G-steerability of Equivariant Neural Networks

for (Hyper-)Nuclear Physics

Andrea Di Donna
May 28, 2025

andrea.didonna@unitn.it
TIFPA - Trento University

Equivariance formal Definition

- We can interpret layers $f(x)$ are functions on feature fields (e.g. $\mathbb{R}^3$):

$$f(x) : \mathcal{F}_{\text{in}} \to \mathcal{F}_{\text{out}}$$

  and treat the subject with Group Theory formalism
- Let a group G act on the input space $\mathcal{F}_{\text{in}}$ via $g \triangleright_{\text{in}} x$, and as $g \triangleright_{\text{out}} f(x)$ on $\mathcal{F}_{\text{out}}$.

Invariant Layer L:
$L(g \triangleright_{in} x) = L(x)$.

$$\begin{tikzcd} \mathcal{F}_{\text{in}} \arrow[dr, "L"] \arrow[dd, "g \triangleright_{\text{in}}"'] & \\ & \mathcal{F}_{\text{out}} \\ \mathcal{F}_{\text{in}} \arrow[ur, "L"'] & \end{tikzcd}$$

Equivariant Layer $\mathcal{T}$:
$\mathcal{T}(g \triangleright_{\text{in}} x) = g \triangleright_{\text{out}} \mathcal{T}(x)$.

$$\begin{tikzcd} \mathcal{F}_{\text{in}} \arrow[r, "\mathcal{T}"] \arrow[d, "g \triangleright_{\text{in}}"'] & \mathcal{F}_{\text{out}} \arrow[d, "g \triangleright_{\text{out}}"] \\ \mathcal{F}_{\text{in}} \arrow[r, "\mathcal{T}"'] & \mathcal{F}_{\text{out}} \end{tikzcd}$$

What's the role of equivariance in Physics?

$$\Psi(\vec{r} - \vec{x}) = e^{-i\hat{p}\cdot\vec{x}}\Psi(\vec{r}) \qquad P_{ij}\mathcal{A}(\cdots \otimes \psi_\alpha(\vec{x}_i) \otimes \cdots \otimes \psi_\beta(\vec{x}_j) \otimes \dots) = -\mathcal{A}(\cdots \otimes \psi_\alpha(\vec{x}_i) \otimes \cdots \otimes \psi_\beta(\vec{x}_j) \otimes \dots)$$

$$\pi^a(x) \longmapsto U_\theta(\pi^a(x)) = D^{ba}(U_\theta)\,\pi^a(x), \qquad U = \exp\Big(i \sum \boldsymbol{\theta} \cdot \boldsymbol{\tau}/2\Big)$$

Intertwiners and Equivariant Layers

UNIVERSITÀ DI TRENTO

Intertwiner (Weiler & Cohen 2019)

A linear map W between representations $\rho_{\text{in}}, \rho_{\text{out}}$ of G is an intertwiner if

$$(\star) \qquad W \rho_{\text{in}}(g) = \rho_{\text{out}}(g)\, W, \qquad \forall g \in G.$$

In terms of matrix representations, this just means that there exists a change of basis (invertible intertwiner) W such that the representations are similar as matrices:

$$\rho_{\text{in}}(g) = W^{-1}\, \rho_{\text{out}}(g)\, W, \qquad \forall g \in G.$$

Symmetric Group S_n: $G = S_n$ acts by permuting the indices of a set $X = \{x_1, \ldots, x_n\}$.

