Group Equivariance in Infrared and Collinear Safe Graph Neural Networks

JHEP 03 (2026) 219

P. Konar, **VN**, M. Spannowsky, D. Srivastava

Karlsruher Institut für Technologie

Institute for Theoretical Physics

Vishal Ngairangbam

vishal.ngairangbam@kit.edu

Inductive Bias I: Infrared and Collinear safety

Int. J. Mod. Phys. A12 (1997) 5411-5529 JHEP 04 (2018) 013 JHEP 01 (2019) 121
Phys. Rev. D 103, 074022 (2021) JHEP 09 (2023) 135

Any IRC safe observable can be expanded in terms of C (calorimeter) - correlators

$$\mathcal{O} \approx \sum_{N=0}^{N_{max}} \mathcal{C}_N^{f_N} \qquad \mathcal{C}_N^{f_N} = \sum_{i_1} \sum_{i_2} \dots \sum_{i_N} E_{i_1} E_{i_2} \dots E_{i_N} \, f_N(\hat{p}_{i_1}, \hat{p}_{i_2}, \dots . \hat{p}_{i_N})$$

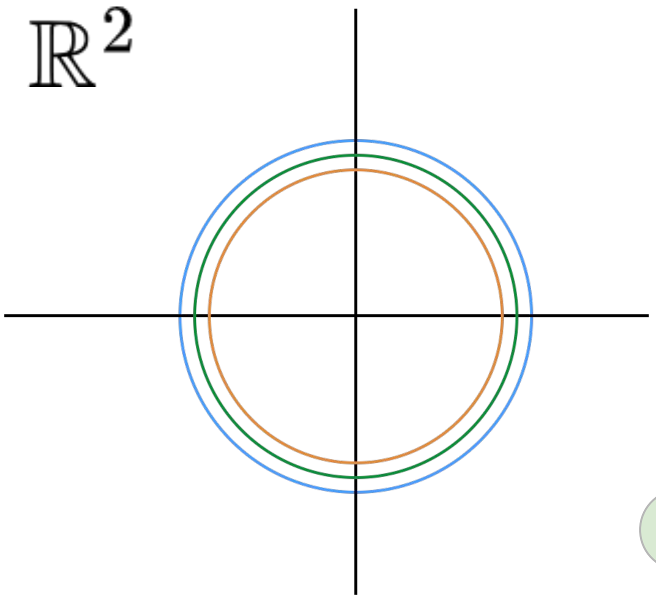
For hadron colliders: $E_i \to p_{Ti}, \hat{p}_i = (y, \phi)$

$$f_N = \text{Permutation invariant with respect to its arguments}$$

$$\sum_{i=1}^{|\mathcal{S}|} \to \mathcal{S} = \{p_1^\mu, p_2^\mu, \dots \dots\}$$ (set of particles/calorimeter hits)

