Symmetric Architecture Design for LHC Phenomenology

Vishal Ngairangbam
IPPP, Durham University

2410.18553 D. Maître, V. N., M. Spannowsky

Durham University

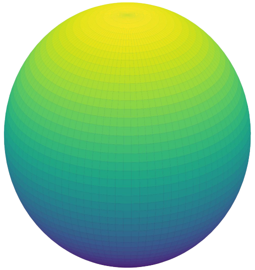
Outline

- Group symmetries in point cloud representation
- Inductive biases a quotient topologies
- Optimal symmetries from the Matrix-Element Method
  - Homogeneous point cloud architectures that assume S_N invariance are suboptimal for most scenarios
  - Largest continuous symmetry can be at most $O(1,1) \oplus O(2)$ and not $O(1,3)$

Point Cloud Representation

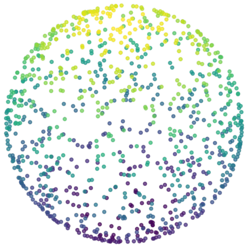
$$S^2 = \{x^2 + y^2 + z^2 = 1 \ : \ (x, y, z) \in \mathbb{R}^3\}$$

Embedded Manifold of
unit radius



$$D = \{(x_1, y_1, y_2) \ : \ i = 1, 2, \ldots, N\} \subset S^2$$

Point Cloud: ***Set of points*** sampled from the embedded manifold



Common 3D rotation on each sample produces a correct subset

$$D' = R(g)\ D \quad , \quad g \in SO(3)$$

Should one always utilise SO(3) invariance?

In an inverse scenario, $D \in \mathbb{R}^{3N}$

$$D = (x_1, y_1, z_1, x_2, y_2, z_2, \ldots, x_N, y_N, z_N)$$

Group Invariant Functions

Target function defines ***(possible)*** symmetries

$$\mathbf{r}_i \mapsto \mathbf{r}'_i = R(g)\,\mathbf{r}_i$$

$$\mathbf{r} = (x, y, z)$$

Target Function
$$f : \mathbb{R}^{3N} \to \{0, 1\}$$

$$D = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$$

$$f(D) = f(D')$$

Group Transformation
$$A : \mathbb{R}^{3N} \times \mathcal{G} \to \mathbb{R}^{3N}$$

Transform and apply function
$$f \circ A : \mathbb{R}^{3N} \to \{0, 1\}$$

$$D \mapsto D' = (\mathbf{r}'_1, \mathbf{r}'_2, \ldots, \mathbf{r}'_N)$$

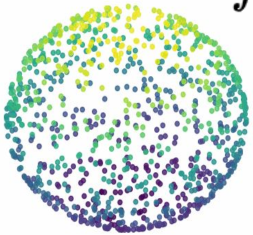
Group Invariant Functions

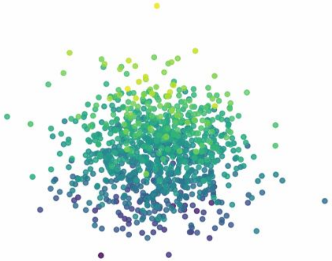
$$r_i = \sqrt{x_i^2 + y_i^2 + z_i^2}$$

Define $f : \mathbb{R}^{3N} \to \{0,1\}$

$$D = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$$

$$f(D) = 1 \implies \text{all } r_i = 1$$





$$f(D) \neq 1 \implies \text{at least one } r_i \neq 1$$

f is SO(3) invariant!

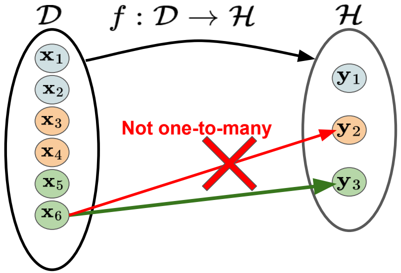
$$f(D) = f(D_\sigma) \implies f \text{ is } S_N \text{ invariant}$$

$$D \mapsto D_\sigma = (\mathbf{r}_{\sigma(1)}, \mathbf{r}_{\sigma(2)}, \ldots, \mathbf{r}_{\sigma(N)})$$

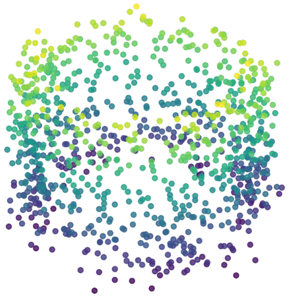
Noisy Binary Classification

Is SO(3) invariance still a good symmetry?

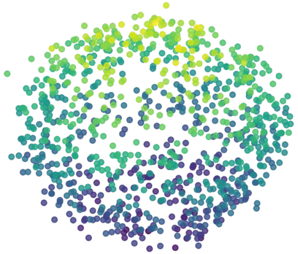
- **Overlap between probability** distributions
  - Gaussian noise model for example
- **Class assignment** becomes **one-to-many**
  - ***not well-defined as a function***



Cylinder: $y = 0$



Sphere: $y = 1$



Noisy Binary Classification

What is the target function?

