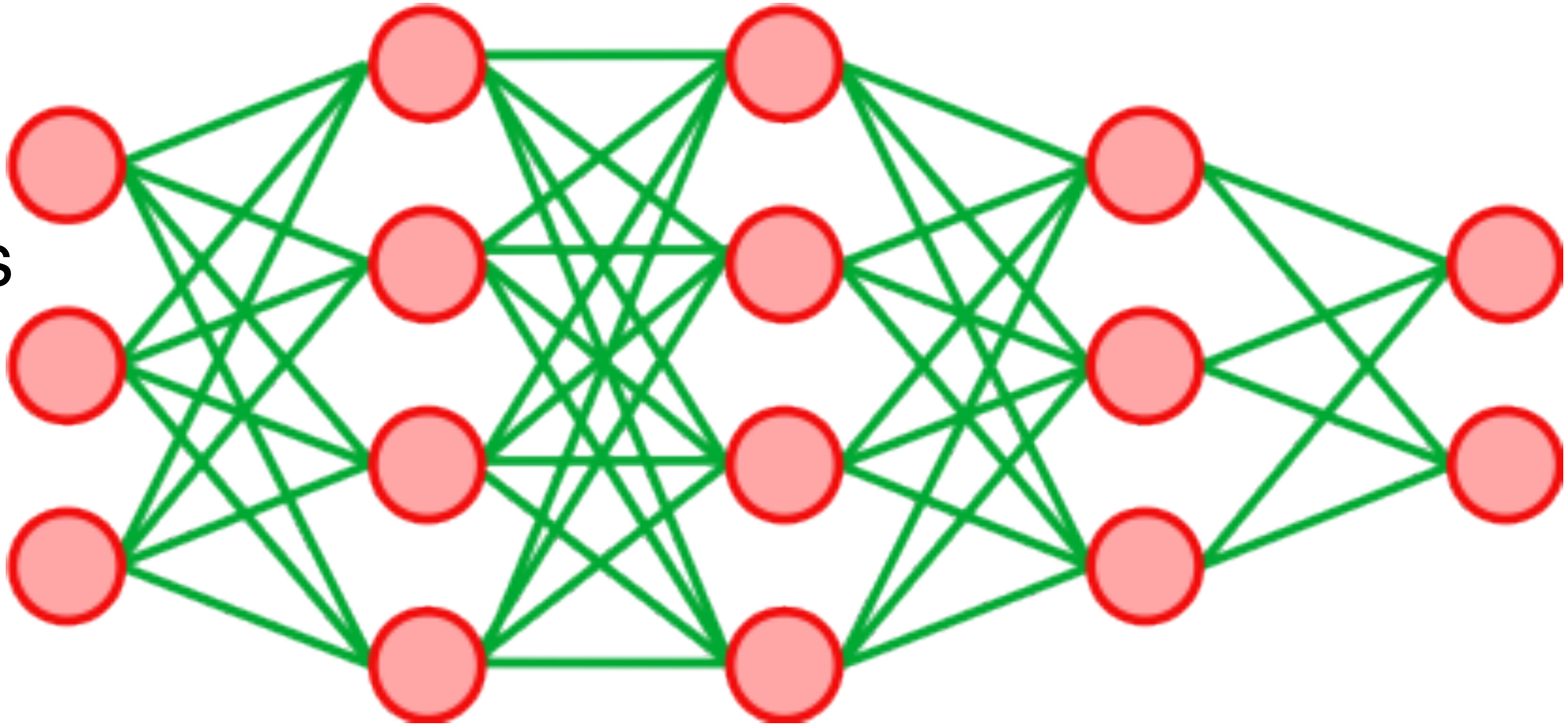
On how to incorporate spacetime symmetries into your neural networks

Patrick Forré

AI4Science Lab, Korteweg-de Vries Institute for Mathematics, University of Amsterdam, The Netherlands

Deep Neural Network (DNN)

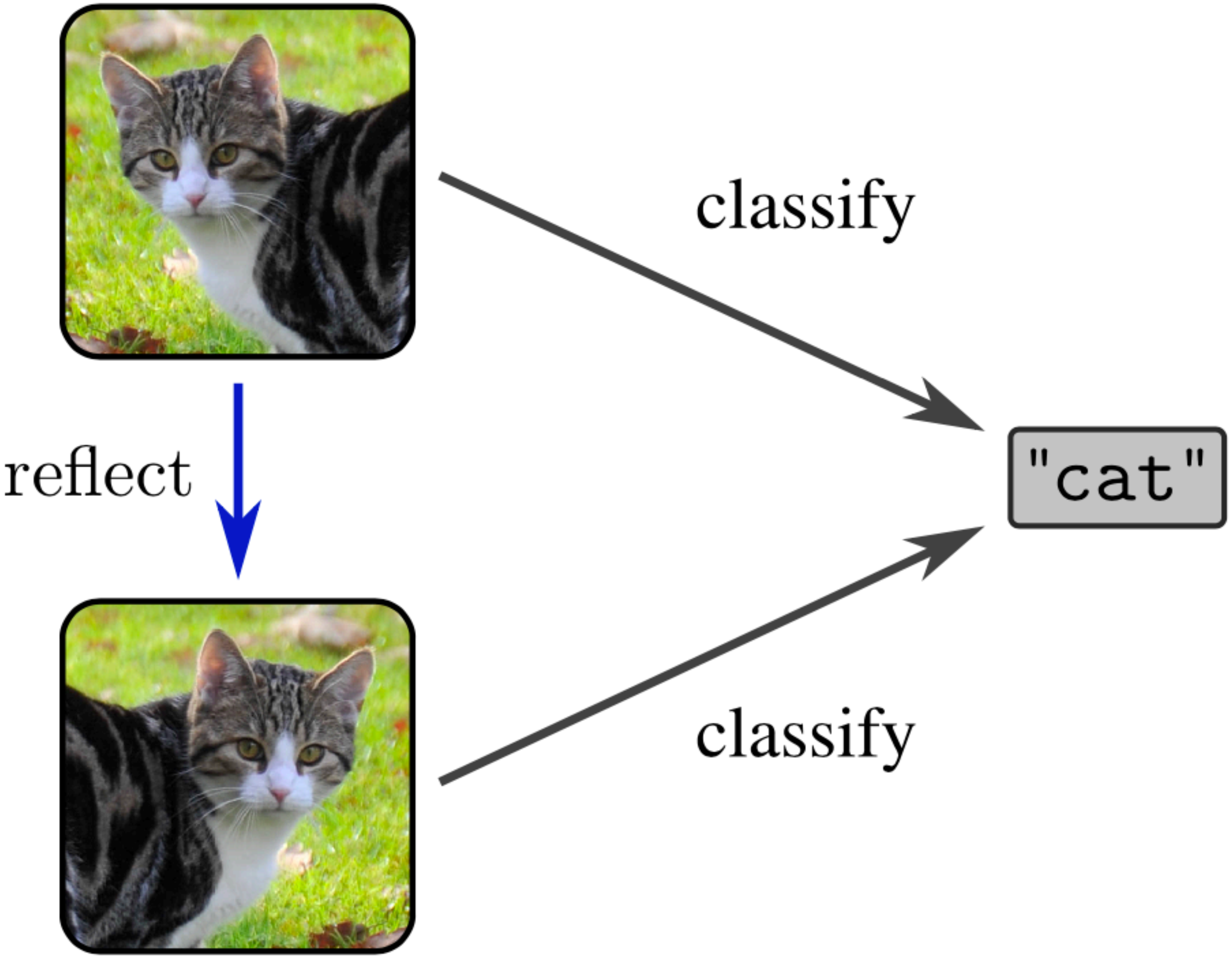
- Input: features



- Output: prediction

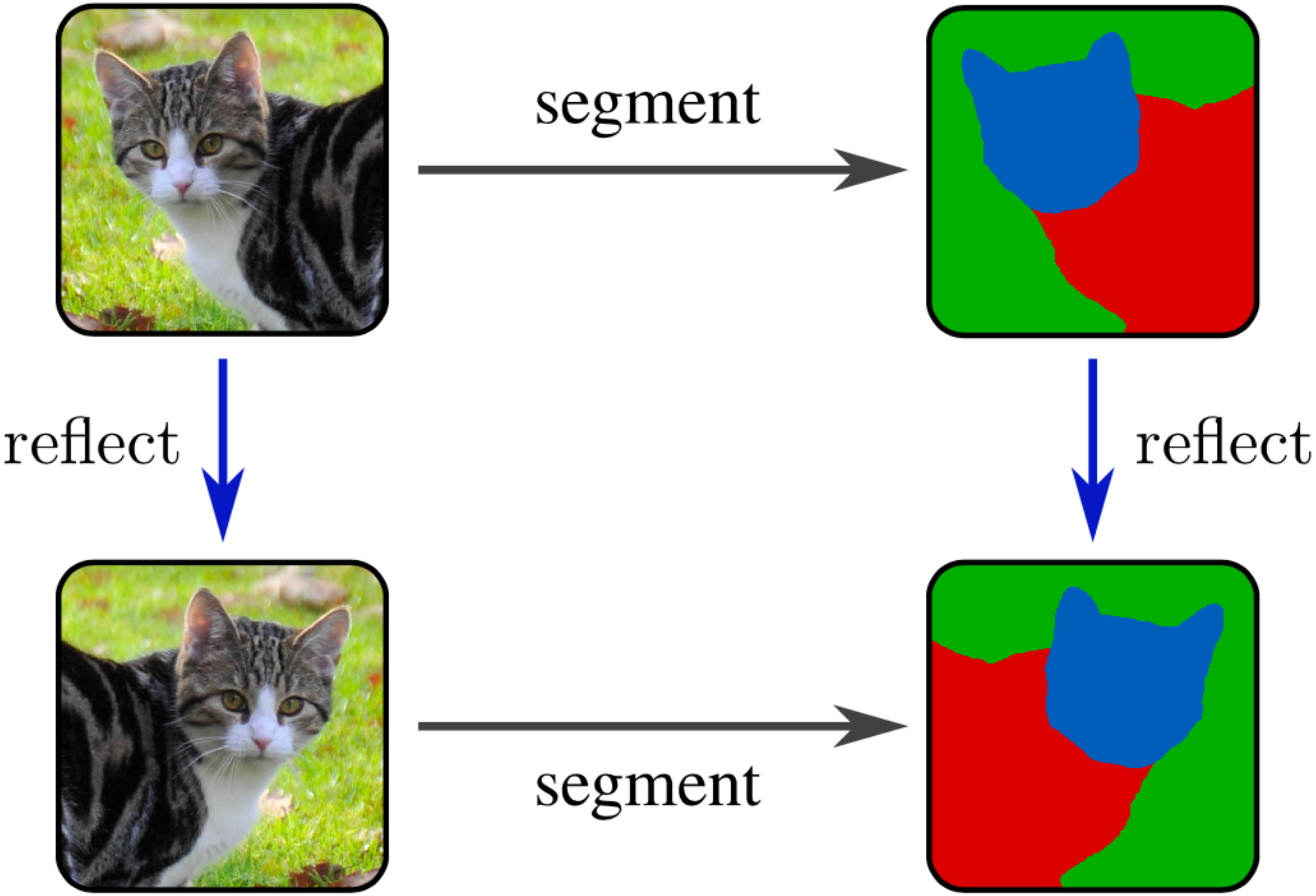
- $F: V_0 \xrightarrow{L_1} V_1 \xrightarrow{L_2} \cdots \xrightarrow{L_l} V_l \xrightarrow{L_{l+1}} \cdots \xrightarrow{L_\ell} V_\ell$
  - V_l finite dimensional vector spaces
  - L_l (affine) linear or simple non-linear maps