$$|\mathcal{S}| \text{ finite but not fixed for each measured jet/event}$$

Inductive Bias II: Group Symmetries



$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

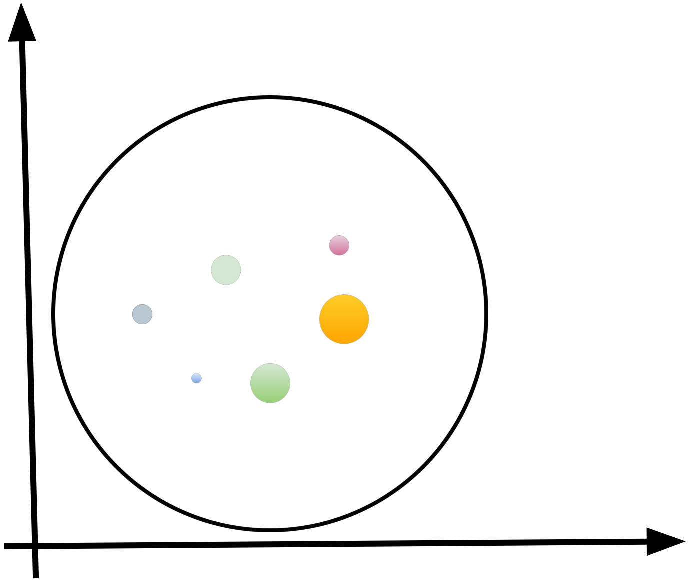
SO(2) group action

$$\mathbf{X}' = \rho(\theta)\ \mathbf{X}$$

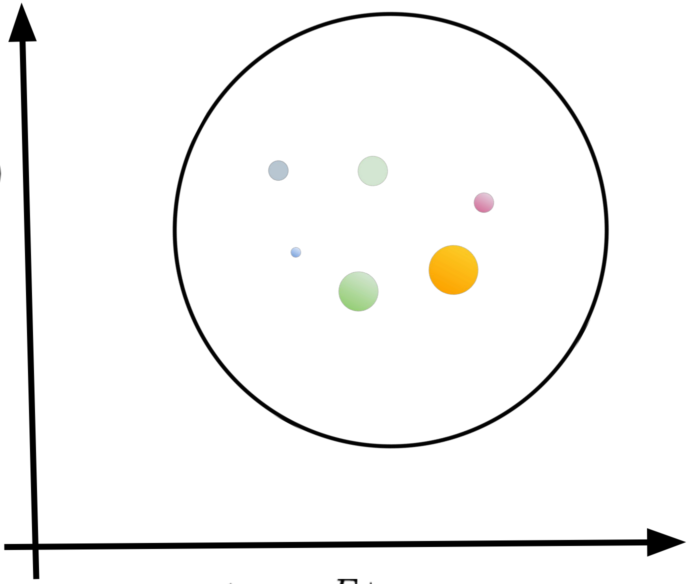
SO(2) invariant target $f(x', y') = f(x, y)$

$\hat{f}(r, \theta) = \text{No bias}$ $\hat{f}(r) = \text{SO(2) biased}$

IRC Safety + Symmetries on the rapidity-azimuth plane



$$\phi = \tan^{-1}(\frac{p_x}{p_y})$$



$$y = \frac{1}{2}\log\frac{E+p_z}{E-p_z}$$

$$z_i = \frac{p_{ti}}{\sum_j p_{tj}}$$

$$(z_i, \mathbf{x}_i) \mapsto (z_i, \mathbf{R}(\theta)\ \mathbf{x}_i + \mathbf{t})$$

Inductive Bias: Function Space

Assumptions on the nature of the problem through which one biases the possible solutions to always follow a certain property (Strong), or prefer (Weak) some solutions over others

$$\Sigma_W = \text{Set of functions represented by a model class}$$

$$W = \text{Parameter space to tune to reach a viable solution } \Theta_0$$

1. The class of functions accessible via Σ_W
2. Given an initial point $\Theta \sim P_W(\Theta)$ prefer some Θ_0 over others

Inductive Bias: Input Domain

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

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

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

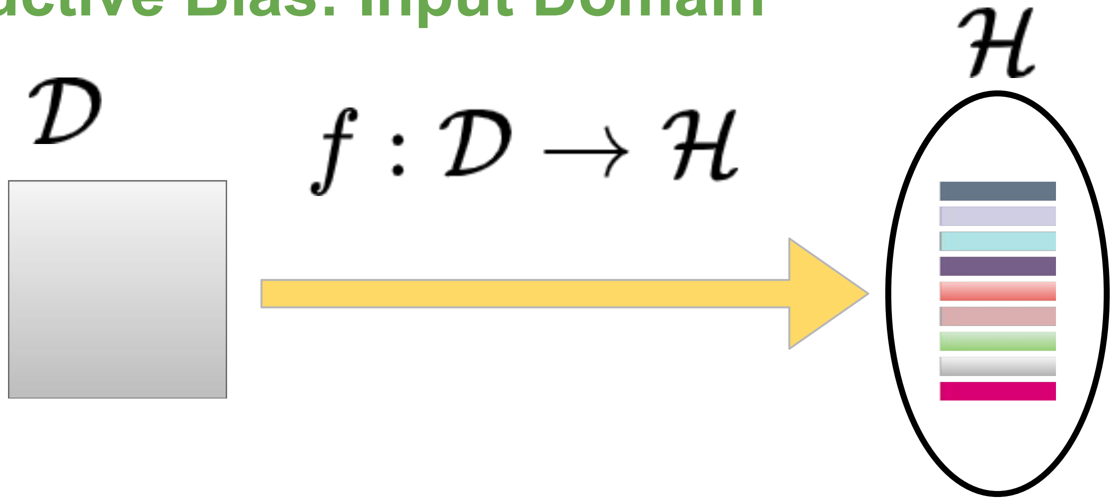
2. Match $\hat{f}([\mathbf{x}]_f) = f([\mathbf{x}]_f) \quad \forall \quad [\mathbf{x}]_f$