Antisymmetrizer – S_n Antisymm Equiv.

$$\boxed{\mathcal{A}\, P_h = \operatorname{sgn}(h)\, \mathcal{A} \qquad \mathcal{A} \;=\; \frac{1}{n!} \sum_{g \in S_n} \operatorname{sgn}(g)\, P_g}$$

Deep Sets – S_n Permutation Invariance.

$$\boxed{W_{\text{sum}} P_h \;=\; \mathbb{1}\, W_{\text{sum}} \qquad \Phi(X) = \rho\Big(\sum_{i=1}^{n} h(x_i)\Big),}$$

UNIVERSITÀ DI TRENTO

Due paradigmi di convoluzione $\Rightarrow$ un unico obiettivo

1. **Group Convolution**

$$(k \star_G f)(g) \;=\; \int_G k(g^{-1}h)\, f(h)\, dh, \qquad k, f : G \to V.$$

$\star_G$ è automaticamente equivariante $\big[(L_{g'}k) \star_G (L_{g'}f) = L_{g'}(k \star_G f)\big]$, ma l'integrale su $SO(3)$ è oneroso.

Osservazione — Cohen et al. (2018)

La doppia integrazione richiesta da $\star_G$ su $SO(3)$ (dominio e kernel) è impraticabile in una rete deep; meglio vincolare analiticamente il kernel in modo steerable e integrare solo sullo *dominio fisico*.

UNIVERSITÀ DI TRENTO

2 G-steerable Convolution (spazio fisico)

$$\boxed{(k \star f)(\mathbf{x}) \;=\; \int_{\mathbb{R}^3} K(\mathbf{x}-\mathbf{y})\, f(\mathbf{y})\, d\mathbf{y}, \quad K(R\hat{\mathbf{r}}) = R\, K(\hat{\mathbf{r}})\, R^{\top}.}$$

L'equivarianza è *impacchettata* nella forma chiusa di K.

- *Group convolution* $\star_G$: integra sia sul dominio che sul gruppo.
- *G-steerable convolution*: integra solo sul dominio; la parte angolare è incorporata analiticamente nel kernel.
- **Equivalenza teorica**: un kernel G-steerable che soddisfa $K(R\hat{r}) = RK(\hat{r})R^{\top}$ implementa la stessa trasformazione indotta da $\star_G$, ma con costo computazionale ridotto.

Kernel Sferico e Steerability

$$K(\mathbf{r}) \;=\; \sum_{J=0}^{2} \varphi_J(r) \sum_{m=-J}^{J} Y_{Jm}\big(\hat{\mathbf{r}}\big)\, Q_{Jm}^{\ell,\ell_in}, \qquad \hat{\mathbf{r}} = \frac{\mathbf{r}}{\|\mathbf{r}\|}.$$

- **Separazione radiale/angolare**:
  $\varphi_J(r)$ è la parte *learnable* (solo raggio), $Y_{Jm}(\hat{\mathbf{r}})$ sono basi fisse sul S^2.
- **Steerability**:
  $$Y_{Jm}\big(R\,\hat{\mathbf{r}}\big) = \sum_{m'=-J}^{J} D^{J}_{m'm}(R)\, Y_{Jm'}(\hat{\mathbf{r}})$$
  $$\Longrightarrow\; K(R\hat{\mathbf{r}}) = R\,K(\hat{\mathbf{r}})\,R^{\top}.$$
- Q_{Jm} sono i coefficienti di Clebsch–Gordan che accoppiano $\ell_{\text{in}} = \ell_{\text{out}} = 1$ ai tre canali angolari $J = 0, 1, 2$.

$$\boxed{K \star f \text{ è automaticamente } SO(3)\text{-equivariante}}$$

Equivalenza dei kernel (TFN $\leftrightarrow$ contratto)

$$\underbrace{\sum_{J=0}^{2} \varphi_J(r) \sum_{m=-J}^{J} Y_{Jm}(\hat{\mathbf{r}})\, Q_m^{(J)}}_{\text{kernel TFN per } \ell_{\text{in}}=\ell_{\text{out}}=1} = \underbrace{a(r)I \;+\; b(r)[\hat{r}]_\times \;+\; c(r)Q(\hat{r})}_{\text{kernel contratto}}$$

