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

Notation and Set-Up

$$\textbf{Invariance:} \quad '(\mathbf{g} \quad \mathbf{x}) = '(\mathbf{x}) \tag{1}$$

Labelled Dataset

m samples; *n* **features**; *k* **labels**

$$\left\{\begin{array}{llll|lll} x_1^{(1)}, & x_1^{(2)}, & \ldots, & x_1^{(n)}, & y_1^{(1)}, & \ldots, & y_1^{(k)} \\ x_2^{(1)}, & x_2^{(2)}, & \ldots, & x_2^{(n)}, & y_2^{(1)}, & \ldots, & y_2^{(k)} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_m^{(1)}, & x_m^{(2)}, & \ldots, & x_m^{(n)}, & y_m^{(1)}, & \ldots, & y_m^{(k)} \end{array}\right.$$

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

$\{\mathbf{y_i}\} = \{\mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_m}\} = \underbrace{\{\psi(\mathbf{x}_i)\}}$

Oracle
(learned or postulated)

Notation and Set-Up

$$\textbf{Invariance:}\quad '(\mathbf{g}\quad \mathbf{x}) = \ '(\mathbf{x}) \qquad (1)$$

Labelled Dataset

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

m **samples**

$x_1^{(1)}; x_1^{(2)}; \ldots; x_1^{(n)};$	$y_1^{(1)}; \ldots; y_1^{(k)}$
$x_2^{(1)}; x_2^{(2)}; \ldots; x_2^{(n)};$	$y_2^{(1)}; \ldots; y_2^{(k)}$
$\vdots \quad \vdots \quad \ddots \quad \vdots$	$\vdots \quad \ddots \quad \vdots$
$x_m^{(1)}; x_m^{(2)}; \ldots; x_m^{(n)};$	$y_m^{(1)}; \ldots; y_m^{(k)}$

$f\mathbf{x_i}g \quad f\mathbf{x_1}; \mathbf{x_2}; \ldots; \mathbf{x_m}g$ where $\mathbf{x_i}\ 2\ \mathbf{V^n}$

$f\mathbf{y_i}g = f\mathbf{y_1}; \mathbf{y_2}; \ldots; \mathbf{y_m}g = \quad f\,\tilde{\ }(\mathbf{x}_i)g$

Oracle
(learned or postulated)

Transformation

Transformation on feature space:

$$\boxed{\mathbf{g} : \mathbf{x}_i \ \ ! \ \ \mathbf{x}_i^0}$$

Transformation is a symmetry if:

$$\boxed{'(\mathbf{x}_i^0) \quad '(g(\mathbf{x}_i)) = \ '(\mathbf{x}_i)}$$

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

Notation and Set-Up

$$\text{Invariance: } ' \, (g \quad x) = \, ' \, (x) \qquad (1)$$

Labelled Dataset

n features   k labels

m samples

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

$f \, x_i g \quad f \quad x_1; x_2; \ldots; x_m g$ where $x_i \, 2 \, V^n$

$f \, y_i g = f \, y_1; y_2; \ldots; y_m g = \quad f \, \tilde{}(x_i) g$

Oracle
(learned or postulated)

Transformation

Transformation on feature space:

$$g : x_i \, ! \quad x_i^0$$

Transformation is a symmetry if:

$$' \, (x_i^0) \quad ' \, (g(x_i)) = \, ' \, (x_i)$$

Goal: Find transformations $g(x_i)$ which preserve the oracle $'$.

In physics, $'$ represents a conserved quantity.

g	'
Time Translation (T_0)	E
Rotation (R_{ij})	Ľ
Lorentz (K)	T

Parameterization of Symmetry Transformations

Linear

$$x^0 = (I + W)x \qquad (2)$$

I identity matrix

W n n matrix to be learned by our method

Parameterization of Symmetry Transformations

Linear

$$x^0 = (I + \quad W)\,x \qquad (2)$$

I identity matrix

W n n matrix to be learned by our method

Figure: Visualization: SU(2) generators for a single layer linear model using theL2-norm oracle ' (x) = jxj.