Invariant predictions tasks



- vanilla DNN could make different predictions when input is rotated or reflected
- goal: build symmetry of input space directly into the DNN

Equivariant prediction tasks



- in such tasks the output should transform correspondingly to the input
- goal: build input/output symmetry directly into your DNNs

Equivariant Deep Neural Network

- DNN: $F: V_0 \xrightarrow{L_1} V_1 \xrightarrow{L_2} \cdots \xrightarrow{L_l} V_l \xrightarrow{L_{l+1}} \cdots \xrightarrow{L_\ell} V_\ell$
- Let G be a group and for each $l = 0,1,\ldots,\ell$ fix:
  - $\rho_l : G \times V_l \to V_l, \quad (g, v) \mapsto \rho_l(g)(v)$ a group action
  - We call (ρ_l, V_l) a **(linear) representation for G**
- The DNN F is G-equivariant if every layer is G-equivariant.
- So focus on one **linear layer** (first): $L: V_{\text{in}} \to V_{\text{out}}$,
  - Equivariance means: $\forall g \in G . \forall v \in V_{\text{in}}$.

$$L\big(\rho_{\text{in}}(g)(v)\big) = \rho_{\text{out}}(g)\big(L(v)\big), \quad \text{or in short:} \quad L \circ \rho_{\text{in}}(g) = \rho_{\text{out}}(g) \circ L$$

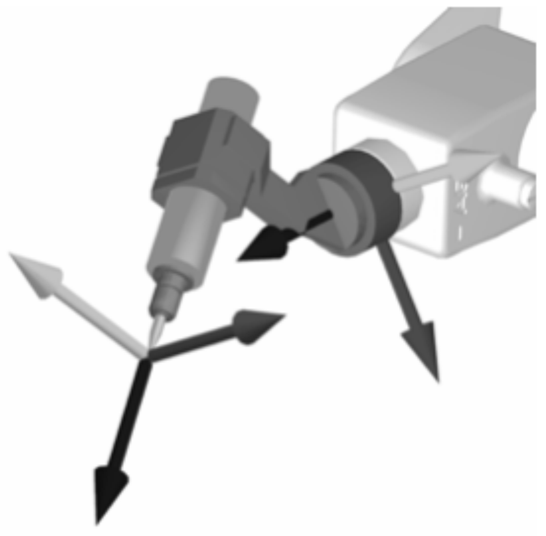
Problems

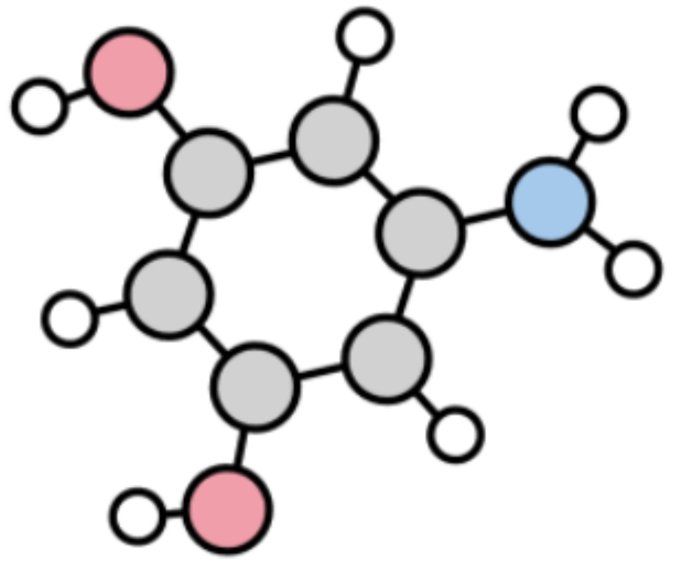
- For every different group G need to study and classify:
  - (irreducible) representations $\rho: G \times V \to V$
  - this explicitly known for finite and compact groups G
  - mainly for $\mathrm{SO}(2), \mathrm{O}(2), \mathrm{SO}(3), \mathrm{O}(3)$ (circular and spherical harmonics)
- To increase expressivity or as partial result often needed:
  - tensor product representations: $V_i \otimes V_j$
  - need to decompose into irreducible representation again:

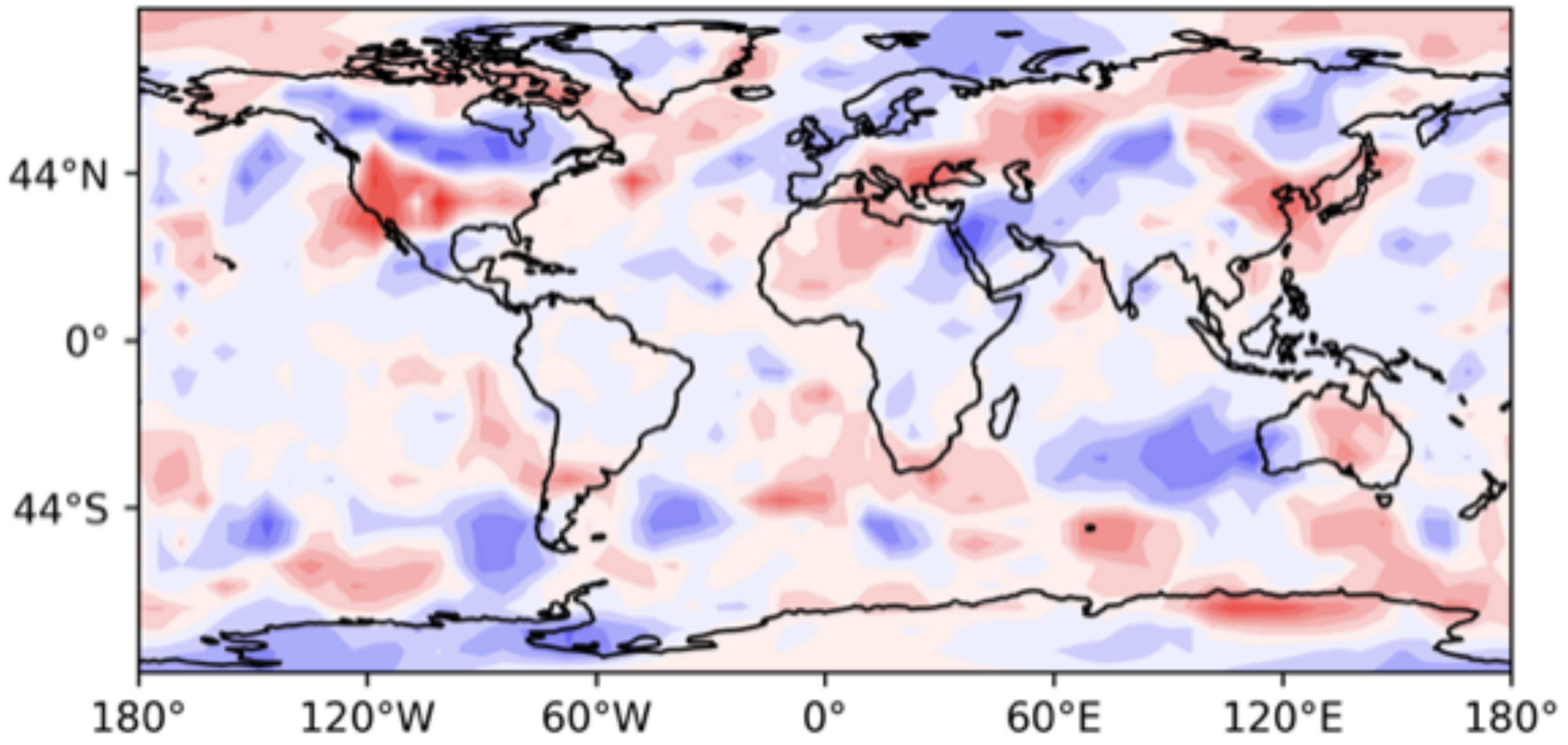
$$V_i \otimes V_j \cong \bigoplus_{k \in K} W_k$$ and the matrices that project on such W_k

Interest in Orthogonal and Pseudo-Orthogonal Transformations

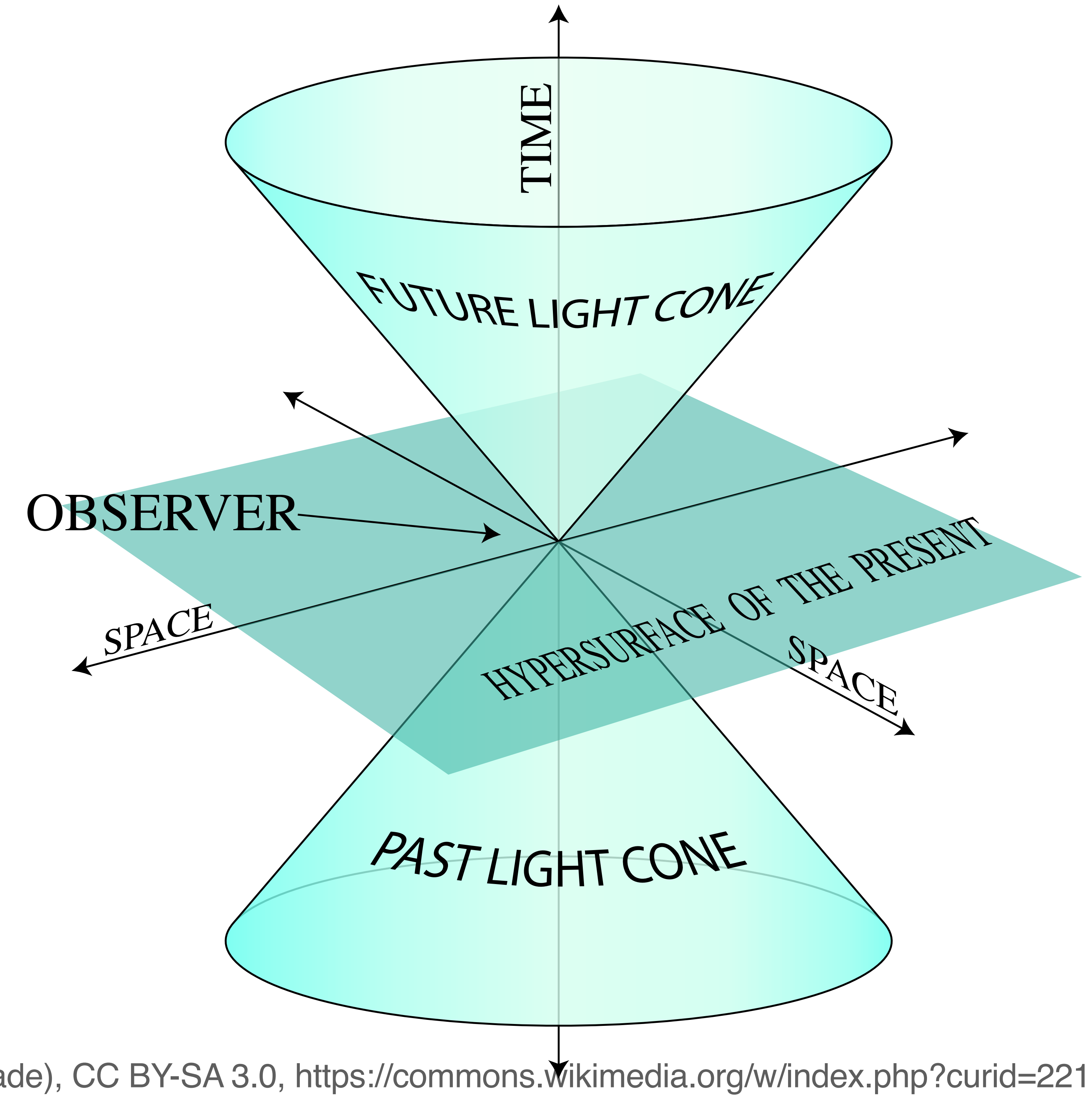
- Application area of **orthogonal transformations**:
  - image and video processing
  - molecular dynamics
  - fluid dynamics
  - robotics







