Deep Learning Symmetries in Physics and Beyond

Roy T. Forestano   Konstantin T. Matchev   Katia Matcheva
Alexander Roman   Eyup B. Unlu   Sarunas Verner

University of Florida

arxiv:2301.05638
arxiv:2302.05383
arxiv:2302.00806
arxiv:2305.xxxxx

2023 Phenomonology Symposium
Tuesday, 9 May 2023

UF UNIVERSITY of FLORIDA

Notation and Set-Up

$$\textbf{Invariance: } \varphi(\mathbf{g} \bullet \mathbf{x}) = \varphi(\mathbf{x}) \tag{1}$$

Labelled Dataset

n **features** k **labels**

m **samples**

$$\boxed{\begin{cases} x_1^{(1)}, x_1^{(2)}, \dots, x_1^{(n)}; & y_1^{(1)}, \dots, y_1^{(k)} \\ x_2^{(1)}, x_2^{(2)}, \dots, x_2^{(n)}; & y_2^{(1)}, \dots, y_2^{(k)} \\ \vdots \quad \vdots \quad \ddots \quad \vdots & \vdots \quad \ddots \quad \vdots \\ x_m^{(1)}, x_m^{(2)}, \dots, x_m^{(n)}; & y_m^{(1)}, \dots, y_m^{(k)} \end{cases}}$$

$$|||$$

$$\{\mathbf{x_i}\} \equiv \{\mathbf{x_1}, \mathbf{x_2}, \dots, \mathbf{x_m}\} \text{ where } \mathbf{x_i} \in \mathbf{V^n}$$

$$\{\mathbf{y_i}\} = \{\mathbf{y_1}, \mathbf{y_2}, \dots, \mathbf{y_m}\} = \underbrace{\{\vec{\varphi}(\mathbf{x}_i)\}}_{\substack{\text{Oracle} \\ \text{(learned or postulated)}}}$$

Notation and Set-Up

$$\textbf{Invariance: } \varphi(\mathbf{g} \bullet \mathbf{x}) = \varphi(\mathbf{x}) \tag{1}$$

Labelled Dataset

n **features** $\quad$ k **labels**

m **samples**

$$\boxed{\begin{cases} x_1^{(1)}, x_1^{(2)}, \ldots, x_1^{(n)}; & y_1^{(1)}, \ldots, y_1^{(k)} \\ x_2^{(1)}, x_2^{(2)}, \ldots, x_2^{(n)}; & y_2^{(1)}, \ldots, y_2^{(k)} \\ \vdots \quad \vdots \quad \ddots \quad \vdots & \vdots \quad \ddots \quad \vdots \\ x_m^{(1)}, x_m^{(2)}, \ldots, x_m^{(n)}; & y_m^{(1)}, \ldots, y_m^{(k)} \end{cases}}$$

$$\equiv$$

$$\{\mathbf{x_i}\} \equiv \{\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_m}\} \text{ where } \mathbf{x_i} \in \mathbf{V^n}$$

$$\{\mathbf{y_i}\} = \{\mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_m}\} = \underbrace{\{\vec{\varphi}(\mathbf{x}_i)\}}_{\substack{\text{Oracle} \\ \text{(learned or postulated)}}}$$

Transformation

Transformation on feature space:

$$\boxed{\mathbf{g} : \mathbf{x}_i \rightarrow \mathbf{x}_i'}$$

Transformation is a symmetry if:

$$\boxed{\varphi(\mathbf{x}_i') \equiv \varphi(g(\mathbf{x}_i)) = \varphi(\mathbf{x}_i)}$$

Goal: Find transformations $\mathbf{g}(\mathbf{x}_i)$ which preserve the oracle φ.