Parameterization of Symmetry Transformations

Linear

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

$\mathbb{I}$ identity matrix

W n n matrix to be learned by our method

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

Non-Linear

$$\mathbf{x} \; ! \quad \underbrace{\qquad}_{\substack{NN \text{ whose parameters are} \\ \text{to be learned by our method}}} \quad ! \quad \mathbf{x}' \text{ or } \frac{\mathbf{x}' \quad \mathbf{x}}{} \tag{3}$$

Parameterization of Symmetry Transformations

Linear

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

$\mathbf{I}$ identity matrix

W n n matrix to be learned by our method

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

Non-Linear

$\mathbf{x}$! $\underbrace{\qquad}$! $\mathbf{x}'$ or $\frac{\mathbf{x}'}{\mathbf{x}}$

NN whose parameters are to be learned by our method (3)

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

Loss Function

Ensure Symmetry $\implies$ Invariance $L_{inv}(G_W; \{x_i\})$

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

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

Loss Function

Ensure Symmetry $\Rightarrow$ Invariance $L_{inv}(G_W; \{x_i\})$

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

$$L_{inv} = h_{inv} \frac{1}{\varepsilon^2 m} \sum_{i=1}^{m} [\varphi(F_W x_i) - \varphi(x_i)]^2 = h_{inv} \frac{1}{\varepsilon^2 m} \sum_{i=1}^{m} [\varphi((\mathbb{I} + \varepsilon W) x_i) - \varphi(x_i)]^2 \quad (4)$$

Ensure non-triviality ($\mathbf{x}' \neq \mathbf{x}$) $\Rightarrow$ Normalization $L_{norm}(G_W; \{x_i\})$

Enforces the normalization condition and finding a non-trivial solution

$$L_{norm} = h_{norm} \left[W_{jk} W_{kj} - 2 \right]^2 \quad (5)$$

Loss Function

Ensure Symmetry $\Rightarrow$ Invariance $L_{inv}(G_W; \{x_i\})$

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

$$L_{inv} = h_{inv} \frac{1}{\varepsilon^2 m} \sum_{i=1}^{m} [\varphi(F_W x_i) - \varphi(x_i)]^2 = h_{inv} \frac{1}{\varepsilon^2 m} \sum_{i=1}^{m} [\varphi((\mathbb{I} + \varepsilon W) x_i) - \varphi(x_i)]^2 \quad (4)$$

Ensure non-triviality ($x' \neq x$) $\Rightarrow$ Normalization $L_{norm}(G_W; \{x_i\})$

Enforces the normalization condition and finding a non-trivial solution

$$L_{norm} = h_{norm} \left[W_{jk} W_{kj} - 2 \right]^2 \quad (5)$$

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

Ensure Symmetry $\implies$ Invariance $L_{inv}(G_W; \{x_i\})$

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

$$L_{inv} = h_{inv} \frac{1}{\epsilon^2 m} \sum_{i=1}^{m} [\phi(F_W x_i) - \phi(x_i)]^2 = h_{inv} \frac{1}{\epsilon^2 m} \sum_{i=1}^{m} [\phi((\mathbb{I} + \epsilon W) x_i) - \phi(x_i)]^2 \quad (4)$$

Ensure non-triviality ($x' \not\approx x$) $\implies$ Normalization $L_{norm}(G_W; \{x_i\})$

Enforces the normalization condition and finding a non-trivial solution

$$L_{norm} = h_{norm} \left[W_{jk} W_{kj} - 2 \right]^2 \quad (5)$$

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

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

Finding Multiple Symmetries

Distinct Transformations =) Orthogonality $L_{orth}(G_W; G'_W)$

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 J

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

Finding Multiple Symmetries

Distinct Transformations $\Rightarrow$ Orthogonality $L_{orth}(G_W; G'_W)$

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 J

$$L_{orth} = h_{orth} \left[W_{jk} W'_{kj} \right]^2 \qquad (6)$$

How many distinct symmetries exist?

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

How many distinct symmetries exist?

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

Example: Rotations in $2D$, $\mathbf{x}$ 2 R^2; ' $= j\mathbf{x}j$

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

Figure: $N_g = 1; 2$ Loss

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

[redacted]

Including a closure term $L_{closure}$ ensures the generators form a closed algebra.

$$L_{clos} = h_{clos} \sum_{\alpha<\beta}^{N_g} \left[[J_\alpha, J_\beta] - \sum_{\gamma=1}^{N_g} a_{[\alpha\beta]\gamma} J_\gamma \right]^2$$

[redacted]

Including a closure term $L_{closure}$ ensures the generators form a closed algebra.

$$L_{clos} = h_{clos} \sum_{\alpha<\beta}^{N_g} \left[[J_\alpha, J_\beta] - \sum_{\gamma=1}^{N_g} a_{[\alpha\beta]\gamma} J_\gamma \right]^2$$

Figure: $N_g = 3$.

Rotations in 4 dimensions $(x \in R^4; \varphi = |x|)$

[redacted]

Including a closure term $L_{closure}$ ensures the generators form a closed algebra.

$$L_{clos} = h_{clos} \sum_{<}^{N_g} \left[[J , J] - \sum_{=1}^{N_g} a_{[\ ]} J \right]^2$$

Figure: $N_g = 3$.

