中国人民大學
RENMIN UNIVERSITY OF CHINA

高瓴人工智能学院
Gaoling School of Artificial Intelligence

Clifford Group Equivariant Neural Networks

David Ruhe[1,2], Johannes Brandstetter[3], Patrick Forré[1,2]

[1]AI4Science Lab, AMLab
[2]University of Amsterdam
[3]Microsoft Research AI4Science

Angxiao Yue

2023-11-09

Outline

- Introduction
- Clifford Algebras
- Theoretical Results
- Methodology
- Experiment
- Conclusion

Outline

- Introduction
- Clifford Algebras
- Theoretical Results
- Methodology
- Experiment
- Conclusion

Introduction

- Equivariant neural networks can be divided into three categories:
  1. Scalarize geometric quantities. Lack of directional information
  2. Regular group representations. Intractable for Lie Group
  3. Irreducible representations. Operate in a steerable spherical harmonics basis. The Clebsch-Gordan coefficients are not trivial to obtain.

Introduction

- Propose **CGENN**: an **equivariant parameterization(3)** of neural networks based on Clifford algebras. Inside the algebra, identify the **Clifford group(2)** and its action, termed the (adjusted) **twisted conjugation(1)**.

  1. Directly transform data in a vector basis.

     Irreducible representations

  2. Preserve geometrically meaningful product structure.

     Scalarize geometric quantities

  3. Readily generalizes to orthogonal groups regardless of the dimension or metric signature of the space

     Regular group representations

Outline

- Introduction
- Clifford Algebras
- Theoretical Results
- Methodology
- Experiment
- Conclusion

Clifford Algebras

Basic definitions and settings

- Let V be a finite-dimensional vector space over a field $\mathbb{F}$, equipped with a quadratic form $q: V \to \mathbb{F}$. In this paper's context, $\mathbb{F} := \mathbb{R}$.
- The Clifford Algebra $Cl(V,\ q)$ generated by V with $v^2 = q(v)$. For $x \in Cl(V,\ q)$, $x = \sum_{i \in I} c_i \cdot v_{i,1} \cdot v_{i,2} \dots v_{i,k_i}$.
- Associated bilinear form is $b(v_1, v_2) = \frac{1}{2}(q(v_1 + v_2) - q(v_1) - q(v_2))$.
- If $n := Dim(V)$, $Dim\big(Cl(V, q)\big) = 2^n$.
- Let $\{e_1,\ e_2,\ \dots,\ e_n\}$ be an orthogonal basis for V. The tuple $(e_A)_{A \subseteq [n]}$ is an orthogonal basis for $Cl(V, q)$, where $[n] = \{1, \dots. n\}$, $e_A := \prod_{i \in A}^{<} e_i$.