Notation and Set-Up

$$\textbf{Invariance: } \varphi(\mathbf{g} \bullet \mathbf{x}) = \varphi(\mathbf{x}) \tag{1}$$

Labelled Dataset

$$
\begin{array}{c|cc}
 & n \textbf{ features} & k \textbf{ labels} \\
m \textbf{ samples} &
\left\{ \begin{array}{cccc}
x_1^{(1)}, x_1^{(2)}, \ldots, x_1^{(n)}; & y_1^{(1)}, \ldots, y_1^{(k)} \\
x_2^{(1)}, x_2^{(2)}, \ldots, x_2^{(n)}; & y_2^{(1)}, \ldots, y_2^{(k)} \\
\vdots \quad \vdots \quad \ddots \quad \vdots & \vdots \quad \ddots \quad \vdots \\
x_m^{(1)}, x_m^{(2)}, \ldots, x_m^{(n)}; & y_m^{(1)}, \ldots, y_m^{(k)}
\end{array} \right. &
\end{array}
$$

$$|||$$

$$\{\mathbf{x_i}\} \equiv \{\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_m}\} \text{ where } \mathbf{x_i} \in \mathbf{V^n}$$

$$\{\mathbf{y_i}\} = \{\mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_m}\} = \underbrace{\{\vec{\varphi}(\mathbf{x}_i)\}}_{\substack{\text{Oracle} \\ \text{(learned or postulated)}}}$$

Transformation

Transformation on feature space:

$$\boxed{\mathbf{g} : \mathbf{x}_i \to \mathbf{x}_i'}$$

Transformation is a symmetry if:

$$\boxed{\varphi(\mathbf{x}_i') \equiv \varphi(g(\mathbf{x}_i)) = \varphi(\mathbf{x}_i)}$$

Goal: Find transformations $\mathbf{g}(\mathbf{x}_i)$ which preserve the oracle φ.

In physics, φ represents a conserved quantity.

g	φ
Time Translation (T_0)	E
Rotation (R_{ij})	$\vec{L}$
Lorentz ($K_{\mu\nu}$)	$T^{\mu\nu}$

Parameterization of Symmetry Transformations

Linear

$$\mathbf{x}' = (\mathbb{I} + \epsilon\mathcal{W})\,\mathbf{x} \qquad (2)$$

$\mathbb{I} \equiv$ identity matrix

$\mathcal{W} \equiv$ $n \times n$ matrix to be learned by our method

Parameterization of Symmetry Transformations

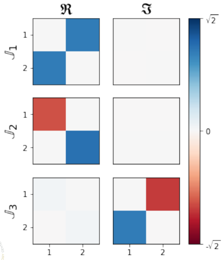
Linear

$$\mathbf{x}' = (\mathbb{I} + \epsilon\mathcal{W})\,\mathbf{x} \qquad (2)$$

$\mathbb{I} \equiv$ identity matrix

$\mathcal{W} \equiv$ $n \times n$ matrix to be learned by our method

Figure: **Visualization:** $SU(2)$ generators for a single layer linear model using the $L2$-norm oracle $\varphi(\mathbf{x}) = |\mathbf{x}|$.



Parameterization of Symmetry Transformations

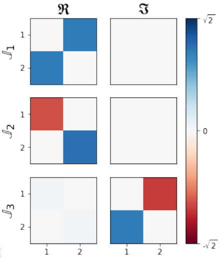
Linear

$$\mathbf{x}' = (\mathbb{I} + \epsilon\mathcal{W})\,\mathbf{x} \qquad (2)$$

$\mathbb{I} \equiv$ identity matrix

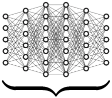
$\mathcal{W} \equiv$ $n \times n$ matrix to be learned by our method

Figure: **Visualization:** $SU(2)$ generators for a single layer linear model using the $L2$-norm oracle $\varphi(\mathbf{x}) = |\mathbf{x}|$.



Non-Linear

$\mathbf{x} \rightarrow$



$\rightarrow$ $\mathbf{x}'$ or $\frac{\mathbf{x}' - \mathbf{x}}{\epsilon}$

NN whose parameters are to be learned by our method

(3)

Parameterization of Symmetry Transformations

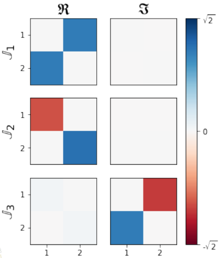
Linear

$$\mathbf{x}' = (\mathbb{I} + \epsilon \mathcal{W})\,\mathbf{x} \qquad (2)$$

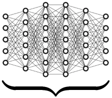
$\mathbb{I} \equiv$ identity matrix

$\mathcal{W} \equiv$ $n \times n$ matrix to be learned by our method

Figure: **Visualization:** $SU(2)$ generators for a single layer linear model using the $L2$-norm oracle $\varphi(\mathbf{x}) = |\mathbf{x}|$.