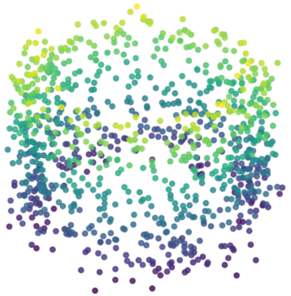
Neyman-Pearson Lemma: Optimal Discriminant is a monotonic function of the likelihood ratio

$$\lambda(D) = \frac{p_1(D)}{p_0(D)}$$

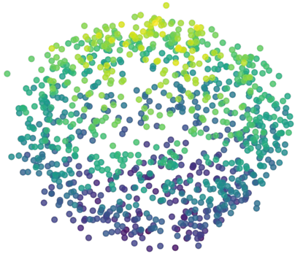
Correct Symmetry = Symmetry Shared by both classes

(upto a possible scale invariance)

Cylinder: $y = 0$

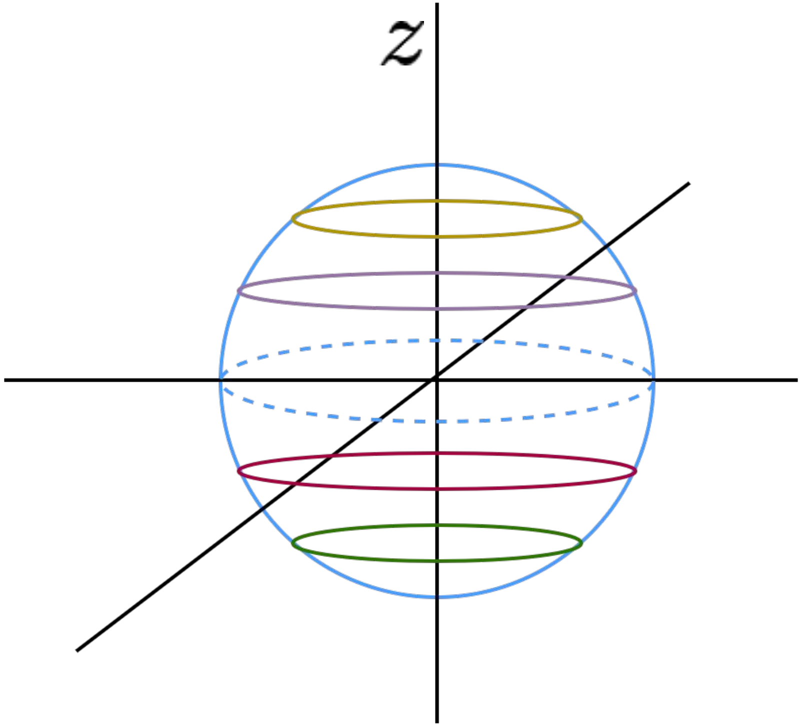


Sphere: $y = 1$



$\mathrm{SO}(2)$ rotations along z-axis is the largest symmetry for cylinder vs sphere

Inverted Hierarchy of Group Invariant Function spaces



Orbit $[\mathbf{x}]_\rho = \{\mathbf{x} \,:\, \mathbf{x}' = \rho(g)\,\mathbf{x}\ \forall g \in \mathcal{G}\}$

Orbits partition the domain $\mathcal{D}$

$\mathcal{G}$-invariance requires fibres to be at least as large as the orbits

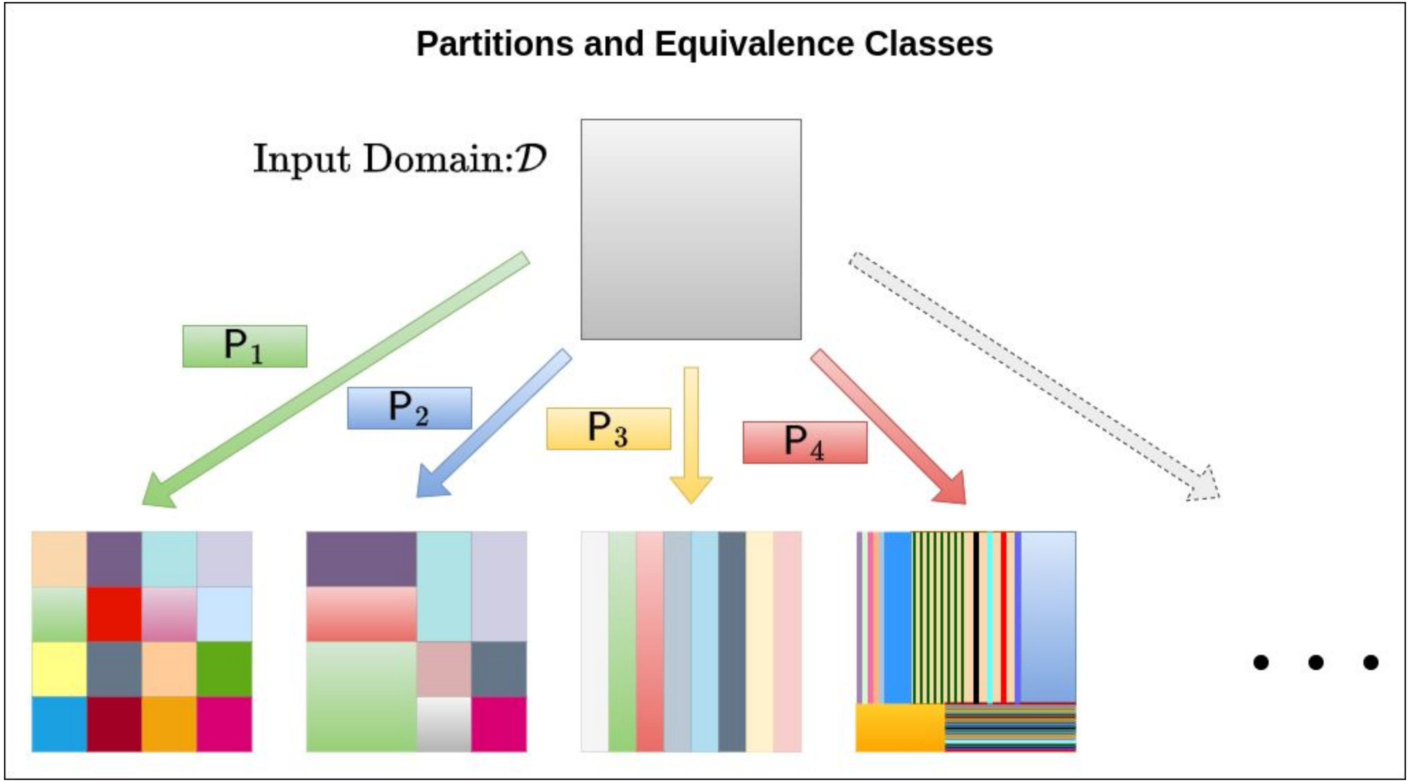
$$\mathbf{y} = f(\mathbf{x}) = f(\rho(g)\,\mathbf{x})$$