Figure: $N_g = 4$.

Rotations in 4 dimensions ($\mathbf{x}$ 2 $\mathbb{R}^4$; ' = $|\mathbf{x}|$)

Closure $L_{clos}(a_{[\;\;]})$

Including a closure term $L_{closure}$ ensures the generators form a closed algebra.

$$L_{clos} = h_{clos} \sum_{<}^{N_g} \left[\mathrm{J} \;\; ; \mathrm{J} \quad \sum_{=1}^{N_g} a_{[\;\;]} \mathrm{J} \right]^2$$

Figure: $N_g = 3$.

Figure: $N_g = 4$.

Rotations in 4 dimensions (x 2 R^4; ' = jxj)

Closure $L_{clos}(a_{[\ \ ]})$

Including a closure term $L_{closure}$ ensures the generators form a closed algebra.

$$L_{clos} = h_{clos} \sum_{<}^{N_g} \left[\left[J\ ; J\ \right] - \sum_{=1}^{N_g} a_{[\ \ ]} J \right]^2$$

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 $'(\mathbf{x}) = \quad x \;\; x \;\; .$

Figure: Lorentz group generators, $O(1;3)$ preserving the Lorentz vector $'(\mathbf{x}) = \quad x \; x$.

Sparsity $L_{sp}(W)$

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

$$L_{sp} = h_{sp} \sum_{j \neq l \,\cup\, k \neq m} {W_{jk} W_{lm}}^{2} \qquad (7)$$

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

!

Figure: Lorentz group generators, $O(1;3)$ preserving the Lorentz vector $'(\mathbf{x}) = \quad x \; x$.

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

Sparsity $L_{sp}(W)$

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

$$L_{sp} = h_{sp} \sum_{j \neq l \cup k \neq m} \left(W_{jk} W_{lm} \right)^2 \qquad (7)$$

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

$!$

Figure: Lorentz group generators, $O(1;3)$ preserving the Lorentz vector $'(\mathbf{x}) = \quad x \; x \;\; .$

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

Sparsity $L_{sp}(W)$

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

$$L_{sp} = h_{sp} \sum_{j \neq l \,[\, k \neq m} \quad W_{jk} W_{lm} \quad ^2 \qquad (7)$$

Figure: $N_g = 8; SU(3)$ Gell-Mann matrices preserving $'(\mathbf{x}) = jxj$.

Understanding the Full Loss Function

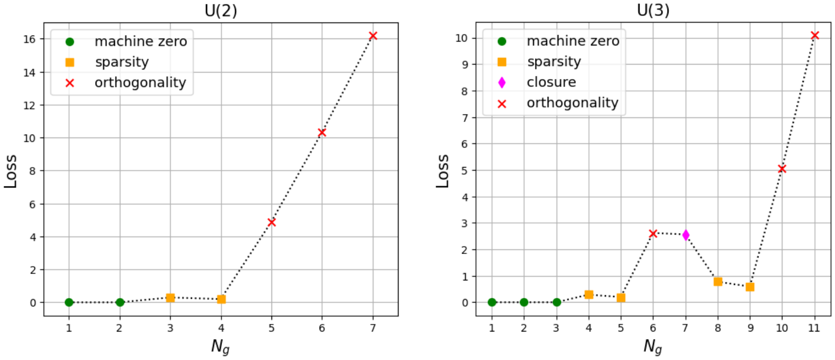

Figure: **The final value of the full loss function as a function of the number of generators N_g for $U(n)$ for $n = 2$ (left panel) and the $n = 3$ (right panel). The colored symbols identify the dominant contribution to the loss. All hyperparameters h_i were fixed to 1 except for $h_{sparsity} = 0{:}05$. 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.

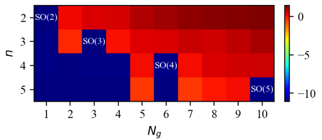

Figure: **Loss function results for $n = 2; 3; 4; 5$ dimensions and $N_g = 1; :::; 10$ generators. The cells are color coded by the base-10 logarithm of the lowest value of the loss attained during training.**

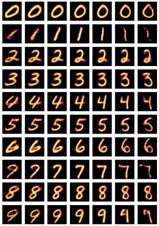

Figure: **Symmetric morphing of images along contours of the 16-dimensional latent flow. The images in the middle column represent the ideal digits in the dataset. The remaining six images in each row are obtained by moving along the contours.**

UF UNIVERSITY of FLORIDA