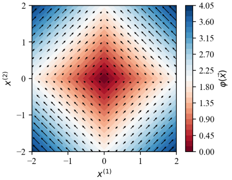


Non-Linear



$$\mathbf{x} \rightarrow \underbrace{\text{NN}}_{\substack{NN \text{ whose parameters are} \\ \text{to be learned by our method}}} \rightarrow \mathbf{x}' \text{ or } \frac{\mathbf{x}' - \mathbf{x}}{\epsilon} \qquad (3)$$

Figure: **Visualization:** Grid vector transformation representation for a deep linear layered model using the $L1$-norm oracle $\varphi(\mathbf{x}) = |\mathbf{x}^{(1)}| + |\mathbf{x}^{(2)}|$.



Loss Function

Ensure Symmetry $\implies$ Invariance $\mathcal{L}_{inv}(\mathcal{G}_{\mathcal{W}}, \{\vec{x}_i\})$

Enforces invariance among a chosen oracle $\vec{\varphi}(\vec{x})$, e.g. l^2-norm $\varphi(\vec{x}) = \sqrt{x_i^* x^i}$,

$$\mathcal{L}_{inv} = h_{inv} \frac{1}{\varepsilon^2 m} \sum_{i=1}^{m} [\vec{\varphi}(\mathcal{F}_{\mathcal{W}} \vec{x}_i) - \vec{\varphi}(\vec{x}_i)]^2 = h_{inv} \frac{1}{\varepsilon^2 m} \sum_{i=1}^{m} [\vec{\varphi}\left(\left(\mathbb{I} + \varepsilon \mathcal{W}\right) \vec{x}_i\right) - \vec{\varphi}(\vec{x}_i)]^2 \qquad (4)$$

Loss Function

Ensure Symmetry $\implies$ Invariance $\mathcal{L}_{inv}(\mathcal{G}_{\mathcal{W}}, \{\vec{x}_i\})$

Enforces invariance among a chosen oracle $\vec{\varphi}(\vec{x})$, e.g. l^2-norm $\varphi(\vec{x}) = \sqrt{x_i^* x^i}$,

$$\mathcal{L}_{inv} = h_{inv} \frac{1}{\varepsilon^2 m} \sum_{i=1}^{m} \left[\vec{\varphi}(\mathcal{F}_{\mathcal{W}} \vec{x}_i) - \vec{\varphi}(\vec{x}_i)\right]^2 = h_{inv} \frac{1}{\varepsilon^2 m} \sum_{i=1}^{m} \left[\vec{\varphi}\left(\left(\mathbb{I} + \varepsilon \mathcal{W}\right) \vec{x}_i\right) - \vec{\varphi}(\vec{x}_i)\right]^2 \quad (4)$$

Ensure non-triviality ($\mathbf{x}' \neq \mathbf{x}$) $\implies$ Normalization $\mathcal{L}_{norm}(\mathcal{G}_{\mathcal{W}}, \{\vec{x}_i\})$

Enforces the normalization condition and finding a non-trivial solution

$$\mathcal{L}_{norm} = h_{norm} \left[\mathcal{W}_{jk} \mathcal{W}^*_{kj} - 2\right]^2 \quad (5)$$

Loss Function

Ensure Symmetry $\implies$ Invariance $\mathcal{L}_{inv}(\mathcal{G}_{\mathcal{W}}, \{\vec{x}_i\})$

Enforces invariance among a chosen oracle $\vec{\varphi}(\vec{x})$, e.g. l^2-norm $\varphi(\vec{x}) = \sqrt{x_i^* x^i}$,