$[\mathbf{x}]_{SO(3)}$ = Sphere of radius $|\mathbf{x}|$

$[\mathbf{x}]_{SO_z(2)}$ = Circles of radius $\sqrt{x^2+y^2}$ around z-axis

$$[\mathbf{x}]_{SO(3)} = \bigcup_{0<r\leq|\mathbf{x}|} [\mathbf{x}_r]_{SO(2)}$$

Partition of Infinite Sets



Inductive Bias: Input Domain

Target function: $f : \mathcal{D} \rightarrow \mathcal{H}$ Approximator : $\hat{f} : \mathcal{D} \rightarrow \mathcal{H}$

$$[x]_f = \text{ Fibres of } f$$

1. Find the partition $[\mathbf{x}]_f$

2. Match $\hat{f}([\mathbf{x}]_f) = f([\mathbf{x}]_f) \quad \forall \quad [\mathbf{x}]_f$

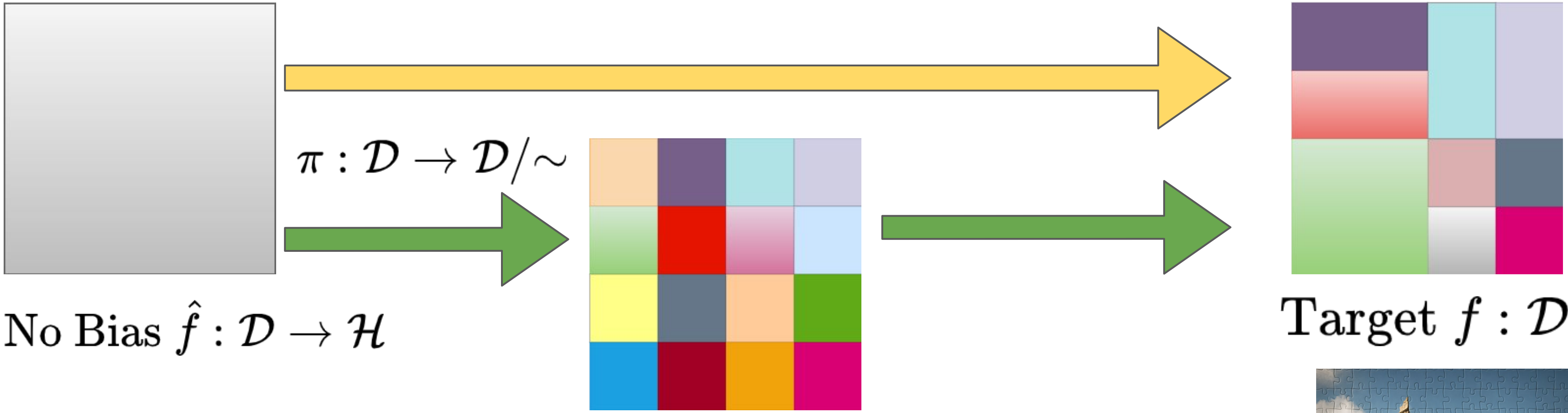
A strong inductive bias is the assumption of a partition $[\mathbf{x}]_\sim$ on $\mathcal{D}$ such that $\hat{f} : \mathcal{D} \rightarrow \mathcal{H}$ cannot be unequal within a single $[\mathbf{x}]_\sim$.

Assume a quotient topology on the domain $\pi : \mathcal{D} \rightarrow \mathcal{D}/\sim$

Correct Inductive Bias

$\mathcal{D}$

$\mathcal{D}/\sim_f = \{f^{-1}(\mathbf{y}) : \mathbf{y} \in \mathrm{Im}(f)\}$



$\pi : \mathcal{D} \to \mathcal{D}/\sim$

No Bias $\hat{f} : \mathcal{D} \to \mathcal{H}$

Target $f : \mathcal{D} \to \mathcal{H}$

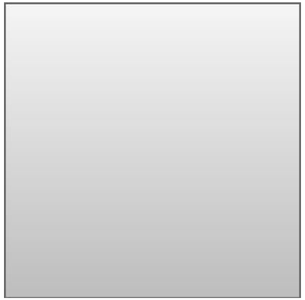
Biased $\hat{f} : \mathcal{D}/\sim \to \mathcal{H}$

Jigsaws generated by ChatGPT

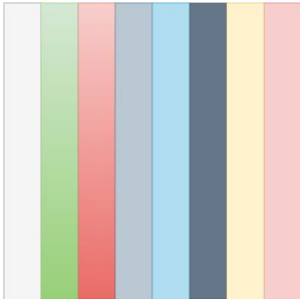
Incorrect Inductive Bias

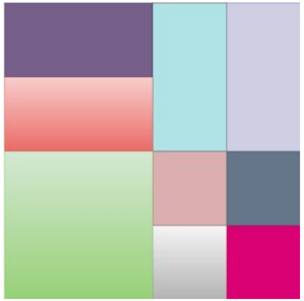
$\mathcal{D}$

$\mathcal{D}/\sim_f = \{f^{-1}(\mathbf{y}) : \mathbf{y} \in \mathrm{Im}(f)\}$



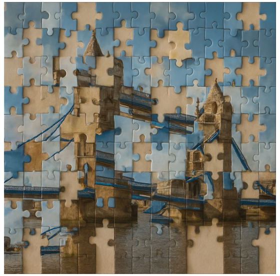
$\pi : \mathcal{D} \to \mathcal{D}/\sim$





