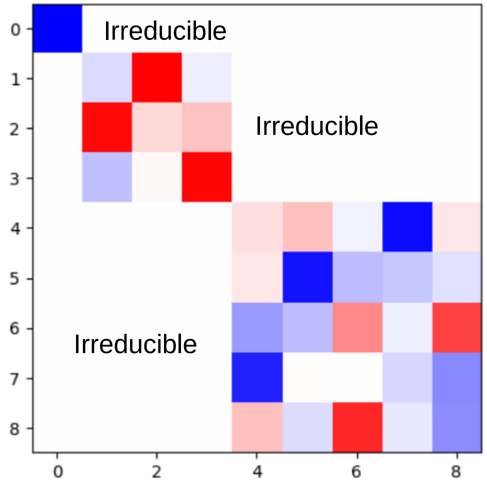
Why? Irreducible representations are fundamental because they represent the "building blocks" of more complex representations. Understanding the irreducible representations of a group provides a basis for understanding all possible representations of the group.

Example: See later.

Preliminary: Representations



Reducible
Representation



Decomposed
into irreps

Stony Brook University

Rotation of Spherical Tensor

To summarize, we have seen that the 9-dimensional rank-2 Cartesian tensor can be decomposed into a 1D, 3D and 5D part which correspond to the $l = 0, 1, 2$ irreps respectively

$$3 \otimes 3 = 1 \oplus 3 \oplus 5.$$

The $\ell = 0$ part is invariant, $\ell = 1$ part rotates as a normal vector in $\mathbb{R}^3$.

For types of higher l, the transformation behaviour under rotations is also known and given by the so-called Wigner D-matrices. The D here stands for Darstellung, the German word for representation. The Wigner D-matrix is how one can represent a rotation in the $2l+1$ dimensional space of a degree l spherical tensor.

$$\vec{T}^{(l)} \rightarrow \mathcal{D}^{(l)}(\mathbf{R})\vec{T}^{(l)},$$

where $\vec{T}^{(l)} \in \mathbb{R}^{(2\ell+1)}$ and $D^{\ell}(R) \in \mathbb{R}^{(2\ell+1)\times(2\ell+1)}$.

Tensor Products of Spherical Tensors

Unfortunately, the tensor product of two spherical tensors $\vec{S}^{(l_1)}$ and $\vec{T}^{(l_2)}$ is generally not a spherical tensor anymore.

Example: Suppose we have a $\ell_1 = 1$ (3 elements) tensor and a $\ell_2 = 2$ (5 elements), the tensor product would have $3 \times 5 = 15$ elements. However, this may not be a $\ell_3 = 7$ (15 elements) tensor.

However, we can decompose the product $\vec{S}^{(l_1)} \otimes \vec{T}^{(l_2)}$ back into spherical tensors.

As a rule, the $(l_1 l_2)$-dimensional tensor product of two spherical tensors of ranks l_1 and l_2 decomposes into:

$$l_1 \otimes l_2 = |l_1 - l_2| \oplus |l_1 - l_2 + 1| \oplus \cdots \oplus (l_1 + l_2 - 1) \oplus (l_1 + l_2)\,.$$

This means the $l_1 l_2$-dimensional product decomposes into exactly one spherical tensor for each rank between the absolute difference $|l_1 - l_2|$ and the sum $l_1 + l_2$.

Example: $|1-2| = 1$ and $1 + 2 = 3$. The 15 elements in the tensor product can be decomposed into a $l = 1$ (3 elements) tensor, a $l = 2$ (5 elements) tensor, and a $l = 3$ (7 elements). In some not so rigorous notation:$1 \otimes 2 = 1 \oplus 2 \oplus 3$.

Tensor Products of Spherical Tensors

The coefficients of the decomposition are given by the Clebsch-Gordan coefficients.

Example: Suppose we with to get the $l = 1$ tensor resulted from the tensor product of $\vec{S}^{(l_1)} \otimes \vec{T}^{(l_2)}$. Each of these three elements is a weighted sum of the 15 elements. So in total, we have $3 \times 5 \times 3 = 45$ coefficients. We denote this change of basis weights by $C^{(l_1,l_2,l_3)}_{(m_1,m_2,m_3)}$, where $-\ell_i \leq m_i \leq \ell_i$.