Clifford Algebras

高瓴人工智能学院
Gaoling School of Artificial Intelligence

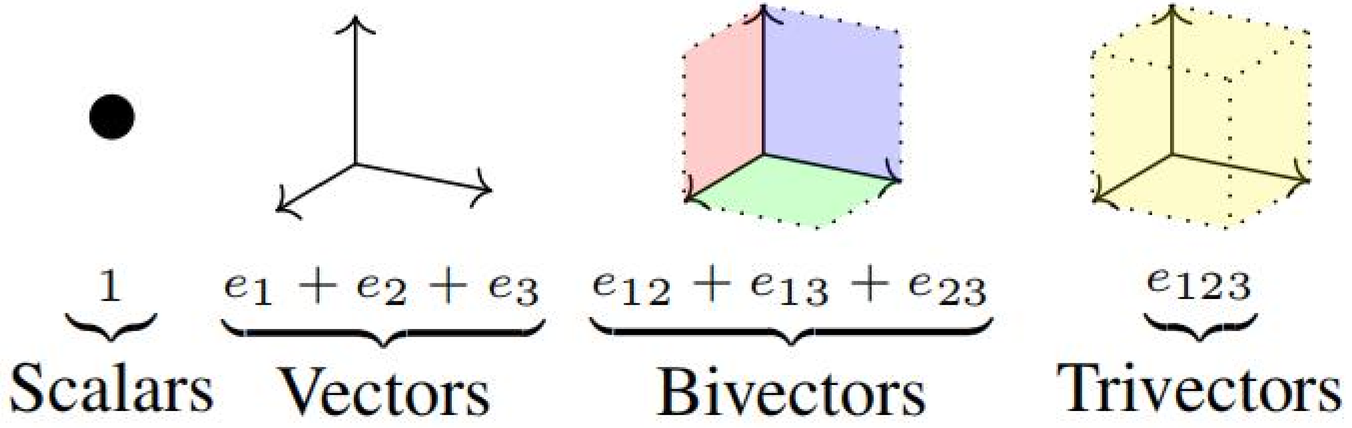
Basic definitions and settings

- We can decompose the algebra into vector subspaces $Cl(V,q)^{(m)}$, $m = 0,1,\ldots,n$, called grades. $Dim\left(Cl(V,q)^{(m)}\right) = C_n^m$.
- $m = 0$: scalars $\mathbb{F}$.
- $m = 1$: vectors V.
- $m = 2$ and $m = 3$ refer to bivectors and trivectors.

Clifford Algebras

■ Basic definitions and settings

- $Cl(V,q) = \oplus_{m=0}^{n} Cl(V,q)^{(m)}$
- For $x \in Cl(V,q)$, we can always write $x = x^{(0)} + x^{(1)} + \ldots + x^{(n)}$.
- $Cl^{[0]}(V,q) = \oplus_{m,\,even}^{n} Cl(V,q)^{(m)}$, $Cl^{[1]}(V,q) = \oplus_{m,\,odd}^{n} Cl(V,q)^{(m)}$, so $x = x^{[0]} + x^{[1]}$.



Outline

Theoretical Results

■ Clifford Group and its Clifford Algebra Representations

➢ Let $Cl^{\times}(V,q)$ denote the group of **invertible** elements of the Clifford algebra. For $w, w^{-1} \in Cl^{\times}(V,q)$, we have $ww^{-1} = w^{-1}w = 1$.

➢ For $w \in Cl^{\times}(V,q)$, define the (adjusted) **twisted conjugation(1)** as follows:

$$\rho(w): Cl(V,q) \to Cl(V,q), \quad \rho(w)(x) = wx^{[0]}w^{-1} + \alpha(w)x^{[1]}w^{-1}$$

where $\alpha(w)$ is called main involution of $Cl(V,q)$ and $\alpha(w) = w^{[0]} - w^{[1]}$.

Theoretical Results

高瓴人工智能学院
Gaoling School of Artificial Intelligence

Clifford Group and its Clifford Algebra Representations

- The map(action) $\rho(w)$ is essential for constructing equivariant neural networks operating on the Clifford algebra.
- When $\rho(w)$ is restricted to a carefully chosen **subgroup** of $Cl^{\times}(V,q)$, many desirable characteristics emerge.
- This subgroup will be called the **Clifford group(2)** of $Cl(V,\ q)$ and we define it as:

$$\Gamma(V,\mathfrak{q}) := \left\{ w \in \mathrm{Cl}^{\times}(V,\mathfrak{q}) \cap \left(\mathrm{Cl}^{[0]}(V,\mathfrak{q}) \cup \mathrm{Cl}^{[1]}(V,\mathfrak{q}) \right) \middle| \forall v \in V,\ \rho(w)(v) \in V \right\}$$

Pay attention: $\cup \neq \oplus$ **!!!**

$$Cl(V,q) = Cl^{[0]}(V,q) \oplus Cl^{[1]}(V,q)$$
$$Cl(V,q) \neq Cl^{[0]}(V,q) \cup Cl^{[1]}(V,q)$$

Theoretical Results

■ Clifford Group and its Clifford Algebra Representations

- Specifically, $\rho(w)$ was ensured to reduce to a reflection when restricted to V.
  $w,\ x \in Cl^{(1)}(V,q) = V, w \in Cl^{\times}(V,q) = V.$

$$\rho(w)(x) = -wxw^{-1} \overset{!}{=} x - 2\frac{\mathfrak{b}(w,x)}{\mathfrak{b}(w,w)}w.$$

- $\rho(w): Cl(V,q) \to Cl(V,q), \quad \rho(w)(x) = wx^{[0]}w^{-1} + \alpha(w)x^{[1]}w^{-1}$

1. *Additivity:* $\rho(w)(x_1 + x_2) = \rho(w)(x_1) + \rho(w)(x_2)$,
2. *Multiplicativity:* $\rho(w)(x_1 x_2) = \rho(w)(x_1)\rho(w)(x_2)$, *and:* $\rho(w)(c) = c$,
3. *Invertibility:* $\rho(w^{-1})(x) = \rho(w)^{-1}(x)$,
4. *Composition:* $\rho(w_2)(\rho(w_1)(x)) = \rho(w_2 w_1)(x)$, *and:* $\rho(c)(x) = x$ *for* $c \neq 0$,
5. *Orthogonality:* $\bar{\mathfrak{b}}(\rho(w)(x_1), \rho(w)(x_2)) = \bar{\mathfrak{b}}(x_1, x_2)$.