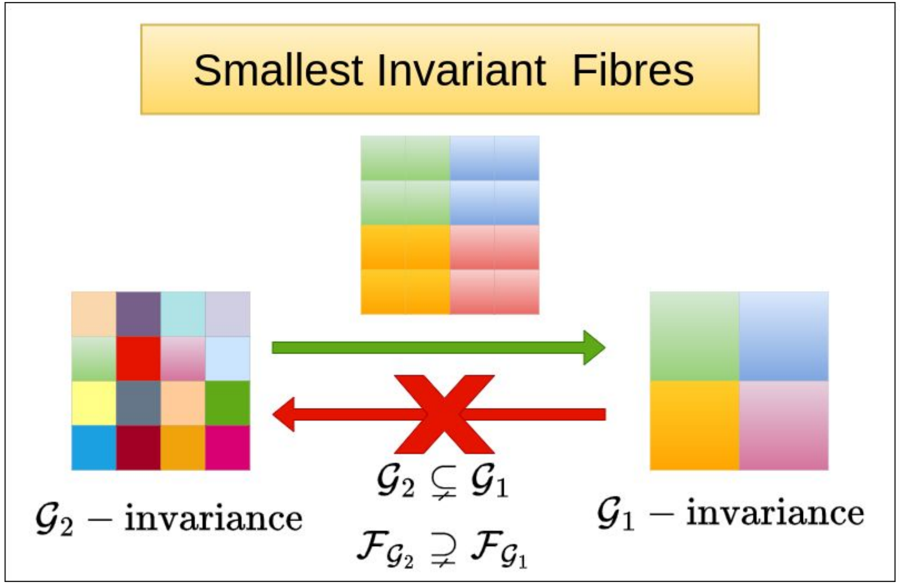
No Bias $\hat{f} : \mathcal{D} \to \mathcal{H}$

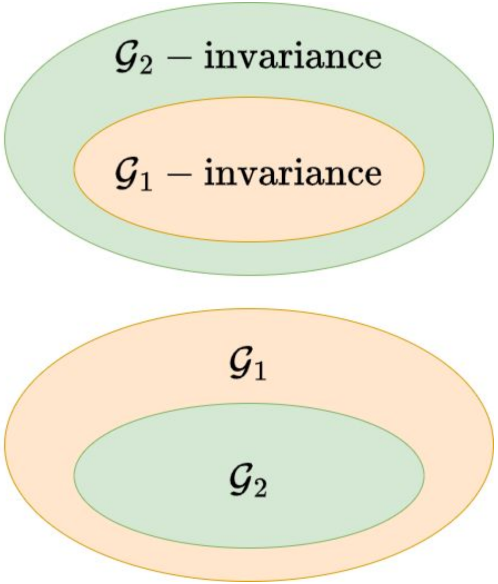
Incorrectly Biased $\hat{f} : \mathcal{D}/\sim \to \mathcal{H}$



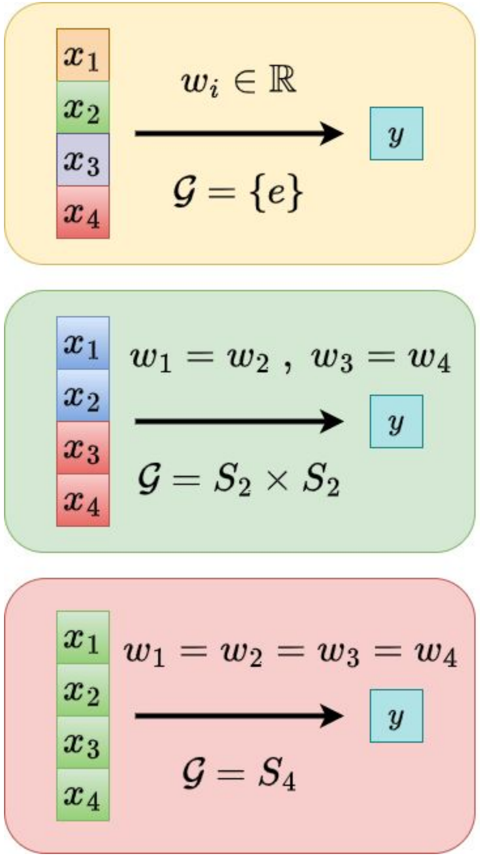
Target $f : \mathcal{D} \to \mathcal{H}$