- Application are of **pseudo-orthogonal transformations**:
  - Lorentz boosts on Minkowski space (relativistic spacetime)
  - mainly in special relativistic physics applications:
    - particle physics
    - electrodynamics
    - relativistic hydrodynamics



Problems

- **Pseudo-orthogonal group** $\mathrm{O}(p,q)$ for $p,q \geq 1$
    - not compact, not connected matrix group
    - irreps are difficult to implement and understand for ML community
- Goal: find convenient (probably not irreducible) representations such that:
    - basis easy to get (not function spaces),
    - action easy to compute,
    - are flexible and expressive,
    - their dimension can be controlled (for computational purposes),
    - work in all dimensions $d \geq 1$ and signatures (p,q).

Proposed Solution:

The Clifford Algebra

Clifford Group Equivariant Neural Networks

David Ruhe
AI4Science Lab, AMLab, API
University of Amsterdam
david.ruhe@gmail.com

Johannes Brandstetter
Microsoft Research
AI4Science
brandstetter@ml.jku.at

Patrick Forré
AI4Science Lab, AMLab
University of Amsterdam
p.d.forre@uva.nl

- Presented at NeurIPS 2023 (oral)
- https://github.com/DavidRuhe/clifford-group-equivariant-neural-networks

Inner Product Spaces

- Let V be a finite dimensional vector space over $\mathbb{R}$ with $d = \dim V$.
- Let $\eta : V \times V \to \mathbb{R}$ be a non-degenerate, symmetric, bilinear form on V.
- We then call (V, η) a **inner product space**
- We abbreviate the corresponding quadratic form:

$$\mathbf{q} : V \to \mathbb{R}, \qquad \mathbf{q}(v) := \eta(v, v)$$

- Note we do **not** require η to be (positive-semi) definite !

Orthonormal Basis

- Let (V, η) be a inner product space with elements $e_1, \dots, e_d \in V$.
- $e = [e_1, \dots, e_d]$ is called an **orthonormal basis** of (V, η) if:
    - $\eta(e_i, e_j) = 0$ for all $i \neq j$ in $[d] := \{1, \dots, d\}$
    - $\eta(e_i, e_i) \in \{\pm 1\}$ for all $i \in [d]$.
- Note that such bases always exist (over $\mathbb{R}$) and the numbers (p, q) of $+1, -1$'s is an invariant of (V, η), called the **signature of** (V, η).

Example - Standard Pseudo-Euclidean Spaces

- Let $V = \mathbb{R}^d$ and $p, q \in \mathbb{N}_0$ with $p + q = d$.
- Consider the $d \times d$ diagonal matrix:

$$\Delta^{p,q} := \mathrm{diag}(\underbrace{+1,\ldots,+1}_{\times p}, \underbrace{-1,\ldots,-1}_{\times q})$$

- and symmetric, bilinear form $\eta^{p,q}$ on (column) vectors $v_1, v_2 \in \mathbb{R}^d$ as:

$$\eta^{p,q}(v_1, v_2) := v_1^\top \Delta^{p,q} v_2$$

- We abbreviate: $\mathbb{R}^{p,q} := (\mathbb{R}^d, \eta^{p,q})$ and call it the **Standard (Pseudo-)Euclidean Space of signature** (p, q)**.**

The Clifford Algebra - Ad Hoc Definition

- Let (V, η) be <u>inner product space</u> with orthonormal basis $e = [e_1, \ldots, e_d]$.
- We define the **Clifford algebra** (aka **geometric algebra**) $\mathrm{Cl}(V, \eta)$ of (V, η) as the vector space spanned by the following formal basis elements:
    - $1_{\mathrm{Cl}(V,\eta)}$
    - $e_1, \ldots, e_d$
    - $e_{1,2}, e_{1,3}, \ldots, e_{i_1,i_2}, \ldots, e_{d-1,d}$ with $i_1 < i_2$
    - $e_{1,2,3}, \ldots, e_{i_1,i_2,i_3}, \ldots, e_{d-2,d-1,d}$ with $i_1 < i_2 < i_3$
    - etc.
    - $e_{1,2,\ldots,d}$
- Note that there are no repeated indices in each element and they need to be in order.

The Clifford Algebra - Short Notations

- We can abbreviate the notation as follows.
- Define for a subset $A \subseteq [d]$ of size $k > 0$ the element:
  - $e_A := e_{i_1,\dots,i_k}$ with $A = \{i_1 < \cdots < i_k\}$,
  - $e_\emptyset := 1_{\mathrm{Cl}(V,\eta)}$
- We interpret the basis elements as products of basis vectors:
  - $e_A = e_{i_1,\dots,i_k} = e_{i_1} \bullet \cdots \bullet e_{i_k}$

Clifford Algebra - The Dimension

- In particular, we have:

$$\dim \mathrm{Cl}(V, \eta) = |\{A \subseteq [d]\}| = 2^d,$$

- and, every element $x \in \mathrm{Cl}(V, \eta)$ can <u>uniquely</u> be written as:

$$x = \sum_{A \subseteq [d]} x_A \cdot e_A \qquad \text{with} \quad x_A \in \mathbb{R}.$$

Geometric Product via the Fundamental Relations

- Let (V, η) be inner product space with orthonormal basis $e = [e_1, \ldots, e_d]$ of signature (p, q). We then extend the **geometric product** by the following relations:

$$e_j \bullet e_i = \begin{cases} -e_i \bullet e_j & \text{if } i < j, \\ +1 & \text{if } i = j \in [1, p], \\ -1 & \text{if } i = j \in [p+1, p+q]\,. \end{cases}$$

The Fundamental Relation

- Let $v_1, v_2 \in V$ then we have the **following fundamental relation** considered as elements in $\mathrm{Cl}(V, \eta)$:

$$v_2 \bullet v_1 = -v_1 \bullet v_2 + 2 \cdot \eta(v_1, v_2) \cdot 1_{\mathrm{Cl}(V,\eta)}$$

The Geometric Product for General Elements

- We can now turn the **Clifford algebra** $\mathrm{Cl}(V, \eta)$ into an algebra via extending the **geometric product**:

$$x \bullet y = \left(\sum_{A \subseteq [d]} x_A \cdot e_A \right) \bullet \left(\sum_{B \subseteq [d]} y_B \cdot e_B \right)$$

$$= \sum_{A \subseteq [d]} \sum_{B \subseteq [d]} (x_A \cdot y_B) \cdot (e_A \bullet e_B)$$

- The element $e_A \bullet e_B$ can now be simplified with the above relations.

Example: The Geometric Product

- Assume $\eta(e_3, e_3) = 1$, then we have:
  - $e_2 \bullet e_3 \bullet e_1 \bullet e_3 = - e_2 \bullet e_1 \bullet e_3 \bullet e_3 = - e_2 \bullet e_1 = e_1 \bullet e_2 = e_{1,2}$.
  - So we get a standard basis element back.
- The same trick can be used to simplify any element $e_A \bullet e_B$:

$$e_A \bullet e_B = \operatorname{sgn}^{A,B} \cdot \prod_{i \in A \cap B} \eta(e_i, e_i) \cdot e_{A \triangle B}$$

Extending the Inner Product

- Let $e = [e_1, \dots, e_d]$ be an orthonormal basis of inner product space (V, η).

- Then the system $(e_A)_{A \subseteq [d]}$ forms an **orthonormal basis** of $\mathrm{Cl}(V, \eta)$ w.r.t. the **extended inner product** $\bar{\eta}$ defined via:

$$\bar{\eta}(e_A, e_B) := \delta_{A,B} \cdot \prod_{i \in A} \eta(e_i, e_i).$$

- This turns $\mathrm{Cl}(V, \eta)$ into an inner product space in itself w.r.t. $\bar{\eta}$.