- La somma sui coefficienti di Clebsch–Gordan $Q_m^{(J)}$ *contratta* gli indici m.
- Produce tre tensori cartesiani ortogonali: I, $[\hat{r}]_\times$, $Q(\hat{r})$.
- Le funzioni radiali corrispondono: $a \propto \varphi_0$, $b \propto \varphi_1$, $c \propto \varphi_2$.
- Stessa legge di trasformazione $\Rightarrow$ stessa equivarianza.

$$\boxed{\text{Kernel TFN} \;\equiv\; \text{Kernel contratto (nostro)}}$$

Il nostro kernel equivariante

$$K(\mathbf{r}) = a(r)I \; + \; b(r)[\hat{\mathbf{r}}]_\times \; + \; c(r)\, Q(\hat{\mathbf{r}}), \qquad \hat{\mathbf{r}} = \frac{\mathbf{r}}{||\mathbf{r}||}.$$

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad [\hat{\mathbf{r}}]_\times = \begin{pmatrix} 0 & -\hat{r}_z & \hat{r}_y \\ \hat{r}_z & 0 & -\hat{r}_x \\ -\hat{r}_y & \hat{r}_x & 0 \end{pmatrix}, \quad Q(\hat{\mathbf{r}}) = 3\begin{pmatrix} \hat{r}_x^2 & \hat{r}_x\hat{r}_y & \hat{r}_x\hat{r}_z \\ \hat{r}_y\hat{r}_x & \hat{r}_y^2 & \hat{r}_y\hat{r}_z \\ \hat{r}_z\hat{r}_x & \hat{r}_z\hat{r}_y & \hat{r}_z^2 \end{pmatrix} - I.$$

- Caso particolare $\ell_{\text{in}} = \ell_{\text{out}} = 1$ del kernel TFN.
- Tre canali angolari $J = 0, 1, 2 \Rightarrow$ tre funzioni radiali a, b, c.
- Equivarianza garantita da $K(R\hat{r}) = R\, K(\hat{r})R^\top \;\; \forall R \in SO(3)$.

UNIVERSITÀ DI TRENTO

Lemma

Equivarianza del termine $b(r)\,[\hat r]_\times f$ Sia $R \in SO(3)$, $\hat r = \frac{r}{|r|}$ e $[\hat r]_{\times,ij} = \varepsilon_{ijk}\,\hat r_k$. Definiamo $g = b(|\,r\,|)\,[\hat r]_\times f \in \mathbb{R}^3$. Allora, per $f' = Rf$ e $\hat r' = R\hat r$,

$$g' \;:=\; b(|\,r\,|)\,[\hat r']_\times f' \;=\; R\,g,$$

cioè g trasforma come un vettore ($\ell = 1$).

Proof.

Poiché $b(|\,r\,|)$ dipende solo da $|\,r\,|$, è invariante per rotazioni. In indici,

$$([\hat r']_\times)_{ij} \;=\; \varepsilon_{ijk}\,\hat r'_k \;=\; \varepsilon_{ijk}\,R_{k\ell}\,\hat r_\ell.$$

Usiamo l'invarianza del simbolo di Levi–Civita sotto rotazioni proprie,

$$R_{ip}R_{jq}R_{kr}\,\varepsilon_{pqr} \;=\; \varepsilon_{ijk}\,,$$

che è equivalente a $[Rv]_\times = R[v]_\times R^\top$. Allora

$$([\hat r']_\times)_{ij} = R_{ip}R_{jq}\,\varepsilon_{pq\ell}\,\hat r_\ell = (R[\hat r]_\times R^\top)_{ij}.$$

Quindi

$$g'_i = b(|\,r\,|)\,([\hat r']_\times)_{ij} f'_j = b(|\,r\,|)\,(R[\hat r]_\times R^\top)_{ij}\,R_{jm} f_m = b(|\,r\,|)\,R_{ip}\,[\hat r]_{\times,pm}\,f_m = (Rg)_i,$$