Inverted Hierarchy of Group Invariant Function spaces

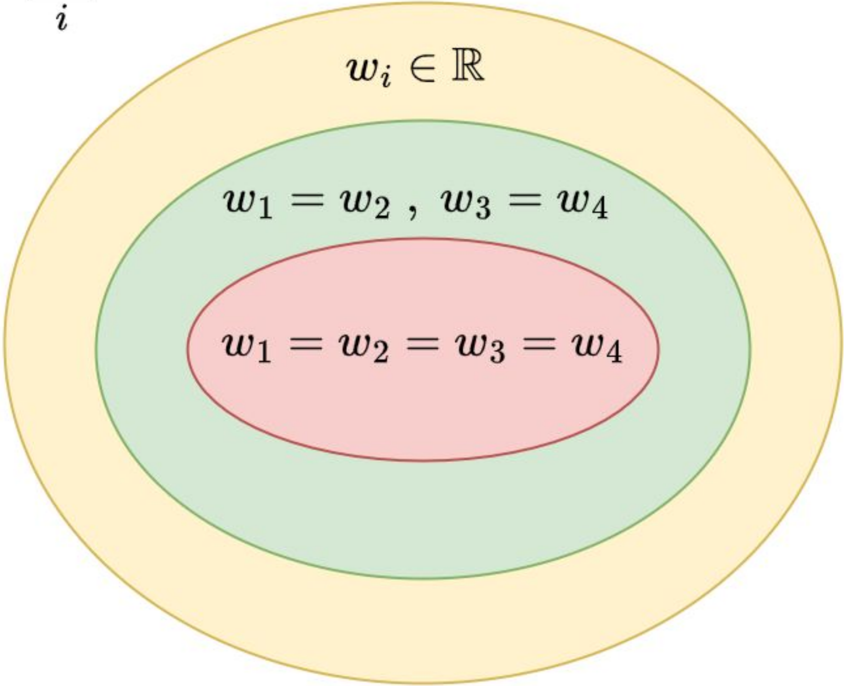




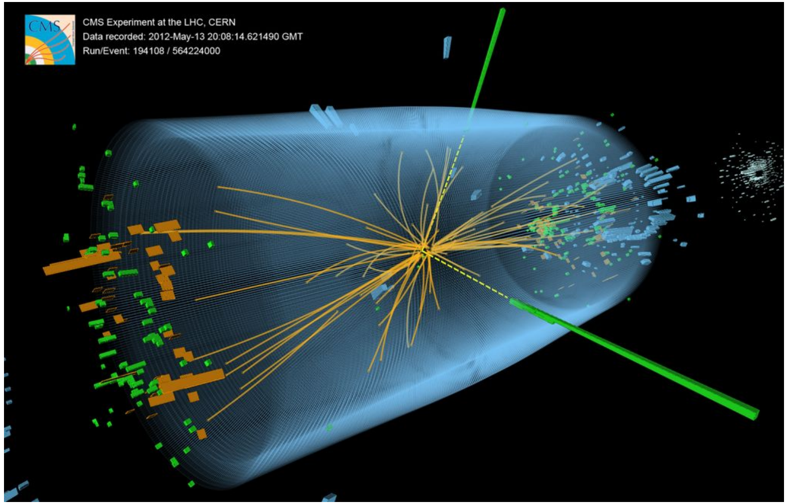
Inverted Hierarchy Example



$$y = \sum_i w_i x_i$$



Point Cloud Representation of Events

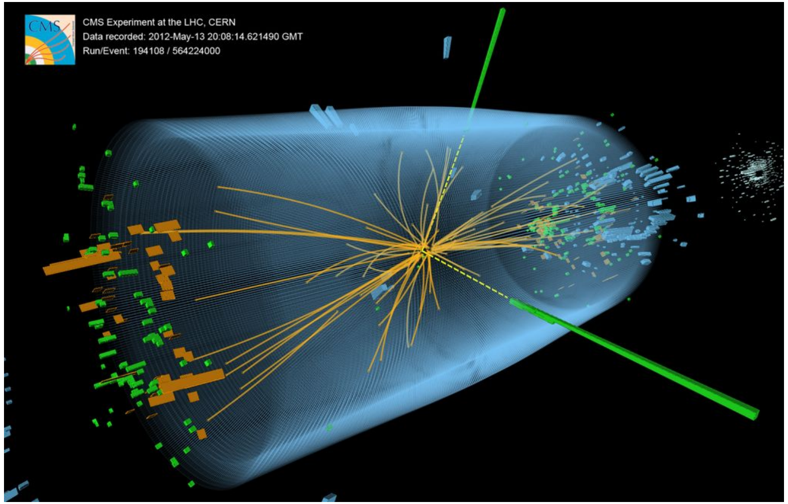


Inherent advantages of point cloud representation

- No assumed order
  - Order has no inherent meaning
  - A set of measured particles
- Variable cardinality
  - Combining event signatures with additionally radiated jets

Should we consider the event representation as a point cloud?

Point Cloud Representation of Events



$$\mathbf{E} = (\mathbf{p}_1 \oplus \mathbf{h}_1, \mathbf{p}_2 \oplus \mathbf{h}_2, \ldots, \mathbf{p}_N \oplus \mathbf{h}_N)$$

$$\mathbf{p}_i \equiv (E_i, \vec{p}_i)$$

$$\mathbf{h}_i \equiv \text{Other information like charge, object type etc.}$$

Is the target function S_N invariant?

Is the target function Lorentz invariant?

Representation of Reconstructed Events

$$\mathbf{E} = (\mathbf{p}_1^\gamma, \mathbf{p}_2^\gamma, \mathbf{p}_3^J)$$

$m_{12} \in [m_H - \Delta, m_H + \Delta] \quad , \quad m_{13} \notin [m_H - \Delta, m_H + \Delta]$

$\implies$ Likely to be Higgs candidate event

$$\mathbf{E}' = (\mathbf{p}_1^\gamma, \mathbf{p}_3^J, \mathbf{p}_2^\gamma)$$

$m_{13} \in [m_H - \Delta, m_H + \Delta] \quad , \quad m_{12} \notin [m_H - \Delta, m_H + \Delta]$