- Furthermore, $\bar{\eta}$ is independent of the chosen orthonormal basis.

k-Vectors

- Let $e = [e_1, \dots, e_d]$ be an <u>orthogonal basis</u> of the quadratic vector space (V, η). For $k = 0, \dots, d$, we define the **grade-k-subspace** of $\mathrm{Cl}(V, \eta)$ as follows:

$$\mathrm{Cl}^{(k)}(V, \eta) := \operatorname{span}\left\{e_A \mid A \subseteq [d], |A| = k\right\}$$

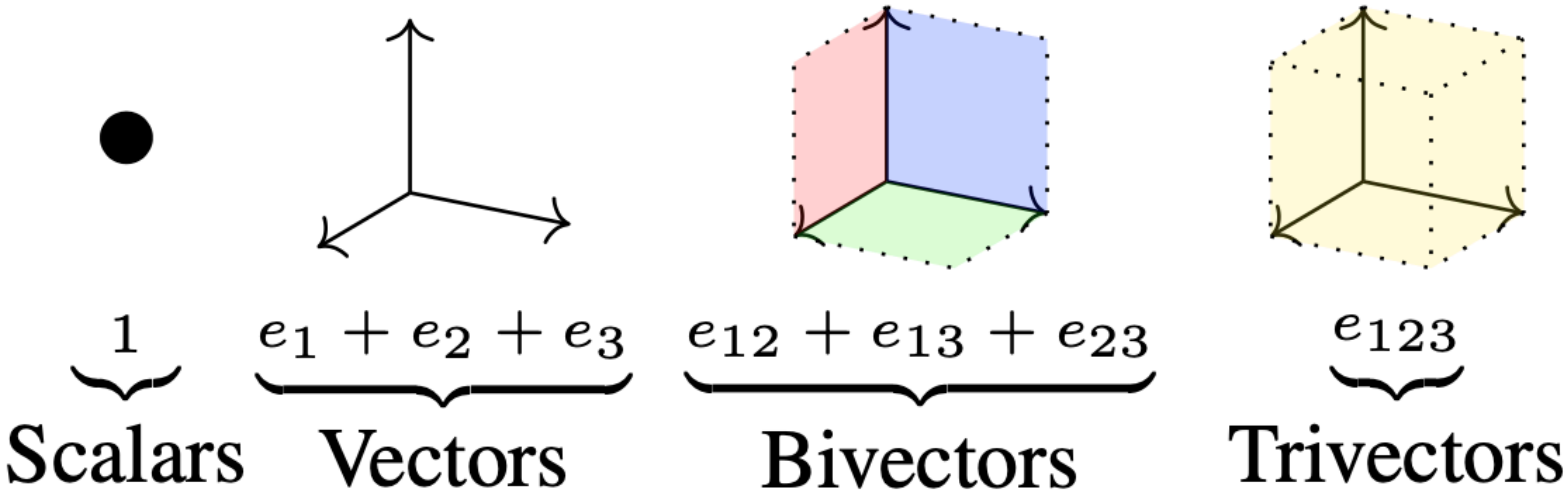
consisting of all k**-vectors** (aka k**-blade**), i.e. elements of the form:

$$x = \sum_{A \subseteq [d], |A| = k} x_A \cdot e_A.$$

- <u>Theorem</u>: $\mathrm{Cl}^{(k)}(V, \eta)$ does <u>not</u> depend on the chosen <u>orthogonal basis.</u>

Example: k-Vectors

- We can identify:
  - $\mathrm{Cl}^{(0)}(V, \eta) = \mathbb{R}$ i.e. 0-vectors = scalars
  - $\mathrm{Cl}^{(1)}(V, \eta) = V$ i.e. 1-vectors = vectors



Grade Decomposition and Grade Projection

- We thus have an **orthogonal direct sum decomposition**:

- $$\mathrm{Cl}(V,\eta) = \bigoplus_{k=0}^{d} \mathrm{Cl}^{(k)}(V,\eta), \qquad \dim \mathrm{Cl}^{(k)}(V,\eta) = \binom{d}{k}$$

- an **orthogonal projection map** onto the <u>grade-k-component</u>:

$$(\_)^{(k)} : \mathrm{Cl}(V,\eta) \to \mathrm{Cl}^{(k)}(V,\eta),$$

$$x = \sum_{A \subseteq [d]} x_A \cdot e_A \mapsto \sum_{A \subseteq [d], |A| = k} x_A \cdot e_A =: x^{(k)}.$$

The (Pseudo-)Orthogonal Group

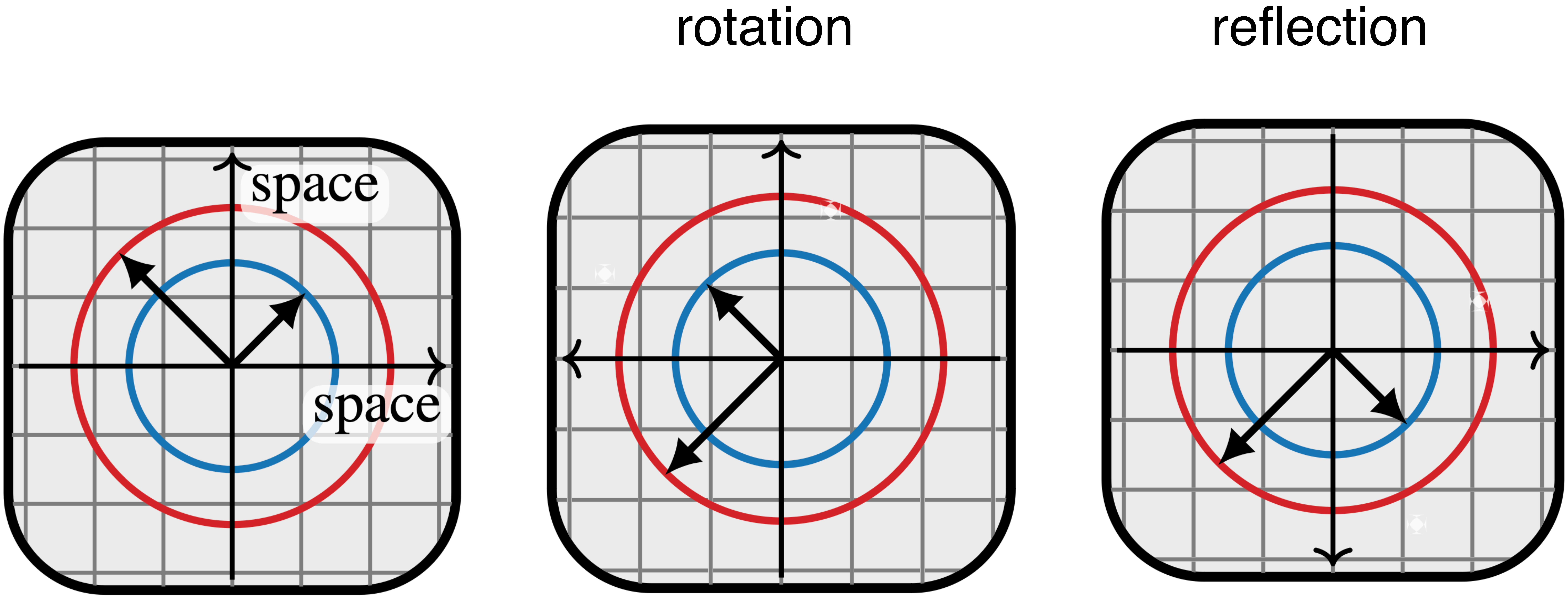
- Let (V, η) be a quadratic vector space / inner product space.
- The **(pseudo-)orthogonal group** of (V, η) is defined as:

$$\mathrm{O}(V, \eta) := \left\{ g \in \mathrm{GL}(V) \mid \forall v \in V\,.\ \eta(gv, gv) = \eta(v, v) \right\}.$$

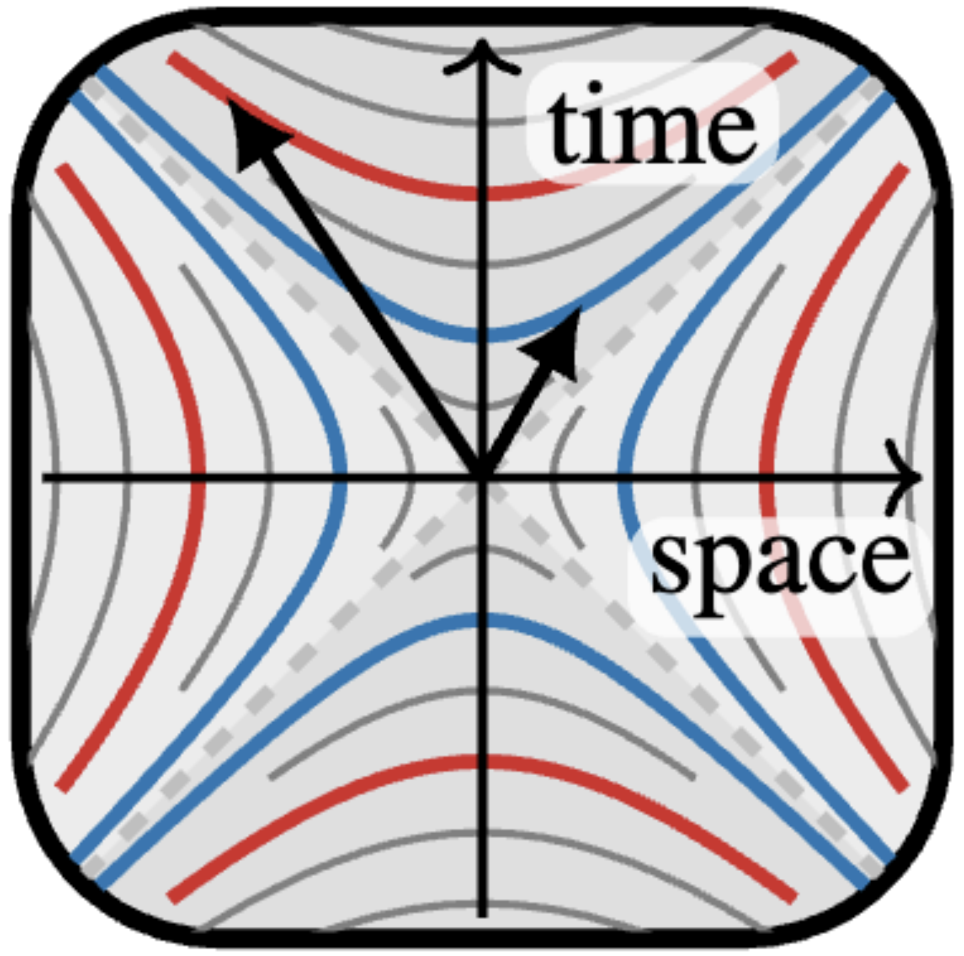
- The **special (pseudo-)orthogonal group** of (V, η) is:

$$\mathrm{SO}(V, \eta) := \left\{ g \in \mathrm{O}(V, \eta) \mid \det g = 1 \right\}.$$