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

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

Inductive Bias: Input Domain

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

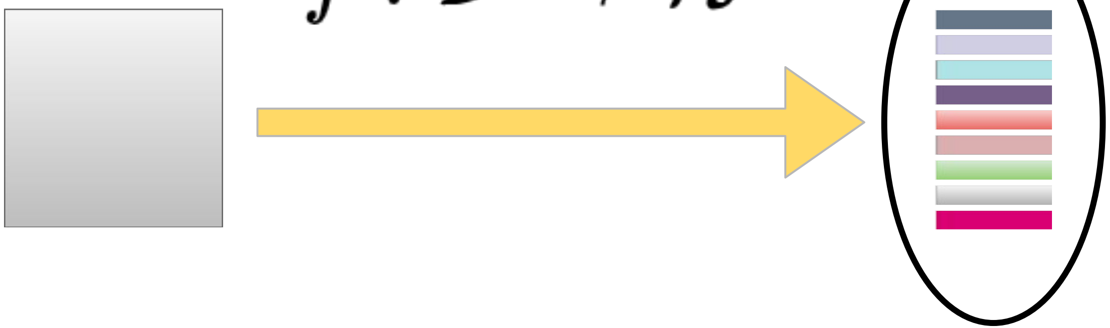


Inductive Bias: Input Domain

$\mathcal{D}$

$$f : \mathcal{D} \rightarrow \mathcal{H}$$

$\mathcal{H}$



Known domain partition where target function must be equal

Target function's value is unknown in most of the domain

Incorrect Inductive Bias

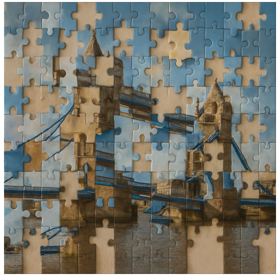
$\mathcal{D}$

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

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

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

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



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

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

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

