Invariance and equivariance in deep network learning: mathematical representation, probabilistic symmetry, Variable Exchangeability, and Sufficient Statistics

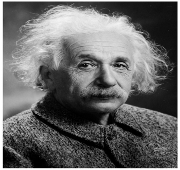
Yueyang Shen

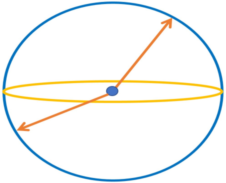
October 20, 2023

Overview

Motivating Example: Symmetry is more fundamental than observational pattern.

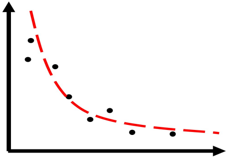
Inverse square law is the only possibility in Einstein's theory of gravity, while Newton only hypothesize the inverse square law.





(Spherical) Symmetry Approach

$$\sim \frac{1}{\text{Surface Area}} \sim \frac{1}{r^2}$$



Data Sciency Approach

$$f(r) \sim \frac{1}{r^2}$$

Symmetry and parsimony emerges from biological structures [1]

- Mathematical perspectives Symmetry is the most effective way to encode biological representation information and more likely to emerge from random mutations.
- Protein Complexes, RNA secondary structures, and model gene regulatory framework exhibit exponential bias to simpler (and symmetric structures).

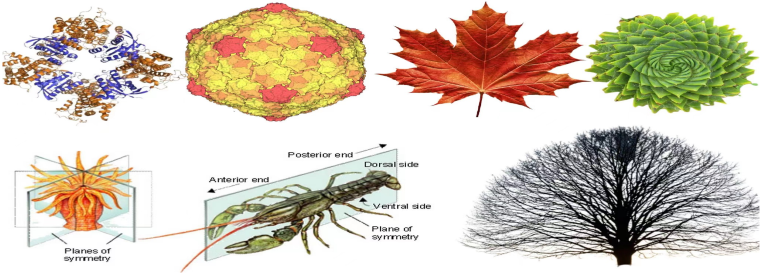


Figure: Twitter thread by chico Camargo

[1] Iain G Johnston et al. "Symmetry and simplicity spontaneously emerge from the algorithmic nature of evolution". In: *Proceedings of the National Academy of Sciences* 119.11 (2022), e2113883119

Overview

Mathematical invariance and equivalence

- Invariance and Equivariance:

$$f \text{ is } \mathfrak{G}\text{- Invariant if } \quad f(\rho(\mathfrak{g})x) = f(x) \quad \forall \mathfrak{g} \in \mathfrak{G}$$

$$f \text{ is } \mathfrak{G}\text{- Equivariant if } \quad f(\rho(\mathfrak{g})x) = \rho(\mathfrak{g})f(x) \quad \forall \mathfrak{g} \in \mathfrak{G}$$

- This exactness can be relaxed Approximate invariance.
- Examples:
  - ▶ We want the *image segmentation* to be Equivariant to translation, i.e., if we shift the object $f(\rho(\mathfrak{g})x)$ then the output of the learning segmentation should be shifted as well $(\rho(\mathfrak{g})f(x))$.
  - ▶ We want the *image classification* to be Invariant to translation, i.e., if we shift the object $f(\rho(\mathfrak{g})x)$ then the output of the class label should be unchanged $(f(x))$

Preliminary - Neural network examples

There are two ways of realizing deep network model invariance:

1. Data Augmentation (data driven)
2. Building invariance into the architecture through weight sharing (model-based).

An effective G-invariant model inference framework is composed of several equivariant functions $f_1^e,\ldots,f_n^e$ followed by a final invariant function f^i, i.e., $f^i \circ f_n^e \circ f_{n-1}^e \circ \ldots, f_1^e$.