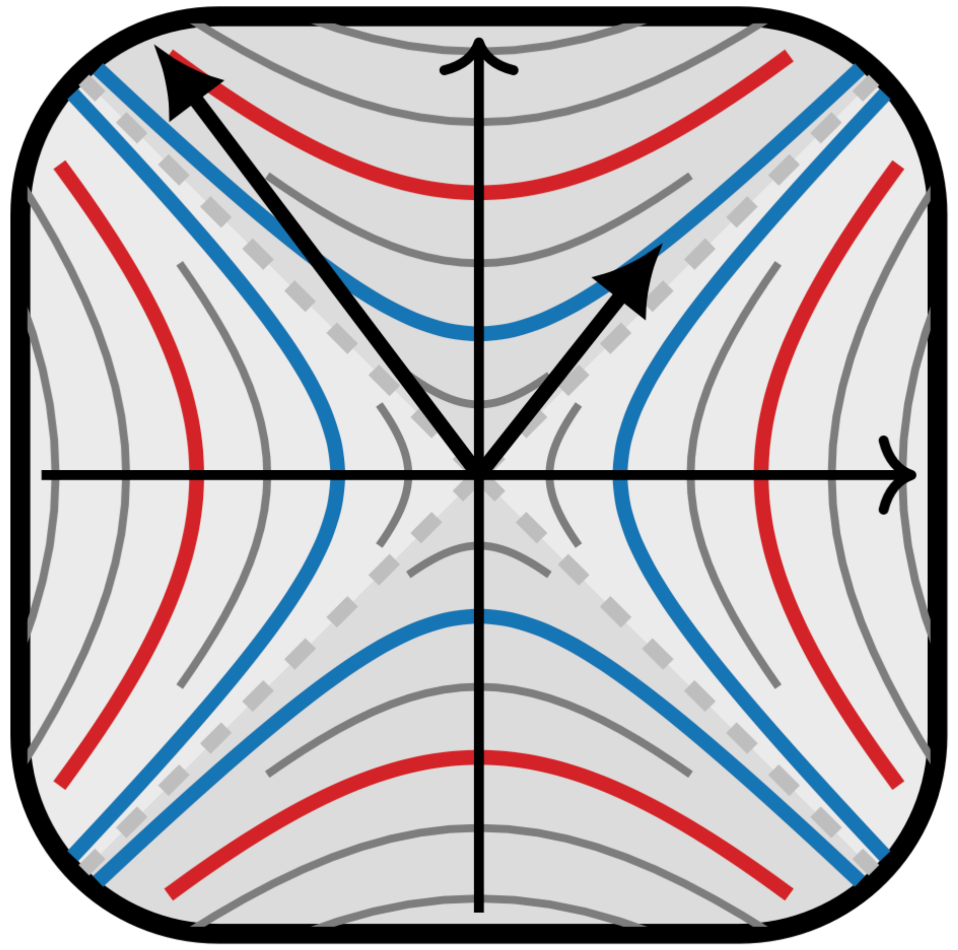
Example: Orthogonal Transformation in Euclidean Space



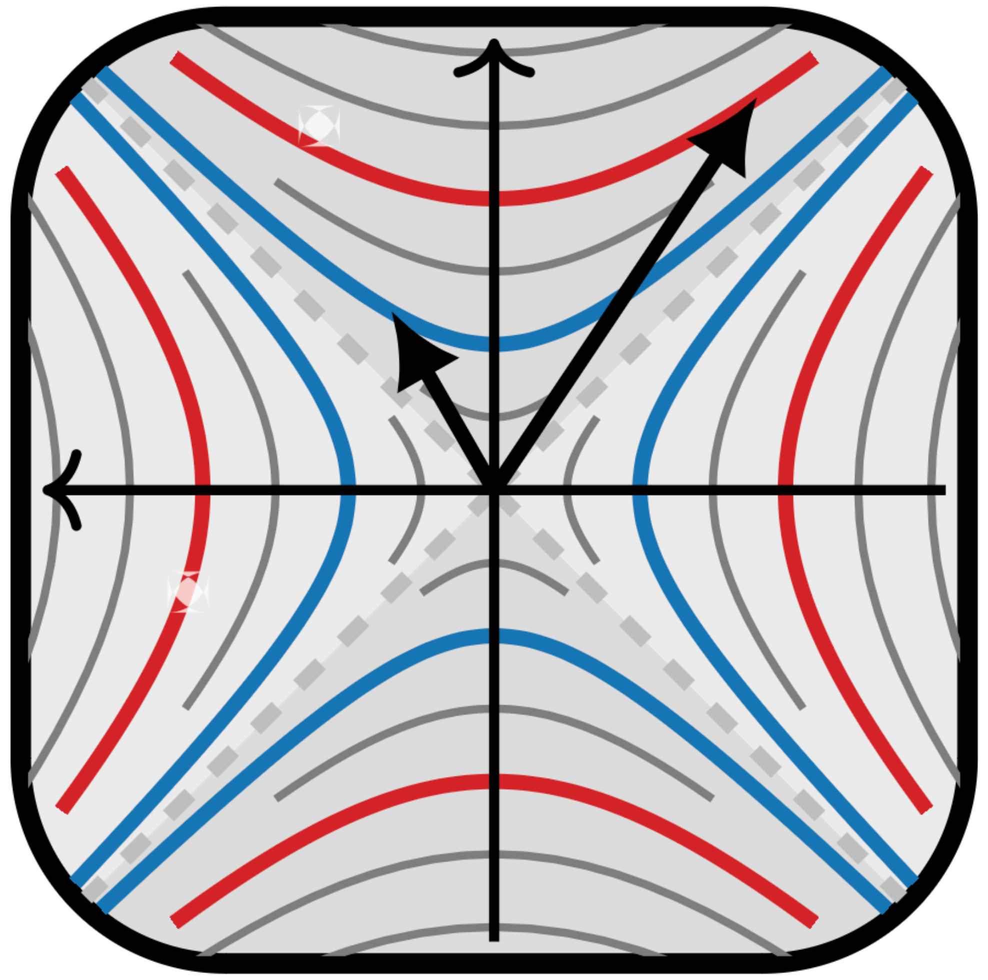
Example: Orthogonal Transformation in Pseudo-Euclidean Space



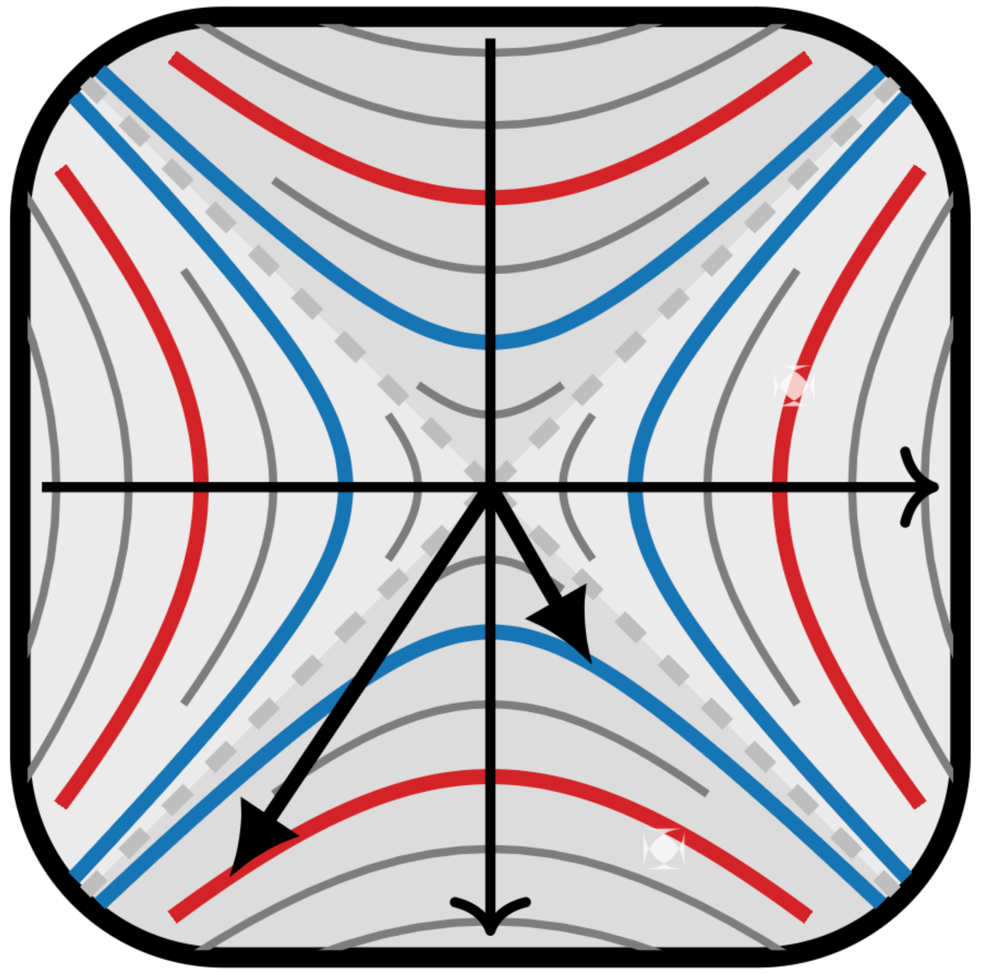
Lorentz boost



space reflection



time reflection



(+ rotations)

The Orthogonal Group Action

- Let (V, η) be a quadratic vector space / inner product space and

$$x = \sum_{i \in I} c_i \cdot v_{i,1} \bullet \cdots \bullet v_{i,l_i} \in \mathrm{Cl}(V, \eta) \quad \text{and} \quad g \in \mathrm{O}(V, \eta).$$

- Then we define the action of g on x as follows:

$$\rho_{\mathrm{Cl}}(g)(x) := \sum_{i \in I} c_i \cdot (g v_{i,1}) \bullet \cdots \bullet (g v_{i,l_i}) \quad \in \mathrm{Cl}(V, \eta)$$

- By the universal property this is well-defined and gives action:

$$\rho_{\mathrm{Cl}} : \ \mathrm{O}(V, \eta) \times \mathrm{Cl}(V, \eta) \to \mathrm{Cl}(V, \eta), \quad (g, x) \mapsto \rho_{\mathrm{Cl}}(g)(x).$$

Orthogonal Group Action - Properties

- We have the following properties for ρ_{Cl}:

- linear: $\rho_{\mathrm{Cl}}(g)(c_1 \cdot x_1 + c_2 \cdot x_2) = c_1 \cdot \rho_{\mathrm{Cl}}(g)(x_1) + c_2 \cdot \rho_{\mathrm{Cl}}(g)(x_2)$
- composing: $\rho_{\mathrm{Cl}}(g_2)\left(\rho_{\mathrm{Cl}}(g_1)(x)\right) = \rho_{\mathrm{Cl}}(g_2 g_1)(x)$
- invertible: $\rho_{\mathrm{Cl}}(g)^{-1}(x) = \rho_{\mathrm{Cl}}(g^{-1})(x)$
- orthogonal: $\bar{\eta}\left(\rho_{\mathrm{Cl}}(g)(x_1), \rho_{\mathrm{Cl}}(g)(x_2)\right) = \bar{\eta}(x_1, x_2)$
- multiplicative: $\rho_{\mathrm{Cl}}(g)(x_1 \bullet x_2) = \rho_{\mathrm{Cl}}(g)(x_1) \bullet \rho_{\mathrm{Cl}}(g)(x_2)$
- grade preserving: $\left(\rho_{\mathrm{Cl}}(g)(x)\right)^{(k)} = \rho_{\mathrm{Cl}}(g)\left(x^{(k)}\right)$

Corollary

- So the following maps are $\mathrm{O}(V,\eta)$-**equivariant**:
  - <u>Geometric product.</u>

$$
\begin{tikzcd}
\mathrm{Cl}(V,\eta) \times \mathrm{Cl}(V,\eta) \arrow[r, "\bullet"] \arrow[d, "\rho_{\mathrm{Cl}}(g) \times \rho_{\mathrm{Cl}}(g)"'] & \mathrm{Cl}(V,\eta) \arrow[d, "\rho_{\mathrm{Cl}}(g)"] \\
\mathrm{Cl}(V,\eta) \times \mathrm{Cl}(V,\eta) \arrow[r, "\bullet"'] & \mathrm{Cl}(V,\eta)
\end{tikzcd}
$$

  - <u>Grade projections.</u>

$$
\begin{tikzcd}
\mathrm{Cl}(V,\eta) \arrow[r, "(\,\cdot\,)^{(k)}"] \arrow[d, "\rho_{\mathrm{Cl}}(g)"'] & \mathrm{Cl}^{(k)}(V,\eta) \arrow[d, "\rho_{\mathrm{Cl}}(g)"] \\
\mathrm{Cl}(V,\eta) \arrow[r, "(\,\cdot\,)^{(k)}"'] & \mathrm{Cl}^{(k)}(V,\eta)
\end{tikzcd}
$$

    - $\mathrm{Cl}^{(k)}(V,\eta)$ are thus sub-representations!