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

Jigsaws generated by ChatGPT

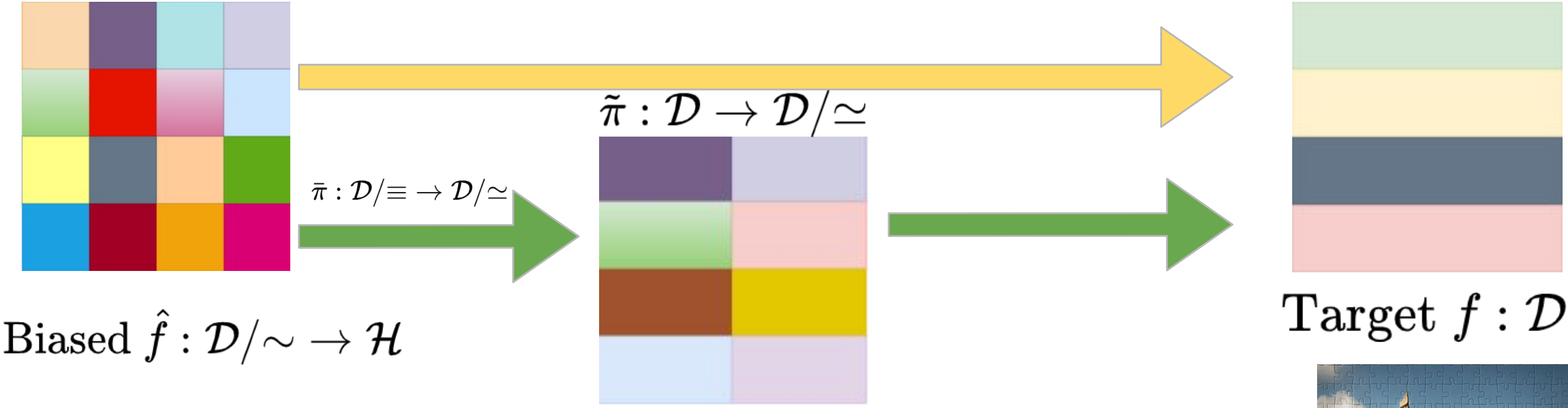
Compatible Inductive Biases

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

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

$\tilde{\pi} : \mathcal{D} \to \mathcal{D}/\simeq$

$\bar{\pi} : \mathcal{D}/\equiv \to \mathcal{D}/\simeq$



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

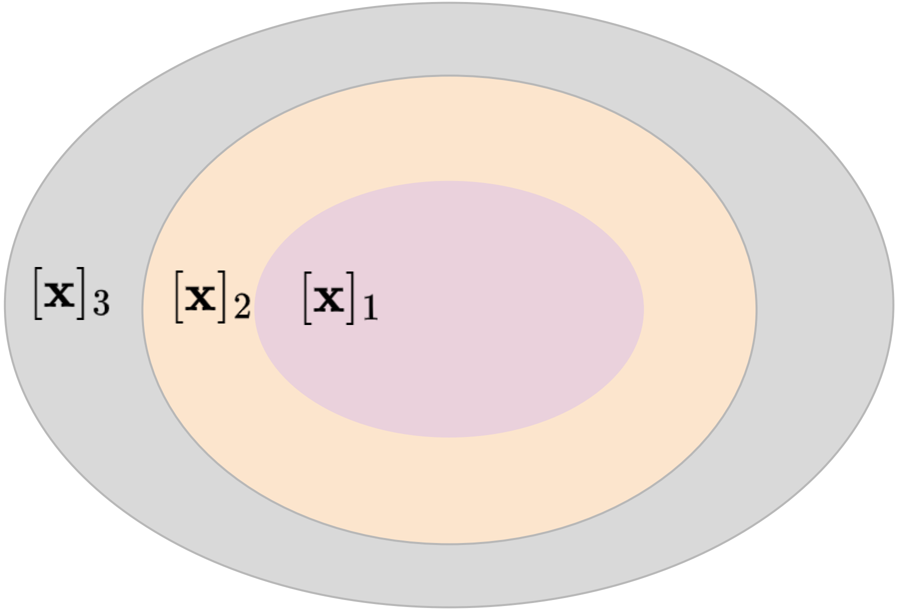
More biased $\tilde{f} : \mathcal{D}/\simeq \to \mathcal{H}$

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

Jigsaws generated by ChatGPT

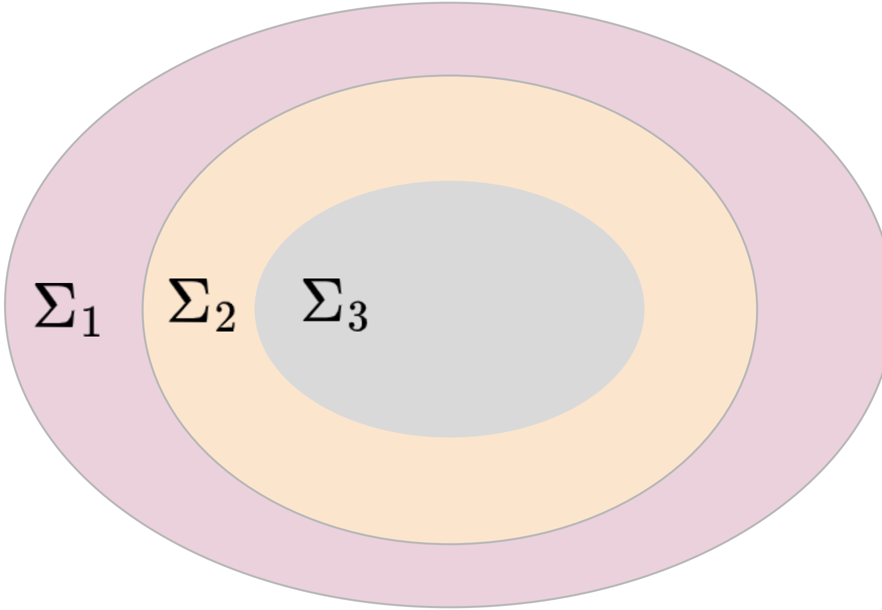
Strength of Inductive biases

Domain Partition $[\mathbf{x}]_\sim$



$$[\mathbf{x}]_1 \subsetneq [\mathbf{x}]_2 \subsetneq [\mathbf{x}]_3$$

Function Space Σ_W



$$\Sigma_1 \supsetneq \Sigma_2 \supsetneq \Sigma_3$$

Infra-red and Collinear (IRC) Safe observables

For an observable $\mathcal{O}_n$ defined on n particles.