Examples:

Vanilla NN	CNN	GNN	Self-attention layers
No symmetry	translation symmetry	Permutation symmetry	Permutation equivariant

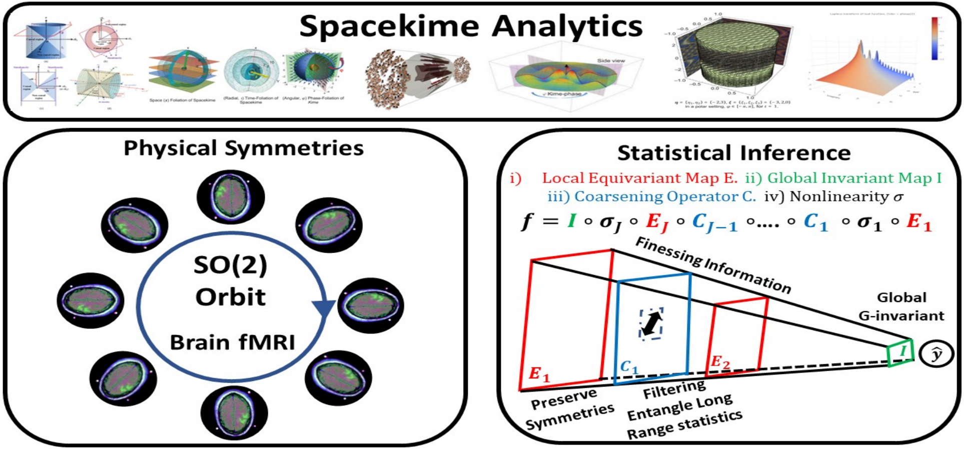
Other 2D/3D geometrical roto-tranlsational symmetries: $E(2), SE(3), SO(3)$

▶: $\mathfrak{G} - CNN$, Steerable CNNs [2], $SE(3)-$ transformer [3]

[2] Maurice Weiler and Gabriele Cesa. "General e (2)-equivariant steerable cnns". In: *Advances in Neural Information Processing Systems* 32 (2019)

[3] Fabian Fuchs et al. "Se (3)-transformers: 3d roto-translation equivariant attention networks". In: *Advances in neural information processing systems* 33 (2020), pp. 1970–1981

Overview



- Mathematical Invariance :(Working paper) Invariance properties of spacekime representations in relation to probabilistic symmetry, variable exchangeability, and sufficient statistics.

Physics example - symmetries, invariance and model equations

Physical Scenarios	Physical Space	(Lie) groups Symmetries /Invariance	Conservation Laws	Equations of motion
Rigid Bodies Rotation	$\mathbb{R}^3$ Dof: 6(3T3R)	SO(3)	Angular momentum (also casimir) $C(\Pi) = 1/2(\Pi_1^2 + \Pi_2^2 + \Pi_3^2)$	$\dot{\Pi} = \Pi \times \Omega$
Shallow water waves kdV	$(\mathbb{R}, t)$ Dof: infinite (field)	Infinite many: Translational, Scaling symmetry $\lambda^2 u(\lambda x, \lambda^3 t)$	Infinite conserved quantities $\int_a^b P_{(2n-1)}(\phi, \phi_x, \ldots)dx$ $P_1 = \phi$ $P_n = -\frac{(dP_{n-1})}{dx} + \sum_{i=1}^{n-2} P_i P_{(n-1-i)}$	$u_t + 6uu_x + u_{xxx} = 0$
Gravity	Spacetime $(M, g_{\alpha\beta})$	Diffeomorphism invariance	Local conservation of energy and momentum (zero divergence) $\nabla_\mu T^{\mu\nu} = 0$	$G_{\alpha\beta} + \Lambda g_{\alpha\beta} - 1/2$ $Rg_{\alpha\beta} = T_{\alpha\beta}$
Incompressible fluid flow	Ω	Diffeomorphism invariance	Volumetric divergence is zero	$\frac{D\Gamma}{Dt} = 0$

Math invariance, probabilistic symmetry, PDE modeling, data science

- **Physics**: (Lie) group symmetries are important in characterizing the model equations of classical physical systems.
- **Deep networks** : Deep neural network invariance is realized by weight sharing schemes (model-based) or emergence from data augmentation (data-driven).
- **Probability and Statistics**: When we introduce stochasticity into the system, classical (deterministic) symmetry is related to probabilistic symmetry where exchangeability and stationarity are primary examples.
- **Representation and information compression**: Sufficiency describes the information that is relevant to the inference and probabilistic symmetry and invariance identify information that is irrelevant to the statistical inference and ideally needs to be compressed.