For example: $C^{(l_1=1,l_2=2,l_3=1)}_{(m_1=1,m_2=2,m_3=1)}$ means the coefficient of $t_1 \times s_2$ in order to get u_1 in the resulting tensor (We have 15 coefficients for u_1). That is

$$u_1 = \sum_{i=-1}^{1} \sum_{j=-2}^{2} C^{(l_1=1,l_2=2,l_3=1)}_{(m_1=i,m_2=j,m_3=1)} t_i s_j$$

$$u_2 = \sum_{i=-1}^{1} \sum_{j=-2}^{2} C^{(l_1=1,l_2=2,l_3=1)}_{(m_1=i,m_2=j,m_3=2)} t_i s_j$$

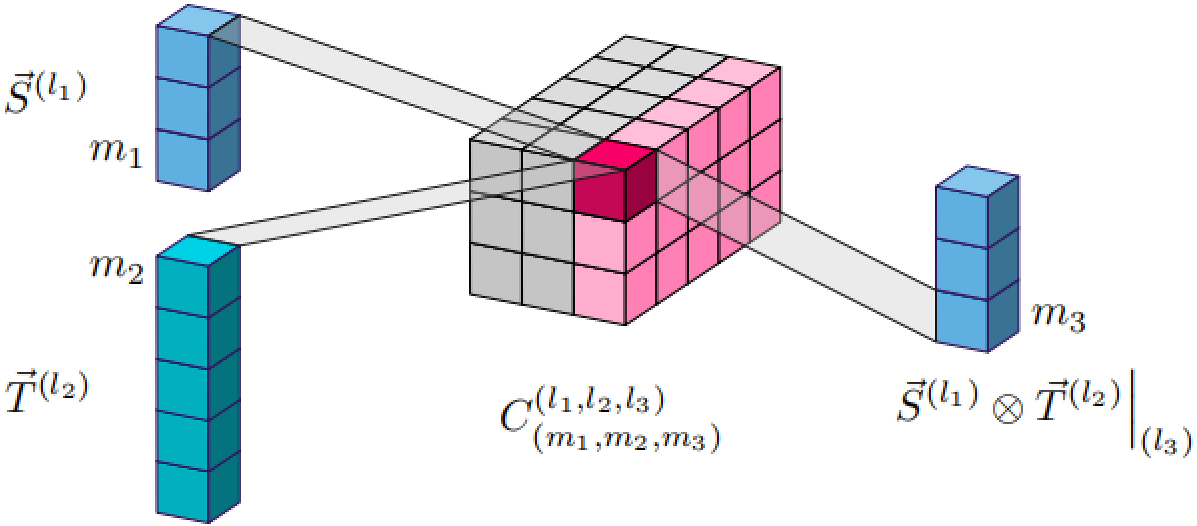
$$u_3 = \sum_{i=-1}^{1} \sum_{j=-2}^{2} C^{(l_1=1,l_2=2,l_3=1)}_{(m_1=i,m_2=j,m_3=3)} t_i s_j.$$

$$\mathbf{T} = \begin{pmatrix} t_{-1} \\ t_0 \\ t_1 \end{pmatrix}, \mathbf{S} = \begin{pmatrix} s_{-2} \\ s_{-1} \\ s_0 \\ s_1 \\ s_2 \end{pmatrix}, \mathbf{T} \otimes \mathbf{S} = \left(\begin{array}{c} u_1 \\ u_2 \\ u_3 \\ \hline u_4 \\ u_5 \\ u_6 \\ u_7 \\ u_8 \\ \hline u_9 \\ u_{10} \\ u_{11} \\ u_{12} \\ u_{13} \\ u_{14} \\ u_{15} \end{array}\right)$$

Similarly, $C^{(l_1=1,l_2=2,l_3=2)}_{(m_1,m_2,m_3)}$ will give the resulting $l = 2$ tensor, etc..