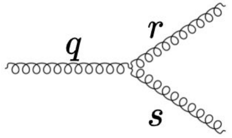
$$z_i = \frac{E_i}{\sum_j E_j}$$

$$\mathcal{O}_{n+1}(p_a, ., p_b, p_r, p_s, p_c, .) \to \mathcal{O}_n(p_a, ., p_b, p_q, p_c, .)$$

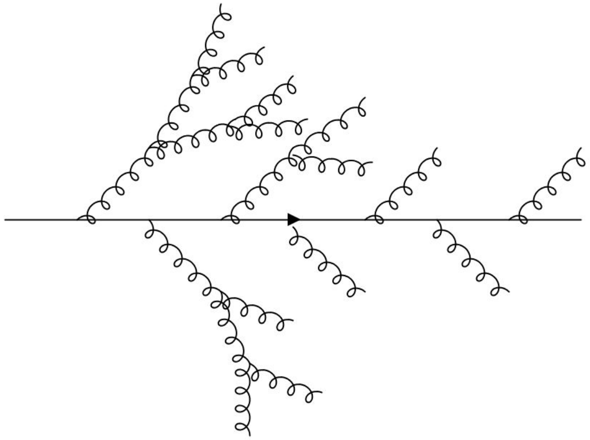
In the infra-red $(z_r \to 0 \text{ or } z_s \to 0)$ or collinear limits $(\Delta_{rs} \to 0)$

$$z_i = \frac{p_{ti}}{\sum_j p_{tj}}$$

For a splitting: $q \to r + s$



$$p_q = (z_q, \hat{p_q})$$
$$p_r = (z_r, \hat{p_r})$$
$$p_s = (z_s, \hat{p_s})$$



Calculable in pQCD!!

=> concentrates on information of the hard interaction

=> relatively insensitive to low-energy non-perturbative effects

Infra-red and Collinear (IRC) Safe observables

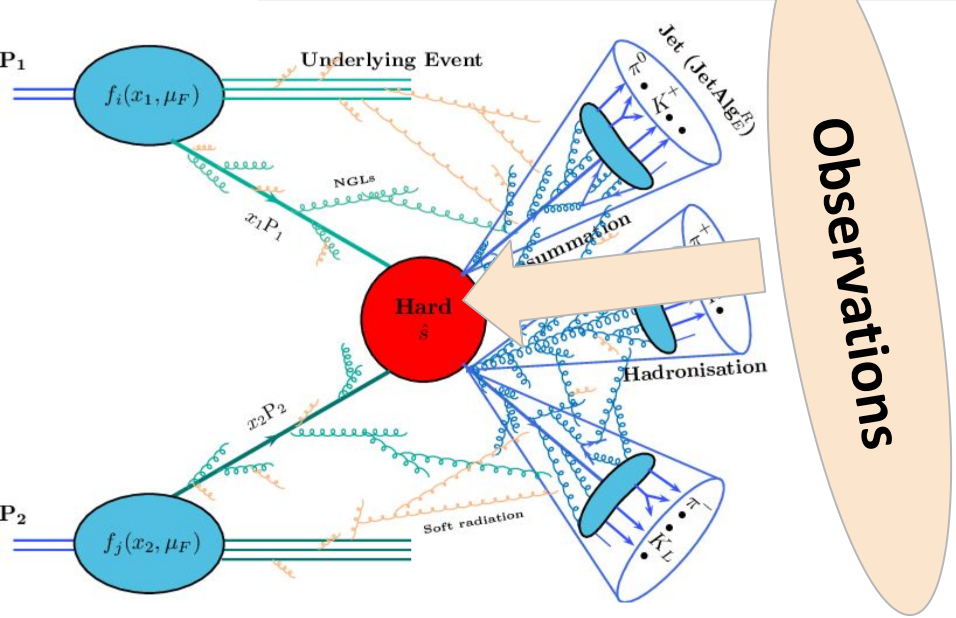
For an observable $\mathcal{O}_n$ defined on n particles.

$$z_i = \frac{E_i}{\sum_j E_j}$$

$$\mathcal{O}_{n+1}(p_a,.,p_b,p_r,p_s,p_c,.) \to \mathcal{O}_n(p_a,.,p_b,p_q,p_c,.)$$

In the infra-red $(z_r \to 0 \text{ or } z_s \to 0)$ or collinear limits $(\Delta_{rs} \to 0)$

$$z_i = \frac{p_{ti}}{\sum_j p_{tj}}$$



Calculable in pQCD!!