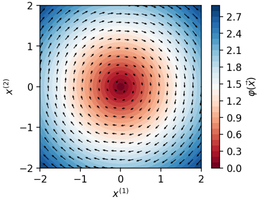
$$\mathcal{L}_{inv} = h_{inv} \frac{1}{\varepsilon^2 m} \sum_{i=1}^{m} \left[\vec{\varphi}(\mathcal{F}_{\mathcal{W}} \vec{x}_i) - \vec{\varphi}(\vec{x}_i)\right]^2 = h_{inv} \frac{1}{\varepsilon^2 m} \sum_{i=1}^{m} \left[\vec{\varphi}\left(\left(\mathbb{I} + \varepsilon \mathcal{W}\right) \vec{x}_i\right) - \vec{\varphi}(\vec{x}_i)\right]^2 \quad (4)$$

Ensure non-triviality ($\mathbf{x}' \neq \mathbf{x}$) $\implies$ Normalization $\mathcal{L}_{norm}(\mathcal{G}_{\mathcal{W}}, \{\vec{x}_i\})$

Enforces the normalization condition and finding a non-trivial solution

$$\mathcal{L}_{norm} = h_{norm} \left[\mathcal{W}_{jk} \mathcal{W}_{kj}^* - 2\right]^2 \quad (5)$$

Figure: Linear: Rotations in $2D$, $\varphi(\vec{x}) = |\vec{x}|$.



Loss Function

Ensure Symmetry $\implies$ Invariance $\mathcal{L}_{inv}(\mathcal{G}_{\mathcal{W}}, \{\vec{x}_i\})$

Enforces invariance among a chosen oracle $\vec{\varphi}(\vec{x})$, e.g. l^2-norm $\varphi(\vec{x}) = \sqrt{x_i^* x^i}$,

$$\mathcal{L}_{inv} = h_{inv} \frac{1}{\varepsilon^2 m} \sum_{i=1}^{m} \left[\vec{\varphi}(\mathcal{F}_{\mathcal{W}} \vec{x}_i) - \vec{\varphi}(\vec{x}_i)\right]^2 = h_{inv} \frac{1}{\varepsilon^2 m} \sum_{i=1}^{m} \left[\vec{\varphi}\left(\left(\mathbb{I} + \varepsilon \mathcal{W}\right) \vec{x}_i\right) - \vec{\varphi}(\vec{x}_i)\right]^2 \quad (4)$$

Ensure non-triviality ($\mathbf{x}' \neq \mathbf{x}$) $\implies$ Normalization $\mathcal{L}_{norm}(\mathcal{G}_{\mathcal{W}}, \{\vec{x}_i\})$

Enforces the normalization condition and finding a non-trivial solution

$$\mathcal{L}_{norm} = h_{norm} \left[\mathcal{W}_{jk} \mathcal{W}^*_{kj} - 2\right]^2 \quad (5)$$

Figure: Linear: Rotations in $2D$, $\varphi(\vec{x}) = |\vec{x}|$.

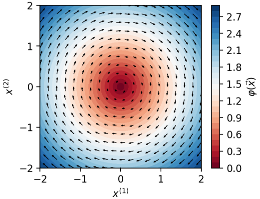
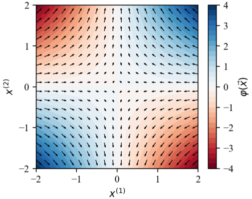


Figure: Non-linear: Squeeze mapping in $2D$, $\varphi(\vec{x}) = x^{(1)} x^{(2)}$.



Finding Multiple Symmetries

Distinct Transformations $\implies$ Orthogonality $\mathcal{L}_{orth}(\mathcal{G}_{\mathcal{W}}, \mathcal{G}'_{\mathcal{W}})$

This is built on intuition from group theory where the generators of different groups obey orthogonality conditions. Enforces the orthogonality condition and finding distinct generators $\mathbb{J}$

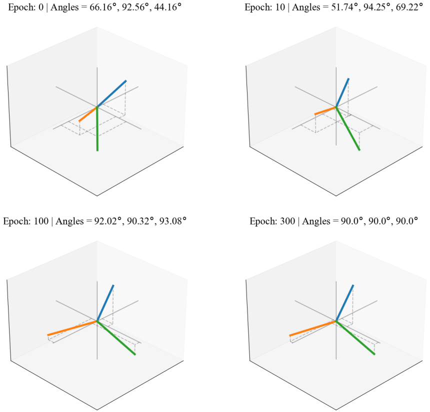
$$\mathcal{L}_{orth} = h_{orth} \left[\mathcal{W}_{jk} \mathcal{W}'^{*}_{kj} \right]^2 \tag{6}$$