UNIVERSITÀ DI TRENTO

Lemma (Equivarianza del termine $c(r)\,Q(\hat{r})\,f$)

Sia $Q_{ij}(\hat{r}) = 3\,\hat{r}_i\hat{r}_j - \delta_{ij}$. Definiamo $h = c(|\,r\,|)\,Q(\hat{r})\,f \in \mathbb{R}^3$. Allora, per $f' = Rf$ e $\hat{r}' = R\hat{r}$,

$$h' \;:=\; c(|\,r\,|)\,Q(\hat{r}')\,f' \;=\; R\,h,$$

cioè h trasforma come un vettore ($\ell = 1$).

Proof.

Poiché $c(|\,r\,|)$ dipende solo da $|\,r\,|$, è invariante per rotazioni. In indici,

$$Q_{ij}(\hat{r}') = 3\,\hat{r}'_i\hat{r}'_j - \delta_{ij} = 3\,R_{ik}\hat{r}_k\,R_{j\ell}\hat{r}_\ell - \delta_{ij} = R_{ik}R_{j\ell}\,(3\,\hat{r}_k\hat{r}_\ell - \delta_{k\ell}) = (RQ(\hat{r})R^\top)_{ij}.$$

Quindi

$$\begin{aligned} h'_i &= c(|\,r\,|)\,Q_{ij}(\hat{r}')\,f'_j = c(|\,r\,|)\,(RQR^\top)_{ij}\,R_{jm}f_m = c(|\,r\,|)\,R_{ik}\,Q_{k\ell}\,(\underbrace{R^\top R}_{I})_{\ell m}\,f_m \\ &= R_{ik}\,(Qf)_k = (Rh)_i. \end{aligned}$$

Ne segue $h' = Rh$. $\square$

Other implementations

ℓ_{out}	$K(\hat{\mathbf{r}}, r)$ (forma del kernel)	Azione su $f_b \in \mathbb{R}^3$	Canali irred. usati ($\ell_{\text{in}}^{=1} \otimes \ell_{\text{ker}} \to \ell_{\text{out}}$)
0	$K_b(\hat{\mathbf{r}}, r) = \alpha(r)\,\hat{r}_b\ (\ell_f = 1)$	$y = K_b f_b = \alpha(r)\,\hat{\mathbf{r}}\cdot\mathbf{f}$	$1 \otimes 1 \to 0 \quad (\ell_f = 1)$
1	$K_{ab}(\hat{\mathbf{r}}, r) = a(r)\,\delta_{ab}\ (\ell_f = 0) + b(r)\,\epsilon_{abc}\,\hat{r}_c\ (\ell_f = 1) + c(r)\,Q_{ab}(\hat{\mathbf{r}})\ (\ell_f = 2)$	$g_a = K_{ab}\,f_b$	$1 \otimes 0 \to 1$ $1 \otimes 1 \to 1$ $1 \otimes 2 \to 1$
2	$K_{ab,c}(\hat{\mathbf{r}}, r) = \alpha(r)\,\mathrm{ST}\big(\hat{r}_a\delta_{bc} + \hat{r}_b\delta_{ac}\big)\ (\ell_f = 1) + \beta(r)\,\mathrm{ST}\big(Q_{ab}(\hat{\mathbf{r}})\,\hat{r}_c\big)\ (\ell_f = 2)$	$T_{ab} = K_{ab,c}\,f_c$	$1 \otimes 1 \to 2$ $1 \otimes 2 \to 2$

Table: Kernel equivariante per $\ell_{\text{in}} = 1 \to \ell_{\text{out}} \in \{0, 1, 2\}$. Si usa $Q_{ab}(\hat{\mathbf{r}}) = 3\,\hat{r}_a\hat{r}_b - \delta_{ab}$ e $\mathrm{ST}(X_{ab}) = \frac{1}{2}(X_{ab} + X_{ba}) - \frac{1}{3}\delta_{ab}\,X_{cc}$. Le funzioni radiali α, β, a, b, c dipendono da $r = \|\mathbf{r}\|$. I tag $(\ell_f = \cdot)$ indicano il grado del pezzo angolare del kernel.