Clifford Group Equivariant Neural Networks

- A **Clifford Group Equivariant Neural Network (CGENN)** is a composition:

$$F: V_0 \xrightarrow{L_1} V_1 \xrightarrow{L_2} \cdots \xrightarrow{L_l} V_l \xrightarrow{L_{l+1}} \cdots \xrightarrow{L_\ell} V_\ell$$

where $V_l = \mathrm{Cl}(V, \eta)^{c_l}$ and $L_l = L: V_{\text{in}} \to V_{\text{out}}$ is either:

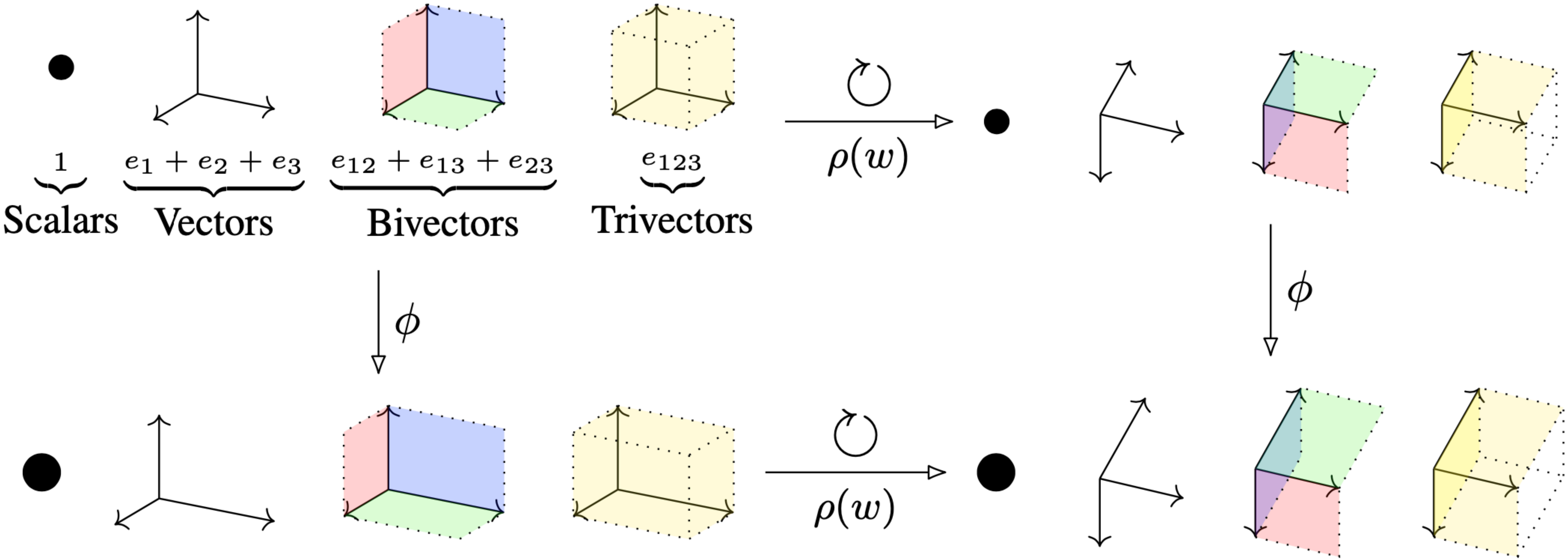
- **Linear layer**: $\forall m \in [c_{\text{out}}]$: $L_m^{(k)}(x_1, \ldots, x_{c_{\text{in}}}) := \sum_{n=1}^{c_{\text{in}}} w_{m,n}^k \cdot x_n^{(k)}$,

- **Geometric Product layer**: $P^{(k)}(x_1, x_2) := \sum_{m=0}^{d} \sum_{n=0}^{d} w_{m,n}^k \cdot \left(x_1^{(m)} \bullet x_2^{(n)}\right)^{(k)}$

- **Nonlinearity**: $A(x) := \sigma(x^{(0)}) \cdot x$, or $A^{(k)}(x^{(k)}) = \sigma(\bar{\mathfrak{q}}(x^{(k)})) \cdot x^{(k)}$.

Theorem - CGENN are $\mathrm{O}(V, \eta)$-equivariant

- <u>Theorem</u>: All CGENN are automatically $\mathrm{O}(V, \eta)$-equivariant.



Remark - CGENNs

- Note that the geometric product layers are non-linear, but still $\mathrm{O}(V, \eta)$-equivariant. We could also simplify them to: $P(x_1, x_2) = w \cdot x_1 \bullet x_2$
- Note that we can just replace all occurring deep learning architectures with corresponding Clifford layers to make them $\mathrm{O}(V, \eta)$-equivariant,
  - e.g. graph or message-passing neural networks.

CGENN - Experiments

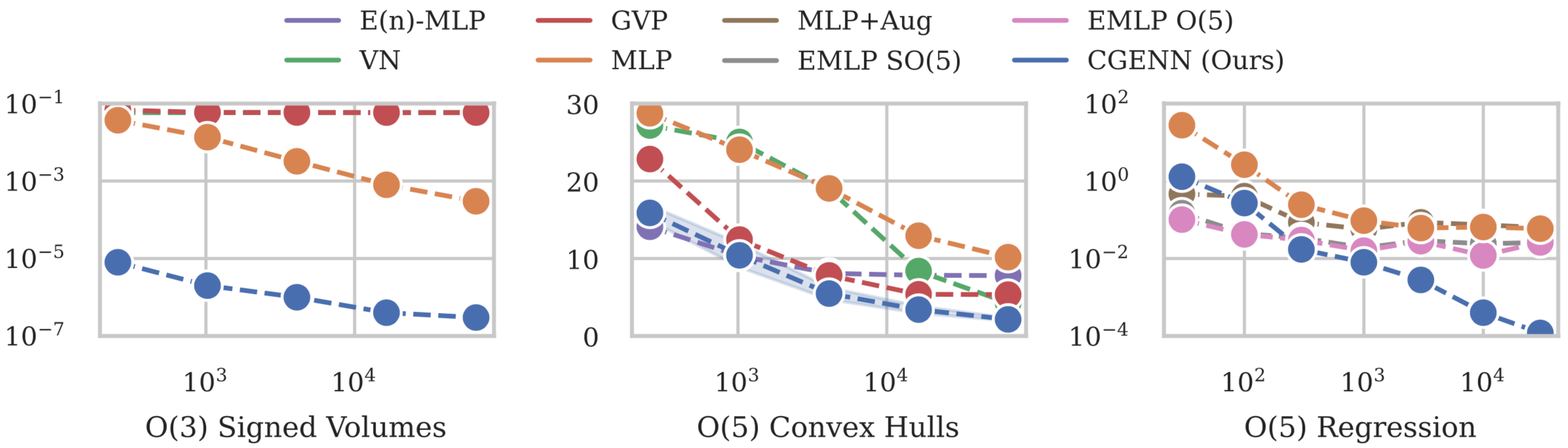


Figure 3: Left: Test mean-squared errors on the $O(3)$ signed volume task as functions of the number of training data. Note that due to identical performance, some baselines are not clearly visible. Right: same, but for the $O(5)$ convex hull task.

Figure 4: Test mean-squared-errors on the $O(5)$ regression task.

CGENN - Top Tagging: $\mathrm{O}(1,3)$-Experiment

- Particle collider experiment at CERN’s ATLAS detector
- produces and measures particle jets
- goal: **predict** if particle jet contains **top quark**
- each jet has ~200 particles, each given with energy-momentum vector
- those live in relativistic spacetime = Minkoswki space $V = \mathbb{R}^{1,3}$
- due to special relativity, prediction should be invariant under Lorenz boosts, i.e. under action of the Lorentz group $\mathrm{O}(1,3)$.
- train/validation/test set: 1.2M, 0.4M, 0.4M entries
- only LGN is steerable, EGNN and LorentzNet scalarized (invariant), rest not

CGENN - Top Tagging: $\mathrm{O}(1,3)$-Experiment

Model	Accuracy (↑)	AUC (↑)	$1/\epsilon_B$ (↑) ($\epsilon_S = 0.5$)	$1/\epsilon_B$ (↑) ($\epsilon_S = 0.3$)
ResNeXt [XGD+17]	0.936	0.9837	302	1147
P-CNN [CMS17]	0.930	0.9803	201	759
PFN [KMT19]	0.932	0.9819	247	888
ParticleNet [QG20]	0.940	0.9858	397	1615
EGNN [SHW21]	0.922	0.9760	148	540
LGN [BAO+20]	0.929	0.9640	124	435
LorentzNet [GMZ+22]	**0.942**	**0.9868**	**498**	**2195**
CGENN	**0.942**	**0.9869**	**500**	**2172**

Table 2: Performance comparison between our proposed method and alternative algorithms on the top tagging experiment. We present the accuracy, Area Under the Receiver Operating Characteristic Curve (AUC), and background rejection $1/\epsilon_B$ and at signal efficiencies of $\epsilon_S = 0.3$ and $\epsilon_S = 0.5$.

- Minkowski metric: $(1, -1, -1, -1)$, group: $G = \mathrm{O}(1,3)$

Equivariant Simplicial Message Passing

Algorithm 1 Shared Simplicial Message Passing

Require: $K, \forall \sigma \in K : h^\sigma, \phi^m, \phi^h$

Repeat:

$$m^\sigma \leftarrow \operatorname{Agg}_{\substack{\tau \in B(\sigma) \\ \tau \in C(\sigma) \\ \tau \in N_\uparrow(\sigma) \\ \tau \in N_\downarrow(\sigma)}} \phi^m(h^\sigma, h^\tau, \dim \sigma, \dim \tau)$$

$$h^\sigma \leftarrow \phi^h\left(h^\sigma, m^\sigma, \dim \sigma\right)$$

- where the message and update functions: ϕ^m, ϕ^h are CGENNs

- C. Liu, D. Ruhe, F. Eijkelboom, P. Forré. **Clifford Group Equivariant Simplicial Message Passing Networks**. ICLR 2024.