Finding Multiple Symmetries

Distinct Transformations $\implies$ Orthogonality $\mathcal{L}_{orth}(\mathcal{G}_{\mathcal{W}}, \mathcal{G}'_{\mathcal{W}})$

This is built on intuition from group theory where the generators of different groups obey orthogonality conditions. Enforces the orthogonality condition and finding distinct generators $\mathbb{J}$

$$\mathcal{L}_{orth} = h_{orth}\left[\mathcal{W}_{jk}\mathcal{W}'^{*}_{kj}\right]^{2} \tag{6}$$



How many distinct symmetries exist?

- Input Parameter $\rightarrow N_g$ (number of generators). We can increase this value to search for more symmetries.

How many distinct symmetries exist?

- Input Parameter $\to N_g$ (number of generators). We can increase this value to search for more symmetries.

Example: Rotations in $2D$, $\mathbf{x} \in \mathbb{R}^2, \varphi = |\mathbf{x}|$

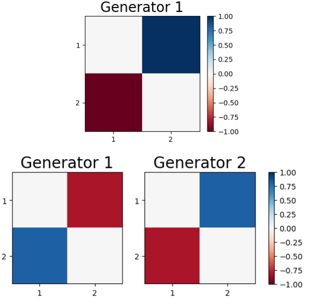


Figure: Success (top). Failure (bottom).

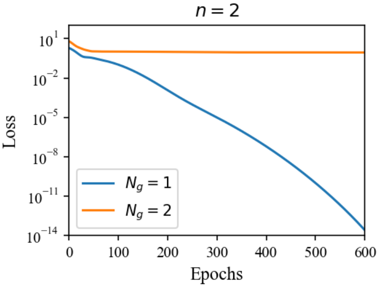


Figure: $N_g = 1, 2$ Loss

Rotations in 4 dimensions ($\mathbf{x} \in \mathbb{R}^4, \varphi = |\mathbf{x}|$)

Closure $\mathcal{L}_{clos}(a^{\gamma}_{[\alpha\beta]})$

Including a closure term $\mathcal{L}_{closure}$ ensures the generators form a closed algebra.

$$\mathcal{L}_{clos} = h_{clos} \sum_{\alpha<\beta}^{N_g} \left[\left[\mathbb{J}_\alpha, \mathbb{J}_\beta \right] - \sum_{\gamma=1}^{N_g} a^{\gamma}_{[\alpha\beta]} \mathbb{J}_\gamma \right]^2$$

Rotations in 4 dimensions ($\mathbf{x} \in \mathbb{R}^4, \varphi = |\mathbf{x}|$)

Closure $\mathcal{L}_{clos}(a^{\gamma}_{[\alpha\beta]})$

Including a closure term $\mathcal{L}_{closure}$ ensures the generators form a closed algebra.

$$\mathcal{L}_{clos} = h_{clos} \sum_{\alpha<\beta}^{N_g} \left[[\mathbb{J}_\alpha, \mathbb{J}_\beta] - \sum_{\gamma=1}^{N_g} a^{\gamma}_{[\alpha\beta]} \mathbb{J}_\gamma \right]^2$$

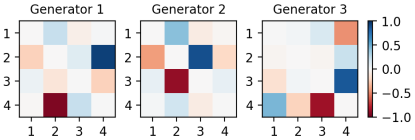


Figure: $N_g = 3$.

Rotations in 4 dimensions ($\mathbf{x} \in \mathbb{R}^4, \varphi = |\mathbf{x}|$)

Closure $\mathcal{L}_{clos}(a^{\gamma}_{[\alpha\beta]})$

Including a closure term $\mathcal{L}_{closure}$ ensures the generators form a closed algebra.

$$\mathcal{L}_{clos} = h_{clos} \sum_{\alpha<\beta}^{N_g} \left[[\mathbb{J}_\alpha, \mathbb{J}_\beta] - \sum_{\gamma=1}^{N_g} a^{\gamma}_{[\alpha\beta]} \mathbb{J}_\gamma \right]^2$$

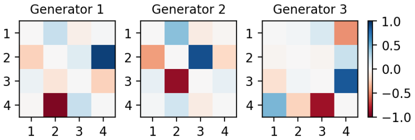


Figure: $N_g = 3$.