Dal continuo al discreto ($\mathbb{R}^3$ campionato)

$$(K \star f)(\mathbf{x}) = \int_{\mathbb{R}^3} K(\mathbf{x}-\mathbf{y})\, f(\mathbf{y})\, d\mathbf{y} \longrightarrow \sum_{j=1}^{N} K(\mathbf{x}-\mathbf{x}_j)\, f_j\, w_j.$$

- **Campionamento fisico**: in chimica e point-cloud il campo f è noto solo su un set finito di posizioni $\{\mathbf{x}_j\}$ (atomi, particelle, voxel).
- **Riemann sum**: l'integrale viene approssimato da una somma con pesi w_j (volume o 1 se densità uniforme).
- **Equivarianza preservata**: se ruoti insieme posizioni e feature, la stessa somma commuta con la rotazione come faceva l'integrale.
- **Efficienza**: evita la quadratura numerica su $SO(3)$; costo $\mathcal{O}(N)$ (o $\mathcal{O}(N^2)$ con tutti i vicini, riducibile via cutoff/sparse).

$$\boxed{\text{Integrale continuo} \approx \text{somma discreta su punti campionati}}$$

Confronto con Tensor Field Networks

	TFN (generale)	**Nostro caso**
Input/Output irreps	$\ell_{\text{in}}, \ell_{\text{out}}$ arbitr.	$\ell_{\text{in}} = \ell_{\text{out}} = 1$
Canali angolari J	$\lvert\ell_{\text{in}} - \ell_{\text{out}}\rvert \leq J \leq \ell_{\text{in}} + \ell_{\text{out}}$	$0, 1, 2$
Parte angolare	$\sum_m Y_J^m \, Q_{Jm}^{\ell k}$	$I, [\hat{r}]_\times, Q$
Radiale learnable	$\hat{\varphi}_J^{\ell k}(r)$	$a(r), b(r), c(r)$
Costo	$\mathcal{O}\big(NC_{in}C_{out}(2\ell_{\max}+1)^2\big)$	$\mathcal{O}\big(NC_{in}C_{out} \times 3\big)$

UNIVERSITÀ DI TRENTO

Sintesi

Il nostro layer è la versione *tascabile* di TFN: mantiene la correttezza teorica (equivarianza $SO(3)$) con soli tre canali angolari. La parte radiale è appresa via MLP 1-D, la parte angolare è codificata in tre tensori fissi – nessun campionamento del gruppo, costo costante.

That's all

Example: The antisymmetrizer

Intertwiner (Weiler & Cohen 2019)

The (unique up to scale) solution of (⋆) is the antisymmetriser

$$\boxed{A \;=\; \frac{1}{n!} \sum_{g \in S_n} \mathrm{sgn}(g)\, P_g}$$

Proof of equivariance.

$$A\, P_h = \frac{1}{n!} \sum_g \mathrm{sgn}(g)\, P_g P_h = \frac{1}{n!} \sum_g \mathrm{sgn}(gh)\, P_{gh}\ \mathrm{sgn}(h)^{-1}$$

$$= \mathrm{sgn}(h)\, \frac{1}{n!} \sum_{g'} \mathrm{sgn}(g')\, P_{g'} = \mathrm{sgn}(h)\, A,$$

which is exactly (⋆) with $W = A$. Because $A^2 = A$, A is a projection onto the totally antisymmetric (one-dimensional) irrep of S_n. Applying A to any tensor produces an antisymmetric object.
Generalisations Pfaffian ansatz for paired states (e.g. superconductors) is built analogously but starts from a two-body skew-symmetric kernel $F(x_i, x_j)$