Overview

Group Invariant Kreuzer Skarke - Problem formulation

Mathematical invariance

- The dataset of reflexive polyhedra in dimension four with 26 vertices, consisting of around $78,000$ polytopes. Learn to predict its Hodge numbers.
- Learn a mapping $f : \mathbb{R}^{4\times 26} \to \mathbb{Z}$, $M \mapsto h^{1,1}(M)$
- Symmetry group: $S_4 \times S_{26}$

Full data augmentation:

$$26! \times 4! = 9.68 \times 10^{27} \tag{1}$$

The full data augmentation is impossible for computation. Previous benchmark is 46.89%[4].

[4] Per Berglund, Ben Campbell, and Vishnu Jejjala. "Machine Learning Kreuzer–Skarke Calabi–Yau Threefolds". In: *arXiv preprint arXiv:2112.09117* (2021)

Group Invariant Kreuzer Skarke

Model	Acc (orig)	Acc (rnd perm)
Random Guess	2.85%	2.85%
Majority Class Only	11.28%	11.28%
MLP with reduced input [2]	46.89%	46.89%
MLP(CNN)	$56.71 \pm 0.38\%$	$30.78 \pm 0.49\%$
MLP(CNN)+π	$71.98 \pm 0.61\%$	$61.28 \pm 0.35\%$
XGBoost [4]	55.02%	29.70%
XGBoost+π	70.63%	59.84%
Vision Transformer	$62.06 \pm 0.44\%$	$43.70 \pm 0.60\%$
Vision Transformer+π	$69.00 \pm 0.48\%$	$61.02 \pm 0.63\%$

- Pointnet and partially invariant with encoder-decoder architecture brings the classification result to high 90% s.

Group invariant Kreuzer Skarke Conclusion

- Building group invariance, or even partial invariance, into string theory Kreuzer Skarke dataset improve model AI performance.
- Non group invariant models benefit from the introduction of a group invariant preprocessing step.

Overview

Constructing In/Equi variant feature representations

- Invariance and Equivariance:

$$f \text{ is } \mathfrak{G}\text{- Invariant if} \quad f(\rho(\mathfrak{g})x) = f(x) \quad \forall \mathfrak{g} \in \mathfrak{G}$$

$$f \text{ is } \mathfrak{G}\text{- Equivariant if} \quad f(\rho(\mathfrak{g})x) = \rho(\mathfrak{g})f(x) \quad \forall \mathfrak{g} \in \mathfrak{G}$$

Theorem

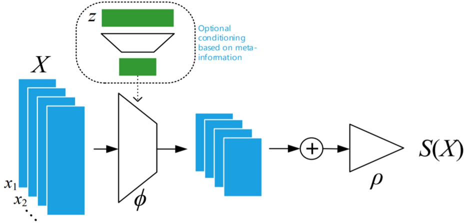
Group convolutions are all you need for equivariance [T. Cohen 1811.02017][E. Bekkers, 1909.12057]

The linear operator $\mathcal{K} \in \mathcal{H}$ is equivariant to group iff:

- *It is a group convolution* $[\mathcal{K}f](y) = \int_X \frac{1}{|g_y|} k(g_y^{-1}x)f(x)dx$
- *The kernel is subject to symmetry constraint for H:* $\forall_{h \in H} \frac{1}{|g_y|} k(hx) = k(x)$ *where* $\mathcal{K} : \underbrace{\mathbb{L}_2(X)}_{\text{2D image}} \rightarrow \underbrace{\mathbb{L}_2(Y)}_{\text{image index on group elements}}$ *and Y is a homogeneous space and quotient space identified by the symmetry H, $Y = \mathfrak{G}/H$ such that for some chosen origin $y_0 \in Y$ $g_y \in \mathfrak{G}$ we have* $\forall_{y \in Y} : y = g_y y_0$

Constructing In/Equi variant feature representations

- Invariance and Equivariance:
- Permutation Invariance and Equivariance[5]:

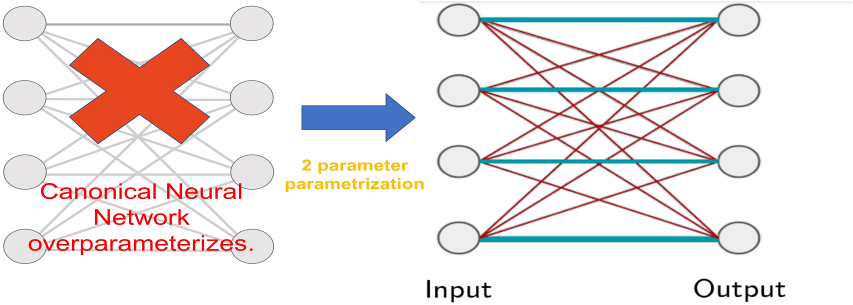


Permuation Invariance (Countable Case):

$$f(x) = \rho\left(\sum_{x \in X} \phi(x)\right)$$

[5] Manzil Zaheer et al. "Deep sets". In: *Advances in neural information processing systems* 30 (2017).