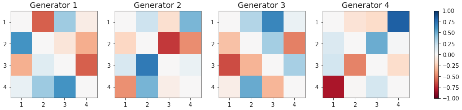
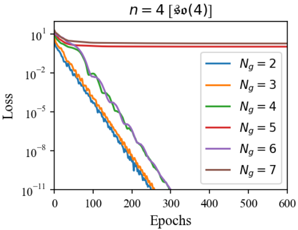


Figure: $N_g = 4$.

Closure $\mathcal{L}_{clos}(a^{\gamma}_{[\alpha\beta]})$

Including a closure term $\mathcal{L}_{closure}$ ensures the generators form a closed algebra.

$$\mathcal{L}_{clos} = h_{clos} \sum_{\alpha<\beta}^{N_g} \left[[\mathbb{J}_\alpha, \mathbb{J}_\beta] - \sum_{\gamma=1}^{N_g} a^{\gamma}_{[\alpha\beta]} \mathbb{J}_\gamma \right]^2$$



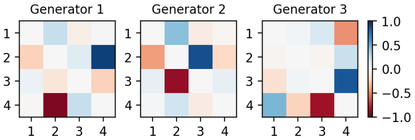


Figure: $N_g = 3$.

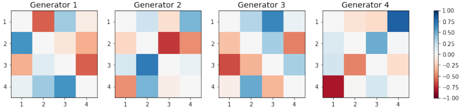


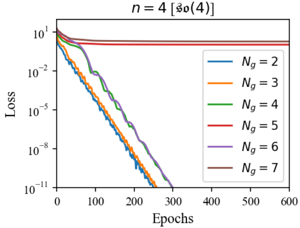
Figure: $N_g = 4$.

Rotations in 4 dimensions ($\mathbf{x} \in \mathbb{R}^4, \varphi = |\mathbf{x}|$)

Closure $\mathcal{L}_{clos}(a^{\gamma}_{[\alpha\beta]})$

Including a closure term $\mathcal{L}_{closure}$ ensures the generators form a closed algebra.

$$\mathcal{L}_{clos} = h_{clos} \sum_{\alpha<\beta}^{N_g} \left[[\mathbb{J}_\alpha, \mathbb{J}_\beta] - \sum_{\gamma=1}^{N_g} a^{\gamma}_{[\alpha\beta]} \mathbb{J}_\gamma \right]^2$$



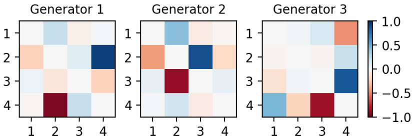


Figure: $N_g = 3$.

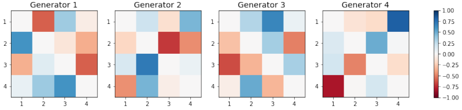


Figure: $N_g = 4$.

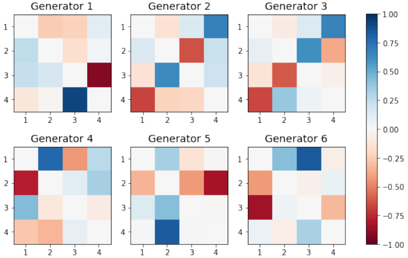


Figure: $N_g = 6$.

Other Examples: Lorentz Group $O(1,3)$ and Unitary Groups $U(n)$

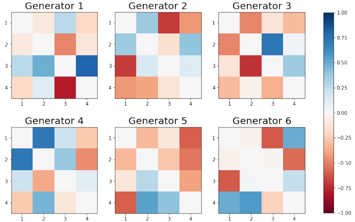


Figure: Lorentz group generators, $O(1,3)$ preserving the Lorentz vector $\varphi(\mathbf{x}) = \eta_\mu^\nu x_\mu x^\nu$.