Theoretical Results

高瓴人工智能学院
Gaoling School of Artificial Intelligence

■ Resulted Equivariance on Clifford Group

➢ All grade projections are Clifford group equivariant. For $w \in \Gamma(V, q)$, $x \in Cl(V, q)$ and $m = 0, .., n$. We have

$$\rho(w)(x^{(m)}) = (\rho(w)(x))^{(m)} .$$

$$\begin{tikzcd}
\mathrm{Cl}(V, \mathfrak{q}) \arrow[r, "(_)^{(m)}"] \arrow[d, "\rho(w)"] & \mathrm{Cl}^{(m)}(V, \mathfrak{q}) \arrow[d, "\rho(w)"] \\
\mathrm{Cl}(V, \mathfrak{q}) \arrow[r, "(_)^{(m)}"] & \mathrm{Cl}^{(m)}(V, \mathfrak{q})
\end{tikzcd}$$

Theoretical Results

■ Resulted Equivariance on Clifford Group

- All polynomials are Clifford group equivariant. Let $F \in \mathbb{F}[T_1, T_2, ..., T_l]$ be a polynomial in l variables with coefficients in $\mathbb{F}$. $w \in \Gamma(V, q)$. Further, consider l varibles $x_1, x_2, ...x_l \in Cl(V, q)$.

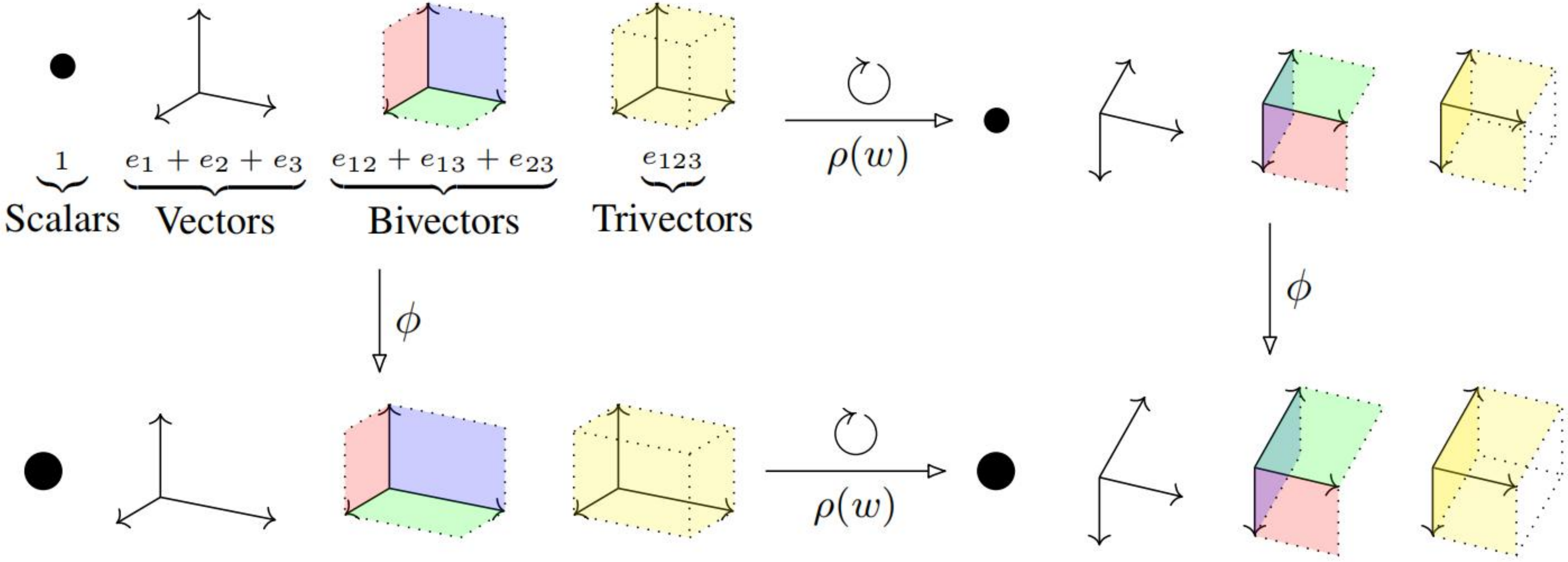
$$\rho(w)\left(F(x_1, \ldots, x_\ell)\right) = F(\rho(w)(x_1), \ldots, \rho(w)(x_\ell)).$$