=> concentrates on information of the hard interaction

=> relatively insensitive to low-energy non-perturbative effects

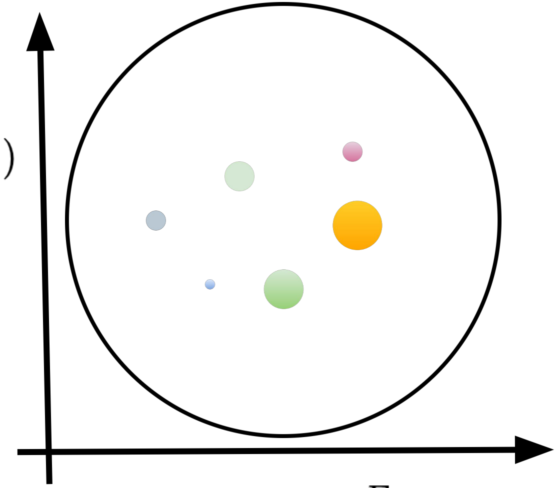
IRC Safety: quotient under indistinguishability of jets/events

Two Jets are equivalent if they can not be **Observationally Distinguished***

$$z_i = \frac{p_{ti}}{\sum_j p_{tj}}$$

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

$$\phi = \tan^{-1}\left(\frac{p_x}{p_y}\right)$$



$$\text{Continuity of } \mathcal{O} : \mathcal{X}/\sim \to \mathcal{Y}$$

$$\mathcal{O}_{n+1}(p_a, .., p_b, p_r, p_s, p_c, .) \to \mathcal{O}_n(p_a, .., p_b, p_q, p_c, .)$$

In the infra-red $(z_r \to 0 \text{ or } z_s \to 0)$ or collinear limits $(\Delta_{rs} \to 0)$

$$y = \frac{1}{2}\log\frac{E+p_z}{E-p_z}$$

$$p_t = \sqrt{p_x^2 + p_y^2}$$

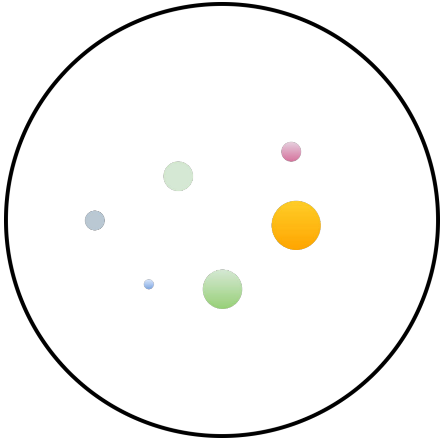
Conservation of Collinearity

$$z_i = \frac{p_i^t}{\sum_j p_j^t}$$

$$h_i^{(0)} \equiv \hat{p}_i$$

Node update Operation

$$h_i^{(1)} = \Phi_\theta(h_1^{(0)}, h_2^{(0)}, \ldots, h_n^{(0)})$$



IRC safe jet features

$$\mathbf{H}_l = \sum_i z_i \, h_i^{(l)}$$

Node update conserves collinearity if:

$$z_q = z_r + z_s \text{ and } \hat{p}_q = \hat{p}_r = \hat{p}_s$$

$$\implies h_q^{(1)} = h_r^{(1)} = h_s^{(1)}$$

$$z_q = z_r + z_s \text{ and } h_q^{(l)} = h_r^{(l)} = h_s^{(l)}$$

$$\implies h_q^{(l+1)} = h_r^{(l+1)} = h_s^{(l+1)}$$

IRC safe

$$\mathbf{H}_0 = \sum_i z_i \, \hat{p}_i$$

JHEP 02 (2022) 060

Front.Artif.Intell. 5 (2022) 943135

JHEP 01 (2024) 113

JHEP 07 (2024) 245 (*)

JHEP 03 (2026) 219 (*)

E(2) and O(2) equivariance and IRC safety

$$\omega_j^{(\mathcal{N}[i])} = \frac{p_T^j}{\sum_{k\in\mathcal{N}[i]} p_T^k}$$

$$\mathbf{m}_{ij}^{(l+1)} = \Phi_e^{(l+1)}(\mathbf{h}_i^{(l)}, \mathbf{h}_j^{(l)}, |\mathbf{x}_i^{(l)} - \mathbf{x}_j^{(l)}|^2) \quad ,$$

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

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

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

$$\mathbf{x}_i^{(\cdot)} = \text{Vector node representation}$$

$$\mathbf{h}_i^{(\cdot)} = \text{Scalar node representation}$$

arxiv:2102.09844 Satorras et. al

Hierarchy of Inductive Biases

On the Jet representation we have:

1. **Permutation Invariant**
2. **Infrared and Collinear Safety**
3. **IRC Safety + O(2) Invariance**
4. **IRC Safety + E(2) Invariance**