- Invariance and Equivariance:
- Permutation Invariance and Equivariance [6]:



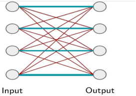
Permuation Equivariance:

$$f_{\Theta}(\mathbf{x}) = \sigma(\Theta \mathbf{x}), \Theta = \lambda \mathbb{I} + \gamma(\mathbf{1}\mathbf{1}^T), \mathbf{1} = [1, \ldots, 1]^T$$

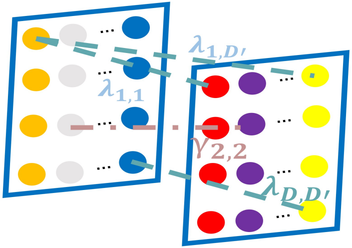
[6] Manzil Zaheer et al. "Deep sets". In: *Advances in neural information processing systems* 30 (2017).

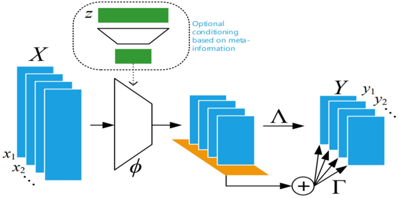
Constructing In/Equi variant feature representations

- Invariance and Equivariance:
- Permutation Invariance and Equivariance [7]:



Permuation Equivariance: $f_\Theta(\mathbf{x}) = \sigma(\Theta\mathbf{x}), \Theta = \lambda\mathbb{I} + \gamma(\mathbf{1}\mathbf{1}^T), \mathbf{1} = [1, \ldots, 1]^T$





$$\Lambda = \begin{pmatrix} \lambda_{1,1} & \cdots & \lambda_{1,D'} \\ \vdots & \ddots & \vdots \\ \lambda_{D,1} & \cdots & \lambda_{1,D'} \end{pmatrix} \Gamma = \begin{pmatrix} \gamma_{1,1} & \cdots & \gamma_{1,D'} \\ \vdots & \ddots & \vdots \\ \gamma_{D,1} & \cdots & \gamma_{1,D'} \end{pmatrix}$$

$$f(x) = \sigma(x\Lambda + \mathbf{1}\mathbf{1}^T x\Gamma)$$

[7] Manzil Zaheer et al. "Deep sets". In: *Advances in neural information processing systems* 30 (2017).

Geometric Deep Learning Blueprint[8]

Learning Stable Representations of high-dimensional data.

- Result 1

pointwise nonlineaity activation

$$\text{If } B \text{ is } \mathfrak{G}\text{-equivariant, then } \boldsymbol{U} = (\boldsymbol{\sigma} \circ B) \text{ is also } \mathfrak{G}\text{-equivariant} \tag{2}$$

▶Corollary: A general $\mathfrak{G}$-invariant family construction can be realized by composing the group average operation A with U, i.e., $A \circ U$ (Theoretical justification for CNN and $\mathfrak{G}$-CNN)

[8] Michael M Bronstein et al. "Geometric deep learning: Grids, groups, graphs, geodesics, and gauges". In: *arXiv preprint arXiv:2104.13478* (2021)

Geometric Deep Learning Blueprint[8]

Learning Stable Representations of high-dimensional data.

- Result 1

pointwise nonlineaity activation

$$\text{If } B \text{ is } \mathfrak{G}\text{-equivariant, then } \boldsymbol{U} = (\boldsymbol{\sigma} \circ B) \text{ is also } \mathfrak{G}\text{-equivariant} \quad (2)$$

▶Corollary: A general $\mathfrak{G}$-invariant family construction can be realized by composing the group average operation A with U, i.e., $A \circ U$ (Theoretical justification for CNN and $\mathfrak{G}$-CNN)

- Result 2: Universal Approximation Theorems for $\mathfrak{G}$-invariant functions

Generalizing group average A, shallow geometric networks are also Universal approximators

[8] Michael M Bronstein et al. "Geometric deep learning: Grids, groups, graphs, geodesics, and gauges". In: *arXiv preprint arXiv:2104.13478* (2021)

Geometric Deep Learning Blueprint[8]

Learning Stable Representations of high-dimensional data.