Other Examples: Lorentz Group $O(1,3)$ and Unitary Groups $U(n)$

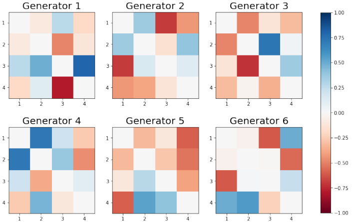


Figure: Lorentz group generators, $O(1,3)$ preserving the Lorentz vector $\varphi(\mathbf{x}) = \eta^{\nu}_{\mu} x_{\mu} x^{\nu}$.

Sparsity $\mathcal{L}_{sp}(\mathcal{W})$

Enforces the learned generators (axes of rotation) to be in the canonical basis (usual axes),

$$\mathcal{L}_{sp} = h_{sp} \sum_{j \neq l \cup k \neq m}^{n} \left[\mathcal{W}_{jk}\mathcal{W}_{lm}\right]^2 \tag{7}$$

Other Examples: Lorentz Group $O(1,3)$ and Unitary Groups $U(n)$

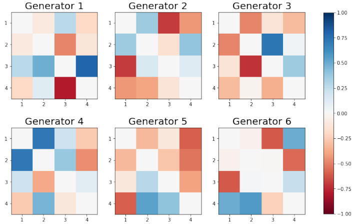


Figure: Lorentz group generators, $O(1,3)$ preserving the Lorentz vector $\varphi(\mathbf{x}) = \eta_\mu^\nu x_\mu x^\nu$.

$\longrightarrow$

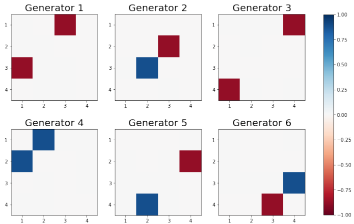


Figure: Canonical representation of $O(1,3)$ with $h_{sp} > 0$.

Sparsity $\mathcal{L}_{sp}(\mathcal{W})$

Enforces the learned generators (axes of rotation) to be in the canonical basis (usual axes),

$$\mathcal{L}_{sp} = h_{sp} \sum_{j \neq l \cup k \neq m}^{n} \left[\mathcal{W}_{jk} \mathcal{W}_{lm}\right]^2 \qquad (7)$$

Other Examples: Lorentz Group $O(1,3)$ and Unitary Groups $U(n)$

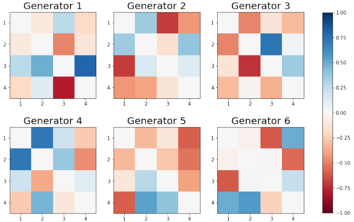


Figure: Lorentz group generators, $O(1,3)$ preserving the Lorentz vector $\varphi(\mathbf{x}) = \eta^{\nu}_{\mu} x_{\mu} x^{\nu}$.

$\longrightarrow$

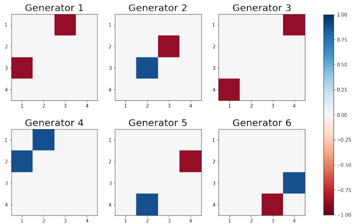


Figure: Canonical representation of $O(1,3)$ with $h_{sp} > 0$.

Sparsity $\mathcal{L}_{sp}(\mathcal{W})$

Enforces the learned generators (axes of rotation) to be in the canonical basis (usual axes),

$$\mathcal{L}_{sp} = h_{sp} \sum_{j \neq l \cup k \neq m}^{n} \left[\mathcal{W}_{jk}\mathcal{W}_{lm}\right]^2 \qquad (7)$$

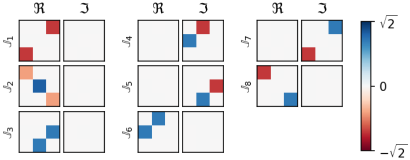
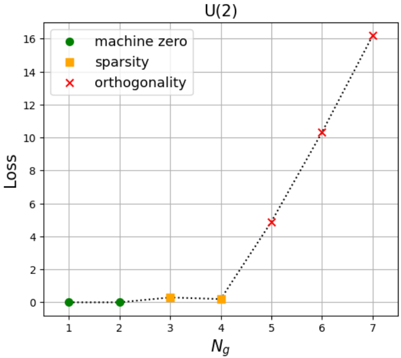


Figure: $N_g = 8, SU(3)$ Gell-Mann matrices preserving $\varphi(\mathbf{x}) = |\mathbf{x}|$.