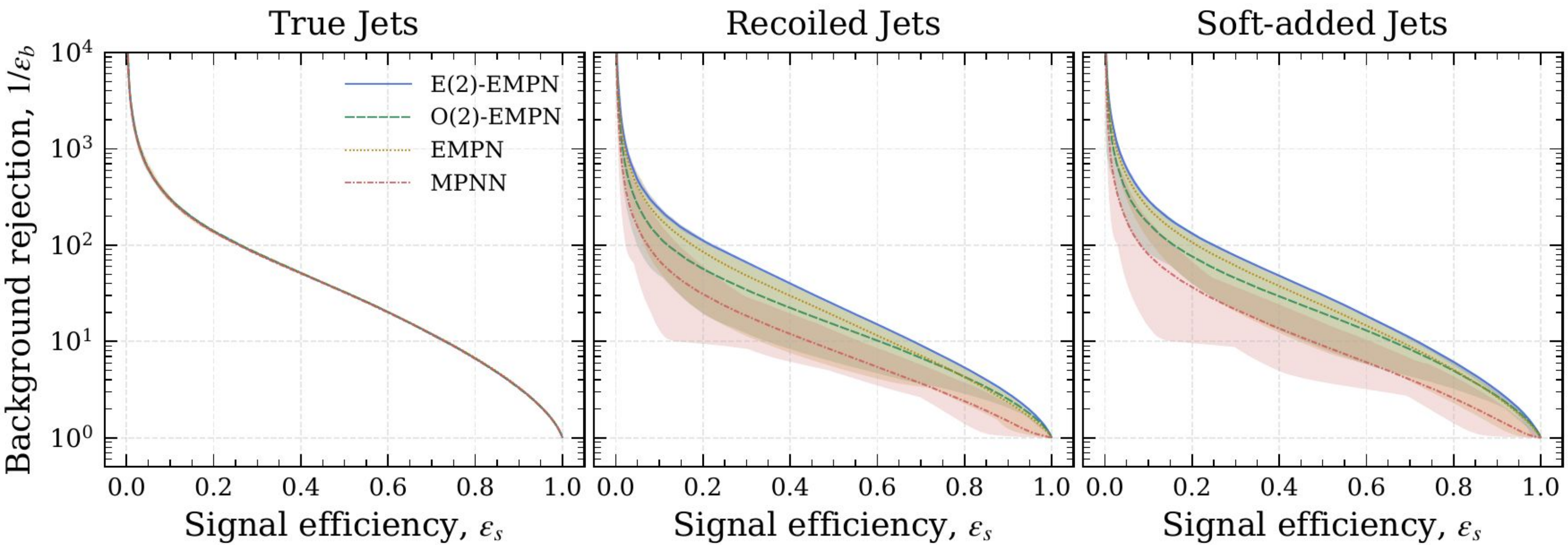
$$[\mathbf{X}]_1 \subsetneq [\mathbf{X}]_2 \subsetneq [\mathbf{X}]_3 \subsetneq [\mathbf{X}]_4$$

$$\mathbf{X} = (z_1, \Delta y_1, \Delta\phi_1, z_2, \Delta y_2, \Delta\phi_2, \ldots, z_n, \Delta y_n, \Delta\phi_n, \ldots)$$

ROC Variability: Quark vs Gluons

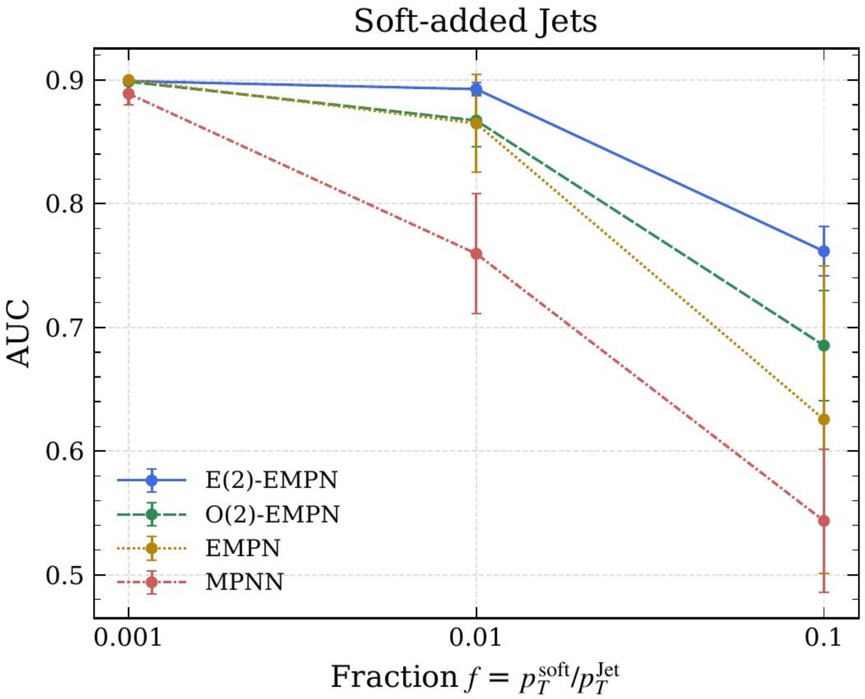
Recoil against a soft emission outside the jet