Stony Brook University

Tensor Products of Spherical Tensors



FAR BEYOND

Preliminary: Spherical Harmonics

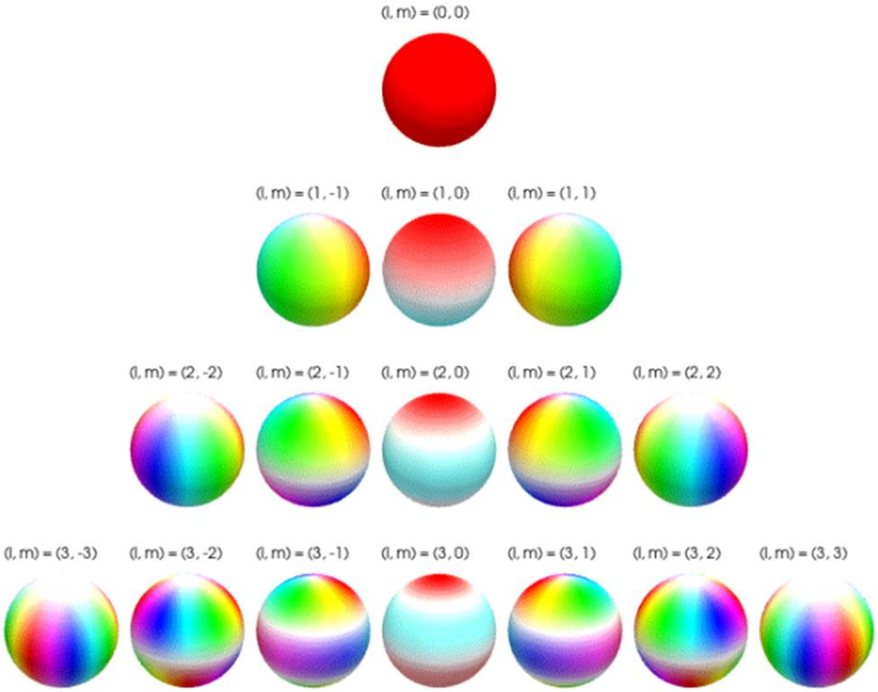
Real spherical harmonics $Y_l^m(\theta, \phi) : S^2 \to \mathbb{R}$ are real-valued functions defined on the surface of a sphere.

$$Y_\ell^m(\theta,\varphi) = (-1)^m \sqrt{\frac{2\ell+1}{4\pi}\frac{(\ell-m)!}{(\ell+m)}}\, P_\ell^m(\cos\theta)e^{im\varphi}$$

Each real spherical harmonic is indexed by two integers: l (degree) and m (order), where $l \geq 0$ and $-l \leq m \leq l$. They are used as an orthonormal basis for representing functions on the sphere. Under fairly general condition (square-integrable on the sphere), any function can be written as a linear combination of spherical harmonics as follows:

$$f(\theta,\varphi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} f_\ell^m Y_\ell^m(\theta,\varphi).$$

Preliminary: Spherical Harmonics



Stony Brook University

Preliminary: Spherical Harmonics

Generally, we can stack all the values from the degree-l spherical harmonics together to get a order-ℓ spherical tensor.

Example: Given a 3D point $v = (x, y, z)$, we can write it as a radial part $||v||$ and a directional part $v/||v||$. The directional part is now defined on S^2, write it as (θ, ϕ). We can get a order-1 tensor with spherical harmonics as: $V^{(1)} = \begin{pmatrix} Y_{l=1}^{m=-1}(\theta,\phi) \\ Y_{l=1}^{m=0}(\theta,\phi) \\ Y_{l=1}^{m=1}(\theta,\phi) \end{pmatrix}$.

For simplicity, we can rewrite (real) spherical harmonics as a vector-valued function for order-ℓ. That is $Y^{\ell}(\cdot) : \mathbb{R}^3 \to \mathbb{R}^{2\ell+1}$ maps an input 3D vector to a $(2\ell + 1)$-dimensional vector representing the coefficients of order- ℓ spherical harmonics bases.

Spherical harmonics function is equivariant to order-ℓ rotations, or so-called order-ℓ $SO(3)$ transformations:

$$Y^{\ell}(R\boldsymbol{c}) = D^{\ell}(R) Y^{\ell}(\boldsymbol{c}),$$

where $\boldsymbol{c}$ is a 3D point.

