

G-steerability of Equivariant Neural Networks

for (Hyper-)Nuclear Physics

Andrea Di Donna
May 28, 2025

andrea.didonna@unitn.it
TIFPA - Trento University



1 Due paradigmi di convoluzione \Rightarrow un unico obiettivo

■ Definizione di equivarianza

Una mappa lineare $F: \mathcal{F}_{\text{in}} \rightarrow \mathcal{F}_{\text{out}}$ è G -equivariante se

$$F(T_g f) = T_g F(f), \quad \forall g \in G, f \in \mathcal{F}_{\text{in}},$$

dove T_g è l'azione del gruppo sul feature field (nel nostro caso: traslazione dello spazio e rotazione del vettore).

■ Group Convolution

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

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

2 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*.

■ G-steerable Convolution (spazio fisico)

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

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 $K(R\hat{\mathbf{r}}) = R K(\hat{\mathbf{r}}) R^T$ implementa la stessa trasformazione indotta da \star_G , ma con costo computazionale ridotto.

$$K(\mathbf{r}) = \sum_{J=0}^2 \varphi_J(r) \sum_{m=-J}^J Y_{Jm}(\hat{\mathbf{r}}) Q_{Jm}^{\ell, \ell_I n}, \quad \hat{\mathbf{r}} = \frac{\mathbf{r}}{\|\mathbf{r}\|}.$$

- **Separazione radiale/angolare:**

$\varphi_J(r)$ è la parte *learnable* (solo raggio), $Y_{Jm}(\hat{\mathbf{r}})$ sono basi fisse sul S^2 .

- **Steerability:**

$$Y_{Jm}(R\hat{\mathbf{r}}) = \sum_{m'=-J}^J D_{m'm}^J(R) Y_{Jm'}(\hat{\mathbf{r}})$$
$$\implies K(R\hat{\mathbf{r}}) = R K(\hat{\mathbf{r}}) R^T.$$

- Q_{Jm} sono i coefficienti di Clebsch–Gordan che accoppiano $\ell_{\text{in}} = \ell_{\text{out}} = 1$ ai tre canali angolari $J = 0, 1, 2$.

$K \star f$ è automaticamente $SO(3)$ -equivariante

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

- 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.

Kernel TFN \equiv Kernel contratto (nostro)

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

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

- 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 $K(R\hat{\mathbf{r}}) = R K(\hat{\mathbf{r}}) R^T \quad \forall R \in SO(3)$.

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 $\hat{r}' = R\hat{r}$,

$$g' := b(|r|) [\hat{r}']_{\times} f' = Rg.$$

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 $[Rv]_{\times} = R[v]_{\times} R^T$. Allora

$$([\hat{r}']_{\times})_{ij} = R_{ip} R_{jq} \varepsilon_{pqr} \hat{r}_{\ell} = (R[\hat{r}]_{\times} R^T)_{ij}.$$

Quindi

$$g'_i = b(|r|) ([\hat{r}']_{\times})_{ij} f'_j = b(|r|) (R[\hat{r}]_{\times} R^T)_{ij} R_{jm} f_m = b(|r|) R_{ip} [\hat{r}]_{\times, pm} f_m = (Rg)_i,$$

dove abbiamo usato $R^T R = I$. Pertanto $g' = Rg$. □

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,

$$Q_{ij}(\hat{r}') = 3 \hat{r}'_i \hat{r}'_j - \delta_{ij} = 3 R_{ik} \hat{r}_k R_{j\ell} \hat{r}_\ell - \delta_{ij} = R_{ik} R_{j\ell} (3 \hat{r}_k \hat{r}_\ell - \delta_{k\ell}) = (RQ(\hat{r})R^T)_{ij}.$$

Quindi

$$\begin{aligned} h'_i &= c(|r|) Q_{ij}(\hat{r}') f'_j = c(|r|) (RQ(\hat{r})R^T)_{ij} R_{jm} f_m = c(|r|) R_{ik} Q_{k\ell} \underbrace{(R^T R)_{\ell m}}_I f_m \\ &= R_{ik} (Qf)_k = (Rh)_i. \end{aligned}$$

Ne segue $h' = Rh$. □

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

Table: Kernel equivariante per $\ell_{\text{in}} = 1 \rightarrow \ell_{\text{out}} \in \{0, 1, 2\}$. Si usa $Q_{ab}(\hat{\mathbf{r}}) = 3 \hat{r}_a \hat{r}_b - \delta_{ab}$ e $\text{ST}(X_{ab}) = \frac{1}{2}(X_{ab} + X_{ba}) - \frac{1}{3} \delta_{ab} X_{cc}$. 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.

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

- **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 $O(N)$ (o $O(N^2)$ con tutti i vicini, riducibile via cutoff/sparse).

Integrale continuo \approx somma discreta su punti campionati

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

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.