$\implies$ Not likely to be Higgs candidate event

$$S_n\text{-invariant} \implies y(\mathbf{E}) = y(\mathbf{E}')$$

Symmetries from Matrix-Element Method

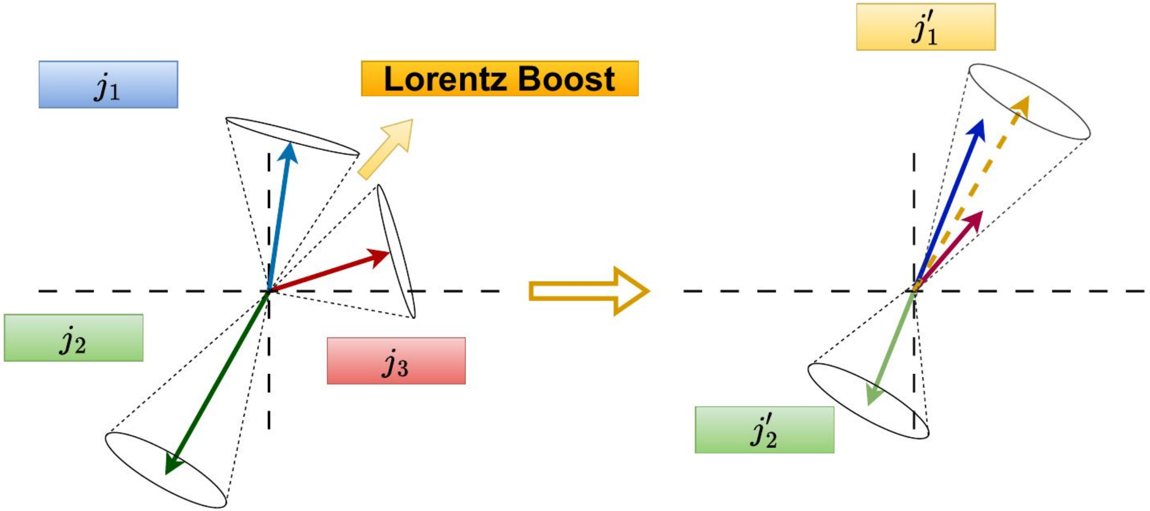
$$p_i(\mathbf{E}|\theta) = \frac{1}{\sigma_i} \int d\Pi_n(\mathbf{P})\, dx_1\, dx_2\, \frac{f_a(x_1) f_b(x_2)}{2E_{cm}\, x_1 x_2} |\mathcal{M}_i(x_1\mathbf{q}_1, x_2\mathbf{q}_2, \mathbf{P}, \theta)|^2\, T(\mathbf{E}, \mathbf{P})$$

$T(\mathbf{E}, \mathbf{P}) = \text{Transfer Function (decides the symmetries)}$

1. Is Lorentz Invariant for a Single Process

2. Adds additional discrete symmetries for experimentally indistinguishable particles

Symmetries from Matrix-Element Method



$$H = \{pp \to jj, pp \to jjj, \ldots\}$$

$$\sigma_H = \sum_i \sigma_i$$

$$p_H(\mathbf{E}|\theta) = \frac{1}{\sigma_H} \sum_{i \in \mathcal{P}_H} \sigma_i \, p_i(\mathbf{E}|\theta)$$

In the sample space of selected events:

Longitudinal Boost invariant but not Lorentz Invariant!

Discrete Symmetries

$$\lambda(\mathbf{E}) = \frac{p_S(\mathbf{E})}{p_B(\mathbf{E})}$$

$S : p\, p \rightarrow W^+\, W^-, W \rightarrow l^+\, \nu, W^- \rightarrow l^-\, \bar{\nu}$

$B : p\, p \rightarrow Z\, Z, Z \rightarrow l^+\, l^-, Z \rightarrow \nu\, \bar{\nu}$

$\lambda(\mathbf{E})$ has no discrete symmetries

$S : p\, p \rightarrow Z\, h\, , Z \rightarrow l^+ l^-, h \rightarrow b\bar{b}$

$B : p\, p \rightarrow Z\, h\, , Z \rightarrow l^+ l^-, Z \rightarrow b\bar{b}$

NWA: $S_2 \times S_2$ invariant $\lambda(\mathbf{E})$

!NWA: S_2 invariant $\lambda(\mathbf{E})$

$S : p\, p \rightarrow hh + \text{jets}, h \rightarrow bb, h \rightarrow bb$ [Non-resonant]

$S : p\, p \rightarrow H, H \rightarrow hh + \text{jets}, h \rightarrow bb, h \rightarrow bb$ [Resonant]

$B : p\, p \rightarrow 4b + \text{jets}$

NWA: $S_2 \times S_2$ invariant $\lambda(\mathbf{E})$

!NWA: S_4 invariant $\lambda(\mathbf{E})$

Di-Higgs to four Bottom Quarks [Resonant]

Architecture/Continuous Group	**Discrete Symmetry**	**Num. Params.**	**AUC**	$\mathbf{R_{30}}$
O(1,1)-Scalar	S_4	293k	0.9652±0.0002	2135±303
		22k	0.9647±0.0005	2000±139
O(1,3)	$S_2 \times S_2$	743k	0.9550±0.0016	865±67
		52k	0.9542±0.0011	864±66
SPA-NET*	S_N	37.9M	0.961 ± 0.001	—

*JHEP 09 (2024) 139 Based on attention mechanism proposed in Phys Rev D 105 112008 for solving combinatorial ambiguities at reconstruction

Public dataset: https://zenodo.org/records/10952296

Di-Higgs to four Bottom Quarks [Non-Resonant]

Architecture/Continuous Group	Discrete Symmetry	Num. Params.	AUC	R_{30}
O(1,1)-Scalar	S_4	293k	0.9160±0.0009	236±20
		22k	0.9165±0.0005	256±25
O(1,3)	$S_2 \times S_2$	743k	0.9000±0.0009	130±7
		52k	0.8995±0.0014	129±10
SPA-NET*	S_N	541k	0.911 ± 0.001	—

*JHEP 09 (2024) 139 Based on attention mechanism proposed in Phys Rev D 105 112008 for solving combinatorial ambiguities at reconstruction

Public dataset: https://zenodo.org/records/10952296

Summary

- Symmetries form the backbone of fundamental physics
  - Shows up as **inductive biases** in machine learning tasks
  - In contrast to industrial ML applications, first-principles demand various exact symmetries of the ***target function***
- Matrix-elements mandate Lorentz invariant final state distributions
  - Finite detector resolution (e.g jet radius) allows at most longitudinal boost invariant target functions for phenomenological tasks
  - Target function for most classification is **not S_N invariant**
  - **S_N invariant** only when the final state consists of a single type of reconstructed object

Backup

Inputs

Decomposed four vector

$$\tilde{\mathbf{x}}_i^{(0)} \equiv (p_x, p_y) \quad \mathbf{x}_i^{(0)} \equiv (p_z, E) \qquad \mathbf{p}_i^{(l)} = \tilde{\mathbf{x}}_i^{(0)} \oplus \mathbf{x}_i^{(l)}$$