$$\begin{array}{ccc}
\overbrace{\mathrm{Cl}(V,\mathfrak{q}) \times \cdots \times \mathrm{Cl}(V,\mathfrak{q})}^{\ell \text{ times}} & \xrightarrow{F} & \mathrm{Cl}(V,\mathfrak{q}) \\
\downarrow \rho(w) \quad \downarrow \rho(w) \quad \downarrow \rho(w) & & \downarrow \rho(w) \\
\mathrm{Cl}(V,\mathfrak{q}) \times \cdots \times \mathrm{Cl}(V,\mathfrak{q}) & \xrightarrow{F} & \mathrm{Cl}(V,\mathfrak{q})
\end{array}$$

Outline

Methodology

■ **Clifford Group Equivariant Neural Networks**



Methodology

■ Linear layers

➢ Let $x_1, x_2, \ldots, x_l \in Cl(V, q)$. Using the fact that a polynomial restricted to the first order constitutes a linear map, we can construct a linear layer by setting

$$y_{c_{\text{out}}}^{(k)} := T_{\phi_{c_{\text{out}}}}^{\text{lin}}(x_1, \ldots, x_\ell)^{(k)} := \sum_{c_{\text{in}}=1}^{\ell} \phi_{c_{\text{out}} c_{\text{in}} k}\, x_{c_{\text{in}}}^{(k)},$$

➢ Where $\phi_{c_{out} c_{in} k} \in \mathbb{R}$ are optimizable coefficients (**equivariant parameterization(3)**). c_{in} and c_{out} denote the input and output channel.

Methodology

■ Geometric Product Layers

➢ In this work, we only consider layers up to second order. Higher-order interactions are indirectly modeled via multiple successive layers. As an example, we take the pair x_1, x_2, their interaction terms take the form $(x_1^{(i)} x_2^{(j)})^{(k)}$, $i, j, k = 0, \ldots, n$.

$$P_\phi(x_1, x_2)^{(k)} := \sum_{i=0}^{n} \sum_{j=0}^{n} \phi_{ijk} \left(x_1^{(i)} x_2^{(j)} \right)^{(k)},$$

➢ Where $\phi_{ijk} \in \mathbb{R}$ are optimizable coefficients (**equivariant parameterization(3)**).We need $(n+1)^3$ parameters for every geometric product between a pair of multivectors.

Methodology

■ Geometric Product Layers

$$P_\phi(x_1, x_2)^{(k)} := \sum_{i=0}^{n} \sum_{j=0}^{n} \phi_{ijk} \left(x_1^{(i)} x_2^{(j)} \right)^{(k)},$$

➢ Parameterizing and computing all second-order terms amounts to l^2. Instead, we first apply a linear map to obtain y_1, ..., y_l. Through this map, the **mixing** (i.e., the terms that will get multiplied) gets learned. That is, we only get l pairs: (x_1, y_1), (x_2, y_2)… (x_l, y_l),

$$z_{c_{\text{out}}}^{(k)} := T_{\phi_{c_{\text{out}}}}^{\text{prod}}(x_1, \ldots, x_\ell, y_1, \ldots, y_\ell)^{(k)} := \sum_{c_{\text{in}}=1}^{\ell} P_{\phi_{c_{\text{out}} c_{\text{in}}}}(x_{c_{\text{in}}}, y_{c_{\text{in}}})^{(k)},$$

Methodology

Nonlinearities

$$x^{(m)} \mapsto \mathrm{ReLU}\left(x^{(m)}\right) \text{ when } m = 0$$

$$x^{(m)} \mapsto \sigma_\phi\left(\bar{\mathfrak{q}}\left(x^{(m)}\right)\right) x^{(m)} \text{ otherwise.}$$

Embedding Data in the Clifford Algebra

- Scalars: mass, charge, temperature…
- Vectors: positions, velocities…

$\mathrm{Cl}^{(0)}(V, \mathfrak{q}) \cong \mathbb{R}$ and $\mathrm{Cl}^{(1)}(V, \mathfrak{q}) \cong V$