- Result 1

pointwise nonlineaity activation

$$\text{If } B \text{ is } \mathfrak{G}\text{-equivariant, then } \boldsymbol{U} = (\boldsymbol{\sigma} \circ B) \text{ is also } \mathfrak{G}\text{-equivariant} \quad (2)$$

▶Corollary: A general $\mathfrak{G}$-invariant family construction can be realized by composing the group average operation A with U, i.e., $A \circ U$ (Theoretical justification for CNN and $\mathfrak{G}$-CNN)

- Result 2: Universal Approximation Theorems for $\mathfrak{G}$-invariant functions

Generalizing group average A, shallow geometric networks are also Universal approximators

- Result 3: Fundamental tension for shallow global invariance and deformation stability.
  Solution: Localized Equivariant Maps (Kernels) - Example given later.

[8] Michael M Bronstein et al. "Geometric deep learning: Grids, groups, graphs, geodesics, and gauges". In: *arXiv preprint arXiv:2104.13478* (2021)

Geometric Deep Learning Blueprint[8]

Learning Stable Representations of high-dimensional data.

- Result 1

pointwise nonlineaity activation

$$\text{If } B \text{ is } \mathfrak{G}\text{-equivariant, then } \boldsymbol{U} = (\boldsymbol{\sigma} \circ B) \text{ is also } \mathfrak{G}\text{-equivariant} \quad (2)$$

▶Corollary: A general $\mathfrak{G}$-invariant family construction can be realized by composing the group average operation A with U, i.e., $A \circ U$ (Theoretical justification for CNN and $\mathfrak{G}$-CNN)

- Result 2: Universal Approximation Theorems for $\mathfrak{G}$-invariant functions

Generalizing group average A, shallow geometric networks are also Universal approximators

- Result 3: Fundamental tension for shallow global invariance and deformation stability.
  Solution: Localized Equivariant Maps (Kernels) - Example given later.
- Building Blocks: i) Local Equivariant Map. ii) Global Invariant Map. iii) Coarsening Operators.

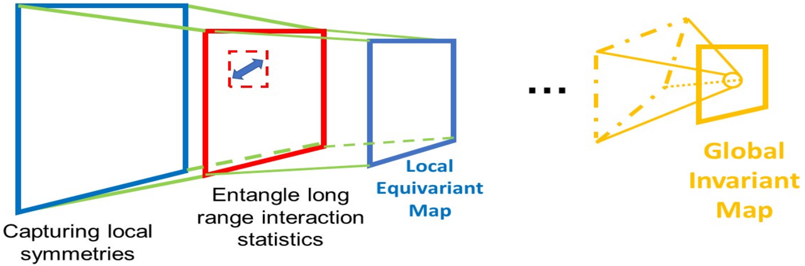
[8] Michael M Bronstein et al. "Geometric deep learning: Grids, groups, graphs, geodesics, and gauges". In: *arXiv preprint arXiv:2104.13478* (2021)

Geometric Deep Learning Blueprint[9]

Need multiple local equivariant Map to entangle long-range action

i) **Local Equivariant Map B.**
ii) **Global Invariant Map A.**
iii) **Coarsening Operators C.**

$$f = A \circ \sigma_J \circ B_J \circ C_{J-1} \circ \cdots \circ C_1 \circ \sigma_1 \circ B_1$$



[9] Michael M Bronstein et al. "Geometric deep learning: Grids, groups, graphs, geodesics, and gauges". In: *arXiv preprint arXiv:2104.13478* (2021)

Approximate equivariance

Invariance/Equivariance Motivation: For practical modeling of the heat equation

$$u_t = \kappa \underbrace{\nabla^2 u_{xx}}_{\text{rotational symmetric}} \tag{3}$$

we may have air bubbles in modeling gas, or the material is not pure in modeling solids, hence the Laplacian cannot guarantee full rotational symmetry .[10]

Definition (Approximate Equivariance/Invariance)