Scalar node features

$$\mathbf{h}_i^{(0)} = (\phi_i, \log p_i^t, \log m_i^t, b_i, \log m_i)$$

$b_i = \text{B-jet tag}$

Scalar edge features

$$\mathbf{e}_{ij}^{(0)} = \left(\log p_{ij}^t, \log\left(p_i^t p_j^t\right), \Delta\eta_{ij}, \Delta\phi_{ij}, \Delta R_{ij}\right)$$

Equivariant Operation

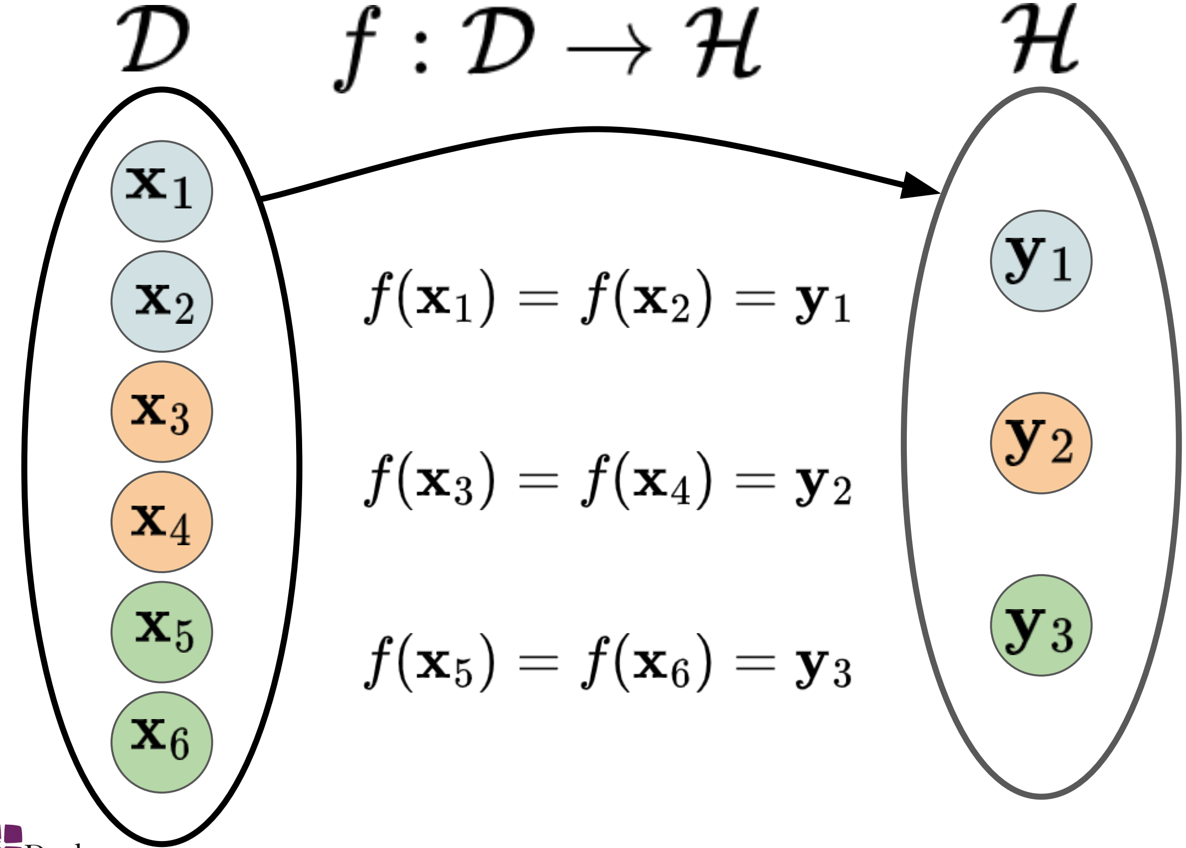
$$\mathbf{m}_{ij}^{(l+1)} = \Phi_e^{(l+1)}\left(\mathbf{h}_i^{(l)}, \mathbf{h}_j^{(l)}, \mathbf{e}_{ij}^{(l)}, |\mathbf{p}_{ij}^{(l)}|^2_{(1,2)}, \langle \mathbf{p}_i^{(l)}, \mathbf{p}_j^{(l)} \rangle_{(1,2)}, |\mathbf{p}_{ij}^{(l)}|^2, \langle \mathbf{p}_i^{(l)}, \mathbf{p}_j^{(l)} \rangle\right) \quad ,$$

$$\mathbf{x}_i^{(l+1)} = \mathbf{x}_i^{(l)} + \sum_{j \in \mathcal{N}(i)} \mathbf{x}_j^{(l)}\, \Phi_x^{(l+1)}\left(\mathbf{m}_{ij}^{(l+1)}\right) \quad ,$$

$$\mathbf{m}_i^{(l+1)} = \frac{1}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} \mathbf{m}_{ij}^{(l+1)} \quad ,$$

$$\mathbf{h}_i^{(l+1)} = \Phi_h^{(l+1)}(\mathbf{h}_i^{(l)}, \mathbf{m}_i^{(l+1)}) \quad .$$

Fibre of Functions



$$f(\mathbf{x}_1) = f(\mathbf{x}_2) = \mathbf{y}_1$$

$$f(\mathbf{x}_3) = f(\mathbf{x}_4) = \mathbf{y}_2$$

$$f(\mathbf{x}_5) = f(\mathbf{x}_6) = \mathbf{y}_3$$

$$f^{-1}(\mathbf{y}_1) = \{\mathbf{x}_1, \mathbf{x}_2\}$$

$$f^{-1}(\mathbf{y}_2) = \{\mathbf{x}_3, \mathbf{x}_4\}$$

$$f^{-1}(\mathbf{y}_3) = \{\mathbf{x}_5, \mathbf{x}_6\}$$

Fibre of Functions

Mutually Exclusive

$$\mathbf{y} \neq \mathbf{y}' \Rightarrow f^{-1}(\mathbf{y}) \cap f^{-1}(\mathbf{y}') = \emptyset$$