Understanding the Full Loss Function



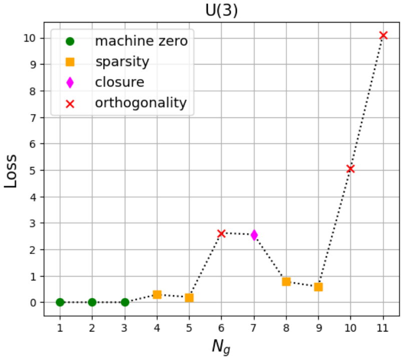


Figure: **The final value of the full loss function as a function of the number of generators N_g for $U(n)$ for $n = 2$ (left panel) and the $n = 3$ (right panel). The colored symbols identify the dominant contribution to the loss. All hyperparameters h_i were fixed to 1 except for $h_{sparsity} = 0.05$. The learning rate was 10^{-3}.**

Summary

ML Symmetries

1. Developed a method for ML symmetries in a labelled dataset.
2. General approach.
3. Finds the complete symmetry group.
4. Can be applied to realistic datasets

Learned $SO(10)$ generators.

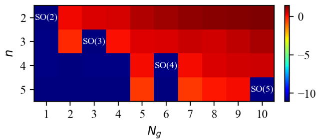


Figure: **Loss function results for $n = 2, 3, 4, 5$ dimensions and $N_g = 1, \ldots, 10$ generators. The cells are color coded by the base-10 logarithm of the lowest value of the loss attained during training.**

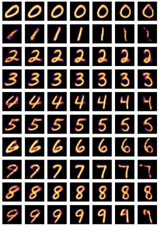
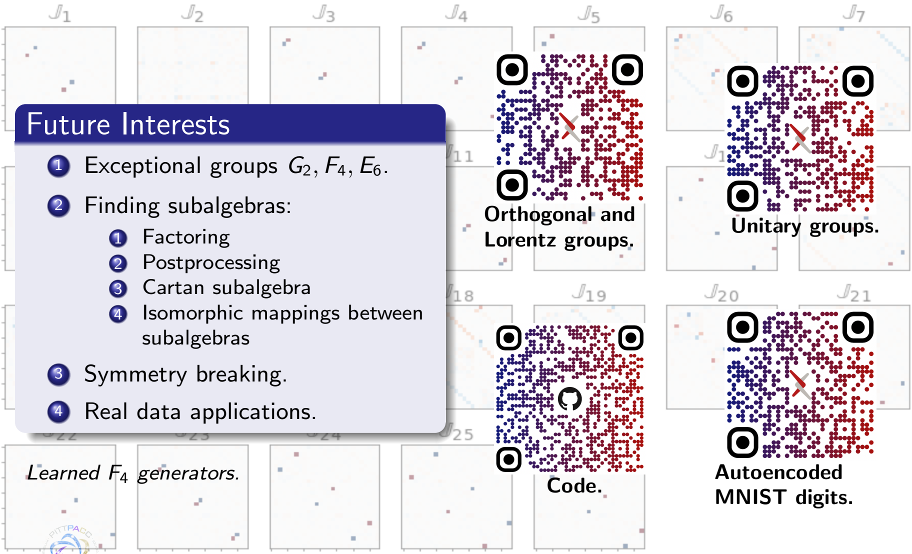


Figure: **Symmetric morphing of images along contours of the 16-dimensional latent flow. The images in the middle column represent the ideal digits in the dataset. The remaining six images in each row are obtained by moving along the contours.**

Outlook



Learned F_4 generators.

Future Interests

1. Exceptional groups G_2, F_4, E_6.
2. Finding subalgebras:
   1. Factoring
   2. Postprocessing
   3. Cartan subalgebra
   4. Isomorphic mappings between subalgebras
3. Symmetry breaking.
4. Real data applications.

Orthogonal and Lorentz groups.

Unitary groups.

Code.

Autoencoded MNIST digits.