- we can embed the data into the scalar and vector subspaces of the Clifford algebra to obtain Clifford features

$x_1, \ldots, x_\ell \in \mathrm{Cl}(V, \mathfrak{q})$

Outline

- Introduction
- Clifford Algebras
- Theoretical Results
- Methodology
- Experiment
- Conclusion

Experiment

■ Experiment setting

- O(3) Experiment: Signed Volumes
- O(5) Experiment: Convex Hulls.
- O(5) Experiment: Regression.
- E(3) Experiment: n-Body System
- O(1, 3) Experiment: Top Tagging

Experiment

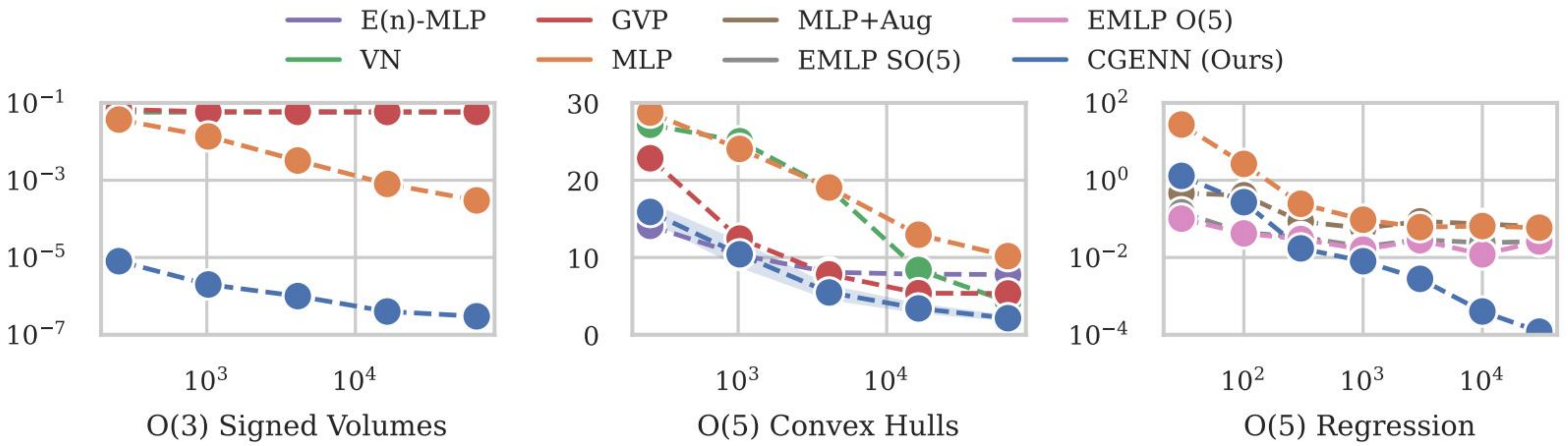


Figure 3: Left: Test mean-squared errors on the O(3) signed volume task as functions of the number of training data. Note that due to identical performance, some baselines are not clearly visible. Right: same, but for the O(5) convex hull task.

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

Experiment

高瓴人工智能学院
Gaoling School of Artificial Intelligence

Method	MSE (↓)
SE(3)-Tr.	0.0244
TFN	0.0244
NMP	0.0107
Radial Field	0.0104
EGNN	0.0070
SEGNN	0.0043
CGENN	**0.0039 ± 0.0001**

Table 1: Mean-squared error (MSE) on the n-body system 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 $\epsilon_S = 0.5$.

Outline

- Introduction
- Clifford Algebras
- Theoretical Results
- Methodology
- Experiment
- Conclusion

Conclusion

- They presented a novel approach for constructing O(n)- and E(n)-equivariant neural networks based on Clifford algebras. After establishing the required theoretical results, they proposed parameterizations of nonlinear multivector-valued maps that exhibit versatility and applicability across scenarios varying in dimension. This was achieved by the core insight that polynomials in multivectors are O(n)-equivariant functions. Theoretical results were empirically substantiated in three distinct experiments, outperforming or matching baselines that were sometimes specifically designed for these tasks.

中国人民大學
RENMIN UNIVERSITY OF CHINA

高瓴人工智能学院
Gaoling School of Artificial Intelligence

Thank You for listening!

Angxiao Yue

2023-11-09