Covers the domain

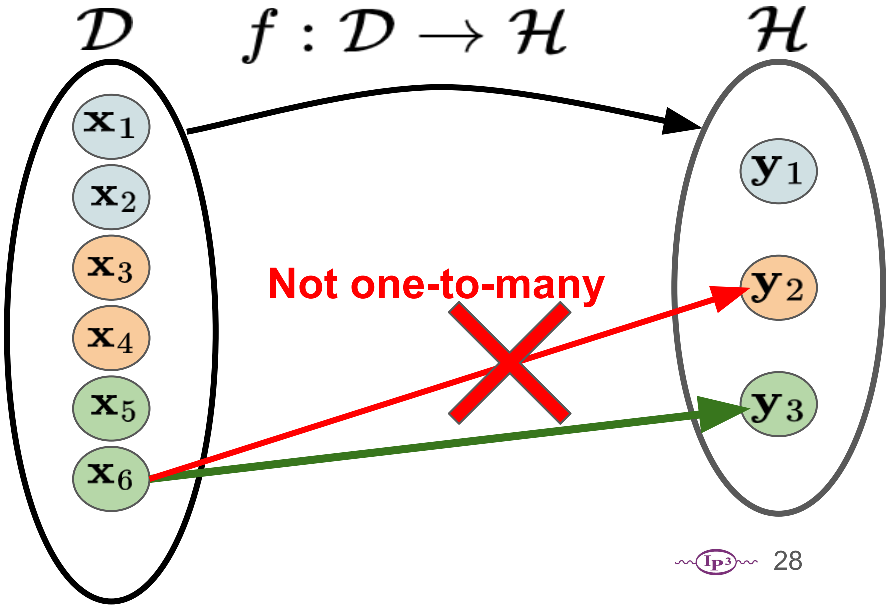
$$\bigcup_{\mathbf{y} \in Im(f)} f^{-1}(\mathbf{y}) = \mathcal{D}$$

Fibres of a function partition the domain!

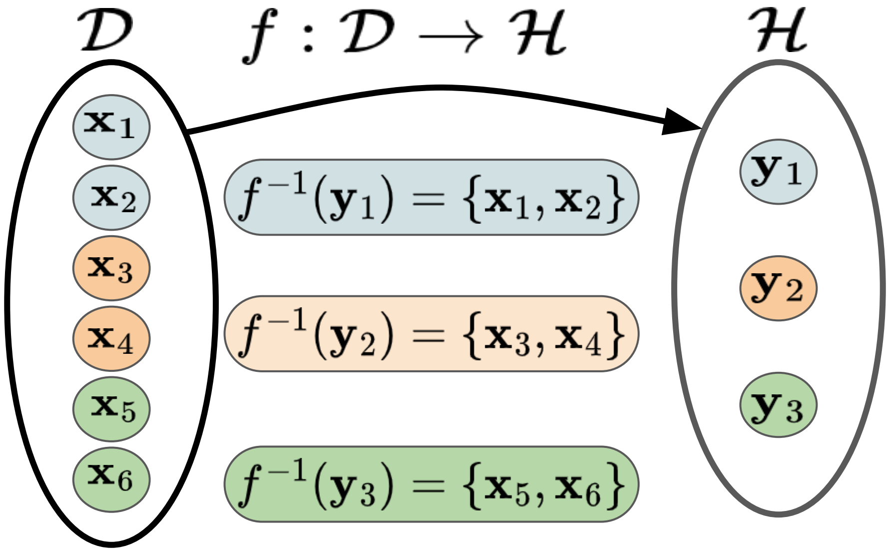
$f^{-1}(\mathbf{y}_1) = \{\mathbf{x}_1, \mathbf{x}_2\}$

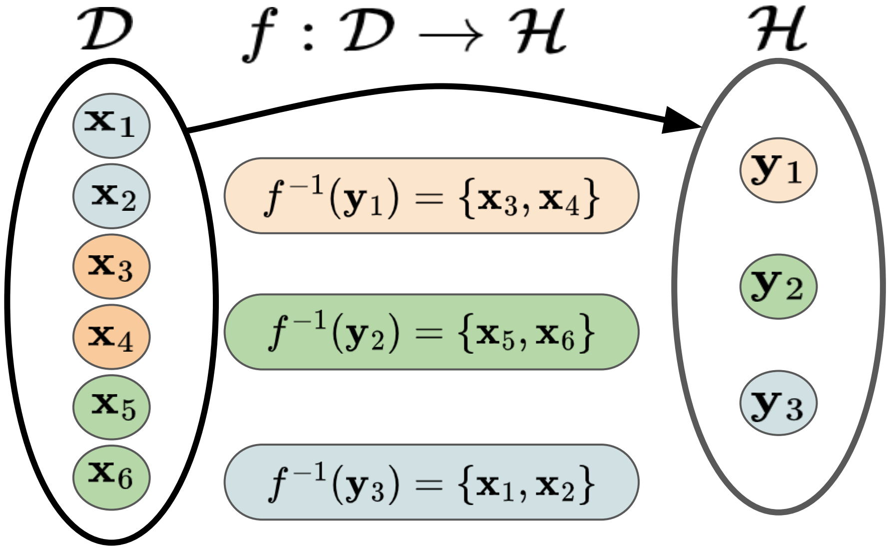
$f^{-1}(\mathbf{y}_2) = \{\mathbf{x}_3, \mathbf{x}_4\}$

$f^{-1}(\mathbf{y}_3) = \{\mathbf{x}_5, \mathbf{x}_6\}$



Fibre of Functions





A unique fibre decomposition corresponds to a ***family of functions***