Equivariant Convolution Layer for Multi-Vector Fields

- Let (V, η) be an <u>inner product space</u> of signature (p, q), $d = p + q$.
- Consider <u>linear layer</u> $L : \Gamma(V, \mathrm{Cl}(V, \eta)^{c_{\mathrm{in}}}) \to \Gamma(V, \mathrm{Cl}(V, \eta)^{c_{\mathrm{out}}})$ given by:

$$L(f)(x) := \int_V K(v)[f(x - v)]\, dv, \qquad \text{with kernel} \quad K = H \circ Y,$$

  - where Y is CGENN and **kernel head** H is linear $\mathrm{O}(V, \eta)$-equivariant top layer:

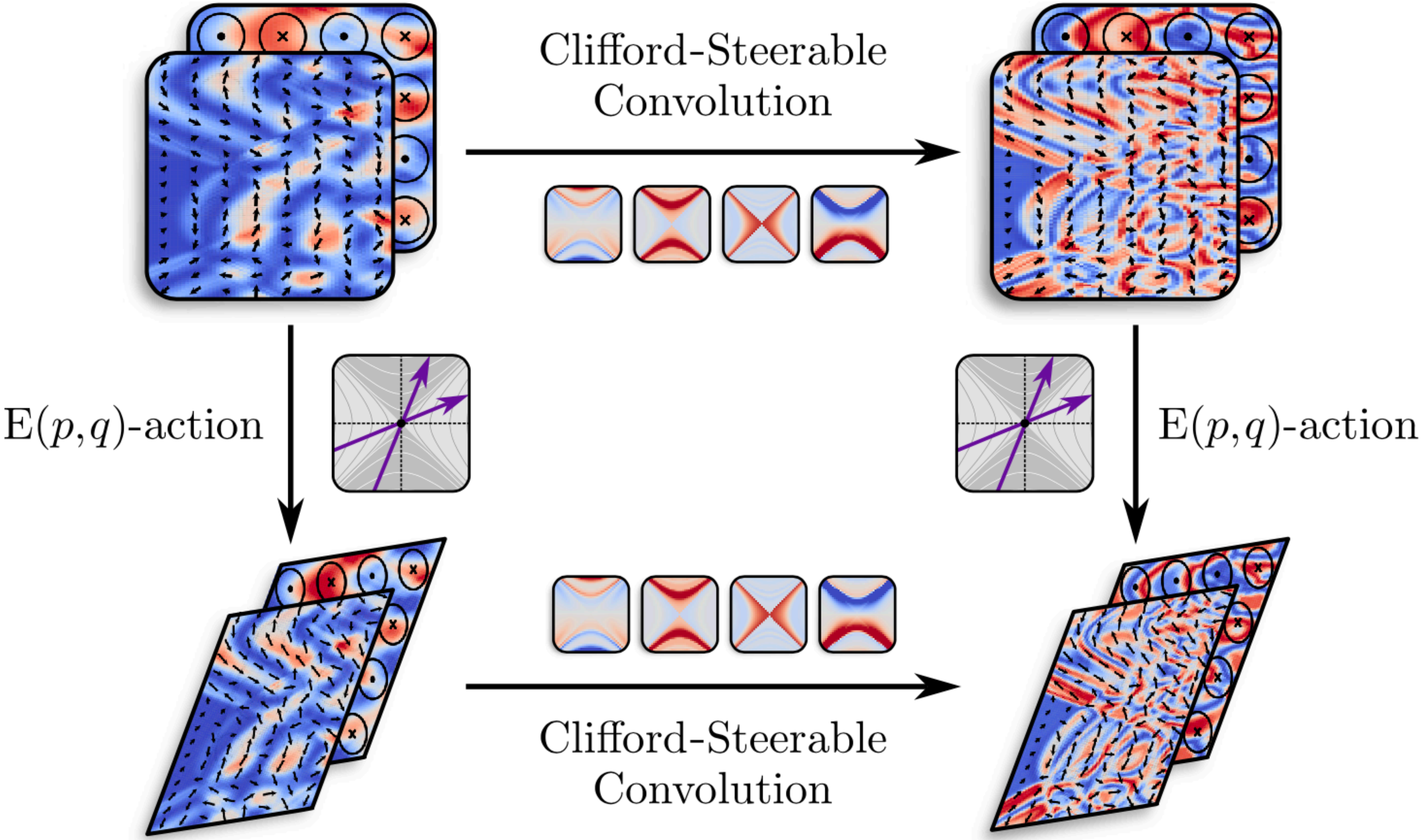
$$K : V \xrightarrow{Y} \mathrm{Cl}(V, \eta)^{c_{\mathrm{out}} \times c_{\mathrm{in}}} \xrightarrow{H} \mathrm{hom}\Big(\mathrm{Cl}(V, \eta)^{c_{\mathrm{in}}}, \mathrm{Cl}(V, \eta)^{c_{\mathrm{out}}}\Big).$$

- Then K is always $\mathrm{O}(V, \eta)$**-steerable** and L is $\mathrm{E}(V, \eta)$**-equivariant:**
  - i.e. for all $(t, g) \in \mathrm{E}(V, \eta)$ and $f \in \Gamma(V, \mathrm{Cl}(V, \eta)^{c_{\mathrm{in}}})$ we have:

$$L\Big((t, g) \rhd f\Big) = (t, g) \rhd L(f).$$

- M. Zhdanov, D. Ruhe, M. Weiler, A. Lucic, J. Brandstetter, P. Forré. **Clifford-Steerable Convolutional Neural Networks**. ICML 2024.

Clifford-Steerable Convolutional Neural Networks



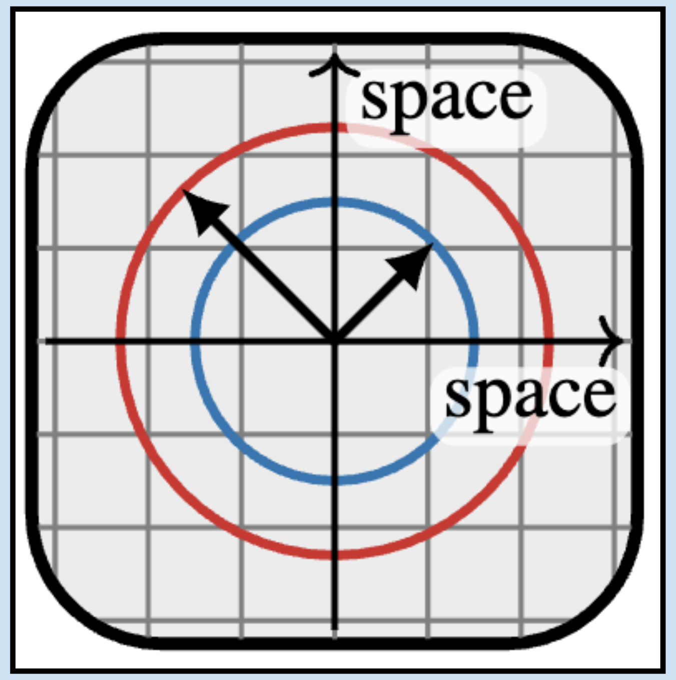
Experiments:

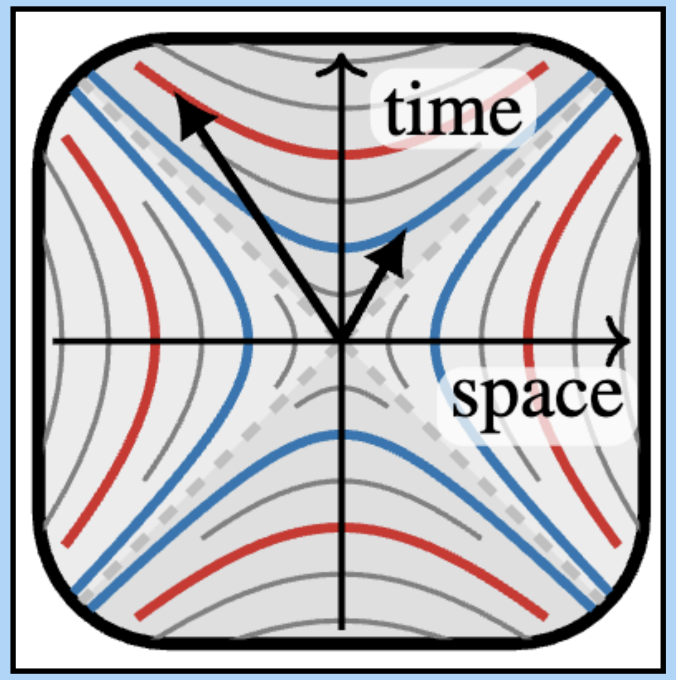
- Forecasting task: trained on past, predict future:
  - Fluid dynamics on $\mathbb{R}^2$ (incompressible Navier-Stokes)
  - Electrodynamics on $\mathbb{R}^3$ (Maxwell equations)
  - Electrodynamics on $\mathbb{R}^{1,2}$ (Maxwell equations)

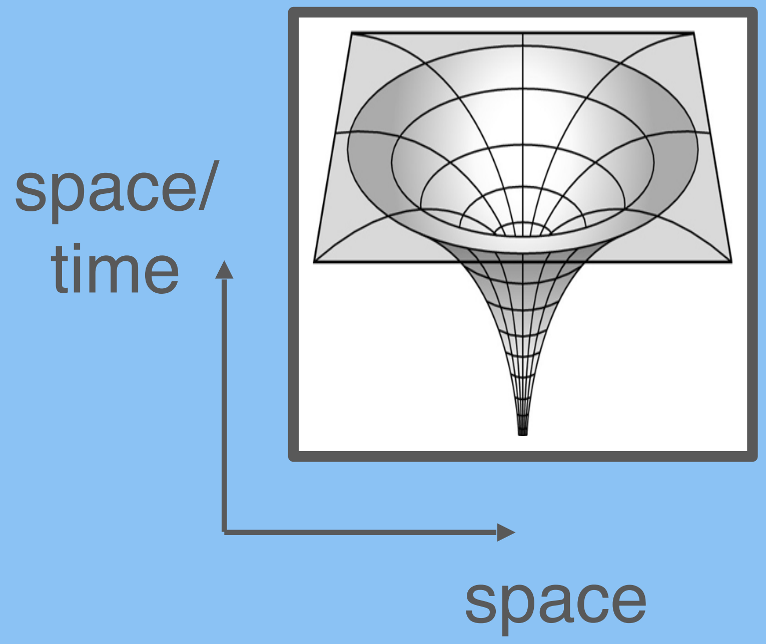
- M. Zhdanov, D. Ruhe, M. Weiler, A. Lucic, J. Brandstetter, P. Forré. **Clifford-Steerable Convolutional Neural Networks**. ICML 2024.