Note:

Spherical harmonics are a set of orthonormal functions defined on the surface of a sphere ($[0, \pi] \times [0, 2\pi]$) just like Fourier Basis. In fact, Fourier basis is called circular harmonics.

FAR BEYOND

Preliminary: Spherical Harmonics

We fix the terms we use:

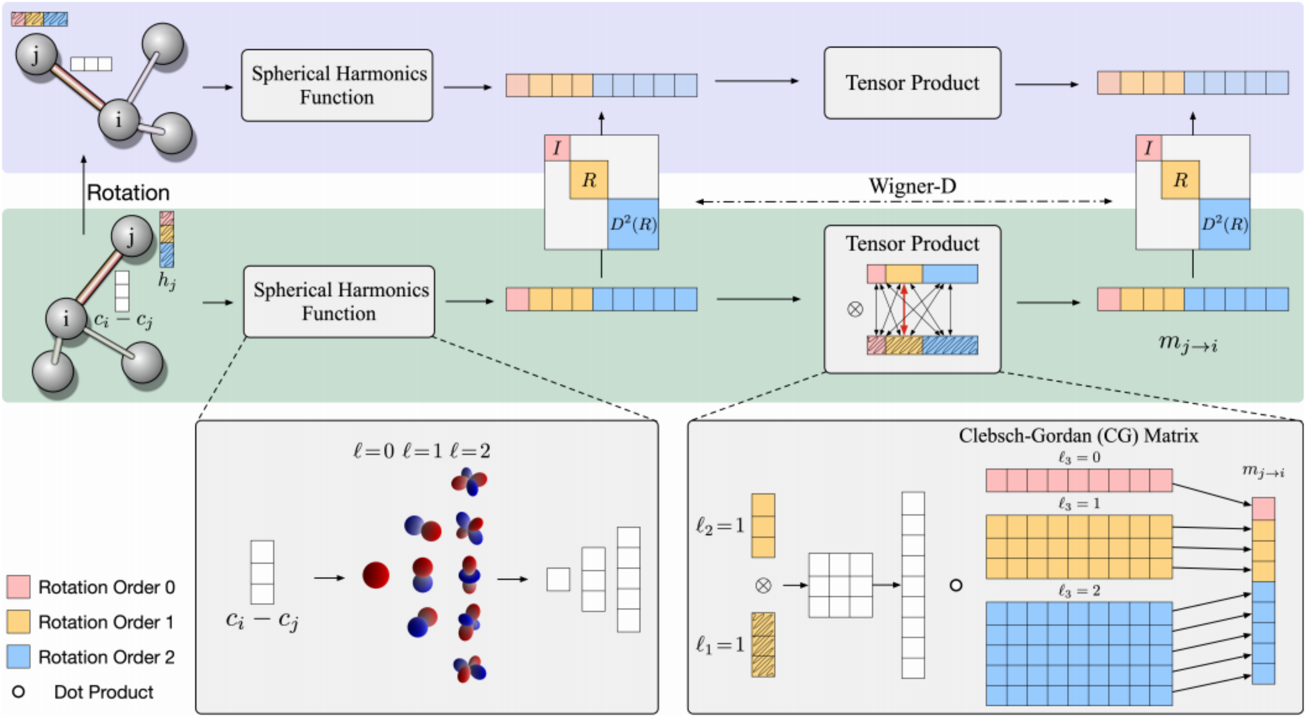
1. **Rank** k Cartesian Tensors: $T^{[k]}$
2. **Order-**ℓ Spherical Tensors: $T^{(l)}$
3. Spherical Harmonics with **degree** ℓ and **order** m: Y_l^m
4. **Order-**ℓ Spherical Harmonics Function that gives the **Order-**ℓ Spherical Harmonics Coefficients: $Y^{\ell}(\cdot)$

It is weird that they use order for degree, but this is the convention.

Caveat: May check if these names are rigorous, these are what I see from other papers, but not necessarily rigorous.

Stony Brook University

Why TP?



FAR BEYOND

Stony Brook University

FAR BEYOND