With $f : X \to Y$ be the inference function and the group homomorphism $\rho_X : G \to GL(X)$ and $\rho_Y : G \to GL(Y)$, then f is ϵ-approximate G-equivariant if for all $g \in G$

$$\|f(\rho_X(g)(x)) - \rho_Y(g)f(x)\| \leq \epsilon \tag{4}$$

ϵ-approximate G-invariant if for all $g \in G$

$$\|f(\rho_X(g)(x)) - f(x)\| \leq \epsilon \tag{5}$$

[10] Rui Wang, Robin Walters, and Rose Yu. "Approximately equivariant networks for imperfectly symmetric dynamics". In: *International Conference on Machine Learning*. PMLR. 2022, pp. 23078–23091

Spacekime Goals

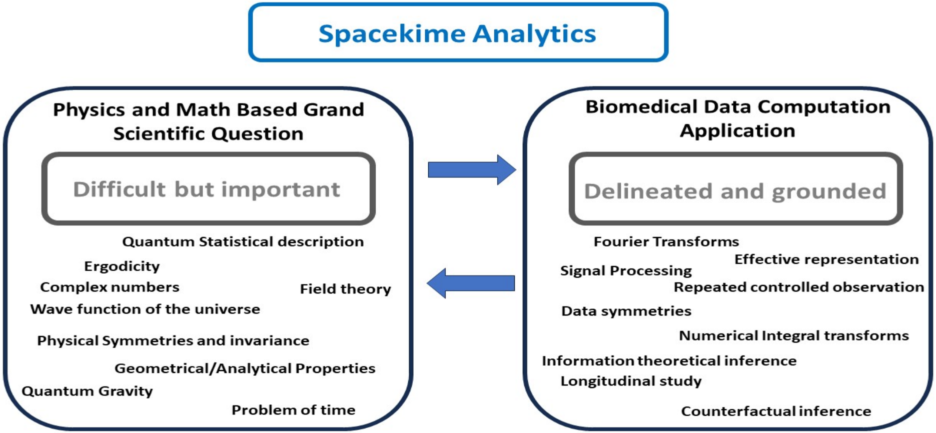


Figure: The high level goal for spacekime

Spacekime Goals

- This project involves two kinds of research question: one is very grand scientific representation models, the other is more practical, pragmatic, theoretical, but computational tractable models.
- The big goals (open-ended) complement with the grounded biomedical computation goals (concrete).
- If some of the physics problem can not be resolved we can find concrete invariances to apply to our biomedical studies.
- The implication of the biomedical theoretical analysis and data computation would have direct benefit to biomedical application.

References

[1] Per Berglund, Ben Campbell, and Vishnu Jejjala. "Machine Learning Kreuzer–Skarke Calabi–Yau Threefolds". In: *arXiv preprint arXiv:2112.09117* (2021).

[2] Per Berglund, Ben Campbell, and Vishnu Jejjala. *Machine Learning Kreuzer–Skarke Calabi–Yau Threefolds*. 2021. DOI: 10.48550/ARXIV.2112.09117. URL: https://arxiv.org/abs/2112.09117.

[3] Michael M Bronstein et al. "Geometric deep learning: Grids, groups, graphs, geodesics, and gauges". In: *arXiv preprint arXiv:2104.13478* (2021).

[4] Tianqi Chen et al. "Xgboost: extreme gradient boosting". In: *R package version 0.4-2* 1.4 (2015), pp. 1–4.

[5] Fabian Fuchs et al. "Se (3)-transformers: 3d roto-translation equivariant attention networks". In: *Advances in neural information processing systems* 33 (2020), pp. 1970–1981.

[6] Iain G Johnston et al. "Symmetry and simplicity spontaneously emerge from the algorithmic nature of evolution". In: *Proceedings of the National Academy of Sciences* 119.11 (2022), e2113883119.

[7] Rui Wang, Robin Walters, and Rose Yu. "Approximately equivariant networks for imperfectly symmetric dynamics". In: *International Conference on Machine Learning*. PMLR. 2022, pp. 23078–23091.

[8] Maurice Weiler and Gabriele Cesa. "General e (2)-equivariant steerable cnns". In: *Advances in Neural Information Processing Systems* 32 (2019).

[9] Manzil Zaheer et al. "Deep sets". In: *Advances in neural information processing systems* 30 (2017).