Steerable Convolutional Neural Networks

prior research	main paper	appendix
Euclidean space	pseudo-Euclidean space	pseudo-Riemannian manifold







classical physics	special relativity	general relativity

• M. Zhdanov, D. Ruhe, M. Weiler, A. Lucic, J. Brandstetter, P. Forré. **Clifford-Steerable Convolutional Neural Networks**. ICML 2024.

Equivariant Attention Layer for Transformers

$$\text{Attention}(Q, K, V)_{i,c'} := \sum_{t=1}^{T} \text{Softmax}_t \left(\sum_{c}^{C} \frac{\bar{\eta}(Q_{i,c}, K_{t,c})}{\sqrt{16C}} \right) \cdot V_{t,c'}$$

- query multi-vectors: Q
- key multi-vectors: K
- value multi-vectors: V
- number of multi-vector channels: C

- J. Brehmer, P. de Haan, S. Behrends, T. Cohen. **Geometric Algebra Transformers**. NeurIPS 2023.
- J. Spinner, V. Bresó, P. de Haan, T. Plehn, J. Thaler, J. Brehmer. **Lorentz-Equivariant Geometric Algebra Transformers for High-Energy Physics**. NeurIPS 2024.

Equivariant Diffusion Model

- $p(z_t^{(m)} \,|\, z^{(m)}) := \mathcal{N}\left(z_t^{(m)} \,|\, \sqrt{\bar{\alpha}_t} \cdot z^{(m)}, (1 - \bar{\alpha}_t) \cdot I\right)$
- $p(z_t \,|\, z) = \prod_{m=0}^{d} p(z_t^{(m)} \,|\, z^{(m)})$
- Loss $:= \mathbb{E}_{t \sim U, x_1 \sim q, \epsilon \sim \mathcal{N}(0,I)} \left[\| \phi_w(x_t(x_1)), t) - \epsilon \|^2 \right]$
  - with CGENN ϕ_w

- C. Liu, S. Vadgama, D. Ruhe, E. Bekkers, P. Forré. **Clifford Group Equivariant Diffusion Models for 3D Molecular Generation**. ICLR 2025 Workshop: Frontiers in Probabilistic Inference.

Summary - CGENNs

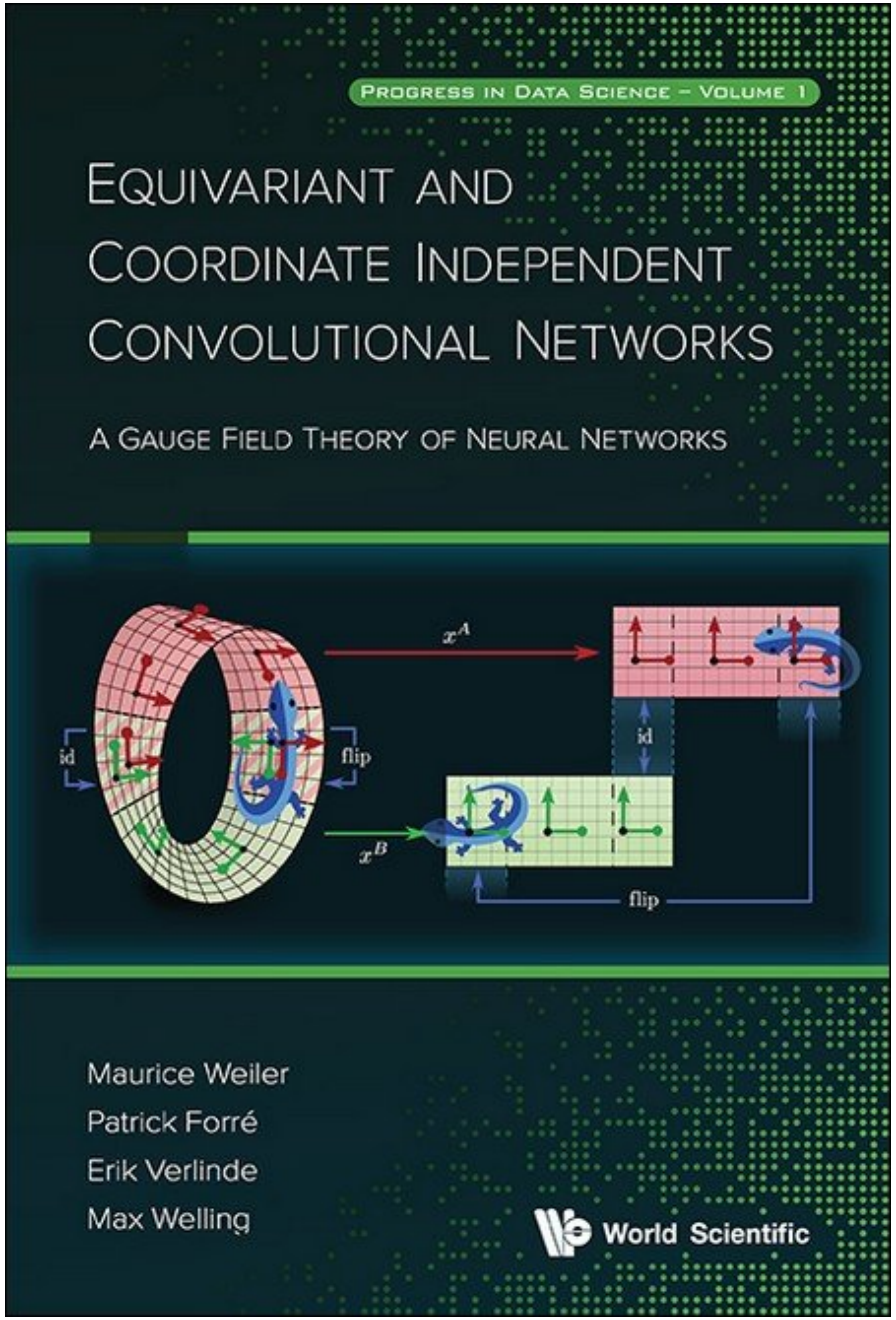
- CGENNs are based on Clifford algebra layers
- process multi-vector features
- incorporate space/spacetime symmetries, i.e. are $\mathrm{O}(V, \eta)$-equivariant
- work in any dimension d and signature (p, q, r) with same code base
- are flexible, expressive, practical
- can be combined with other architectures (graph NN, simplicial message passing NN, convolutional neural networks, diffusion models, transformers, etc.)
- perform well in comparison to other methods on $\mathrm{O}(V, \eta)$-equivariant prediction tasks
- can be improved by more efficient implementations

References

- M. Weiler, P. Forré, E. Verlinde, M. Welling. **Equivariant and Coordinate Independent Convolutional Networks - A Gauge Field Theory of Neural Networks**. Text book, World Scientific, 2025, https://doi.org/10.1142/14143.
- M. Weiler, P. Forré, E. Verlinde, M. Welling. **Coordinate Independent Convolutional Networks - Isometry and Gauge Equivariant Convolutions on Riemannian Manifolds**. arXiv:2106.06020.
- D. Ruhe, J. Brandstetter, P. Forré. **Clifford Group Equivariant Neural Networks**. NeurIPS 2023 (oral).
- C. Liu, D. Ruhe, F. Eijkelboom, P. Forré. **Clifford Group Equivariant Simplicial Message Passing Networks**. ICLR 2024.
- M. Zhdanov, D. Ruhe, M. Weiler, A. Lucic, J. Brandstetter, P. Forré. **Clifford-Steerable Convolutional Neural Networks**. ICML 2024.
- C, Liu, D. Ruhe, P. Forré. Multivector Neurons: **Better and Faster O(n)-Equivariant Clifford Graph Neural Networks**. ICML 2024 Workshop: Geometry-grounded Representation Learning and Generative Modeling.
- C. Liu, S. Vadgama, D. Ruhe, E. Bekkers, P. Forré. **Clifford Group Equivariant Diffusion Models for 3D Molecular Generation**. ICLR 2025 Workshop: Frontiers in Probabilistic Inference.
- M. Zhdanov, N. Iqbal, E. Bekkers, P. Forré. **AdS-GNN - a Conformally Equivariant Graph Neural Network**. arXiv:2505.12880.
- M. Zhdanov, N. Hoffmann, G. Cesa. **Implicit Neural Convolutional Kernels for Steerable CNNs**. NeurIPS 2022.
- S. Buchholz, G. Sommer. **On Clifford Neurons and Clifford Multi-Layer Perceptrons**. Neural Networks, 2008, 21(7), 925-935.
- J.R. Vallejo, E. Bayro-Corrochano. **Clifford Hopfield Neural Networks**. IEEE 2008.
- J. Brandstetter, R. van den Berg, M. Welling, J.K. Gupta. **Clifford Neural Layers for PDE Modeling**. ICLR 2023.
- D. Ruhe, J.K. Gupta, S. de Keninck, M. Welling, J. Brandstetter. **Geometric Clifford Algebra Networks**. ICML 2023.
- J. Brehmer, P. de Haan, S. Behrends, T. Cohen. **Geometric Algebra Transformers**. NeurIPS 2023.
- J. Spinner, V. Bresó, P. de Haan, T. Plehn, J. Thaler, J. Brehmer. **Lorentz-Equivariant Geometric Algebra Transformers for High-Energy Physics**. NeurIPS 2024.
- J. Brehmer, V. Bresó, P. de Haan, T. Plehn, H. Qu, J. Spinner, J. Thaler. **A Lorentz-Equivariant Transformer for All of the LHC**, SciPost 2025.

Main References

- M. Weiler, P. Forré, E. Verlinde, M. Welling. **Equivariant and Coordinate Independent Convolutional Networks - A Gauge Field Theory of Neural Networks**. Text book, World Scientific, 2025, https://doi.org/10.1142/14143.
- M. Weiler, P. Forré, E. Verlinde, M. Welling. **Coordinate Independent Convolutional Networks - Isometry and Gauge Equivariant Convolutions on Riemannian Manifolds**. arXiv:2106.06020.
- D. Ruhe, J. Brandstetter, P. Forré. **Clifford Group Equivariant Neural Networks**. NeurIPS 2023 (oral).
- C. Liu, D. Ruhe, F. Eijkelboom, P. Forré. **Clifford Group Equivariant Simplicial Message Passing Networks**. ICLR 2024.
- M. Zhdanov, D. Ruhe, M. Weiler, A. Lucic, J. Brandstetter, P. Forré. **Clifford-Steerable Convolutional Neural Networks**. ICML 2024.
  - => last 3 papers also presented at AGACSE 2024.



Thank you for your attention!