Additional wide-angle soft emission within the jet

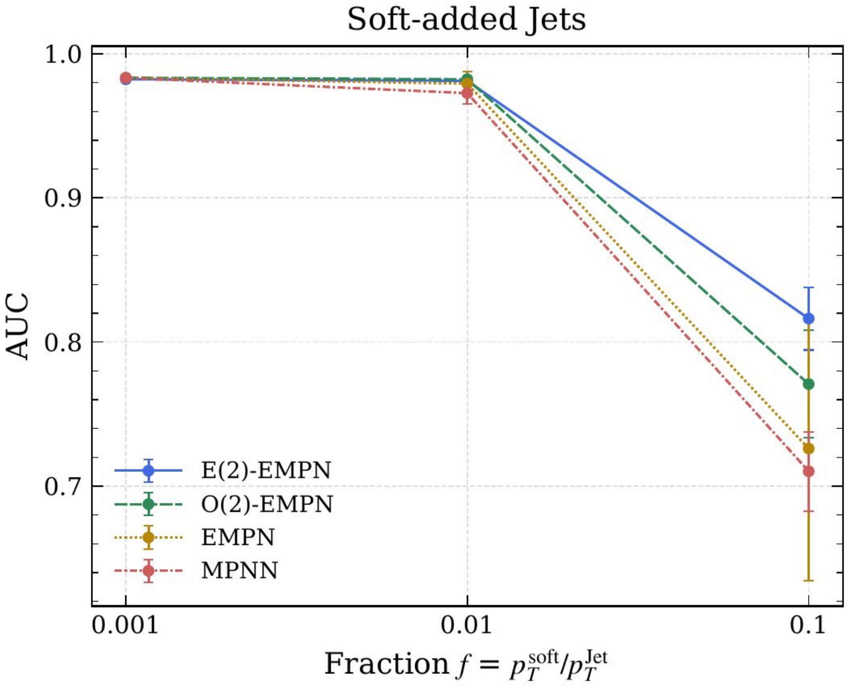


AUC Variability

Quark vs Gluon (R=0.4)



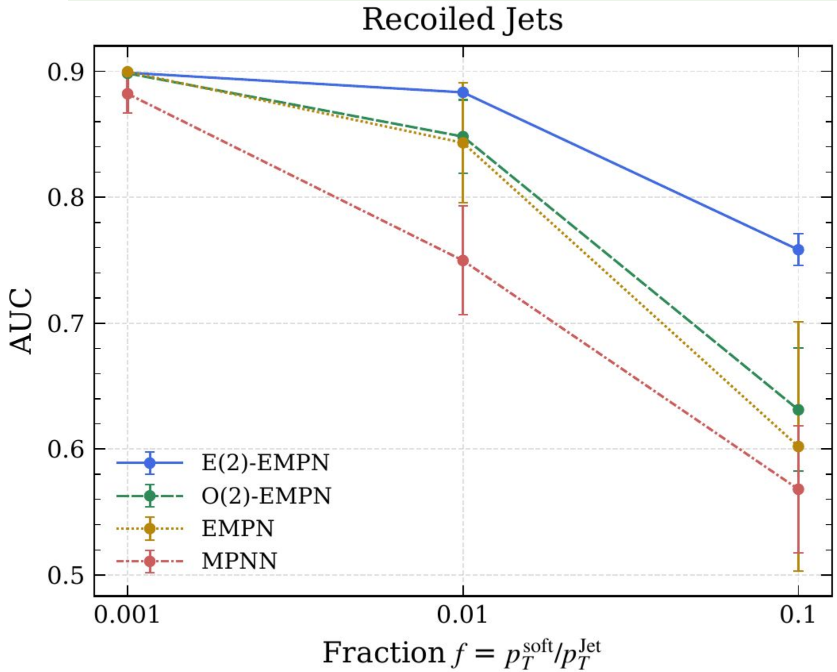
Top vs QCD (R=0.8)



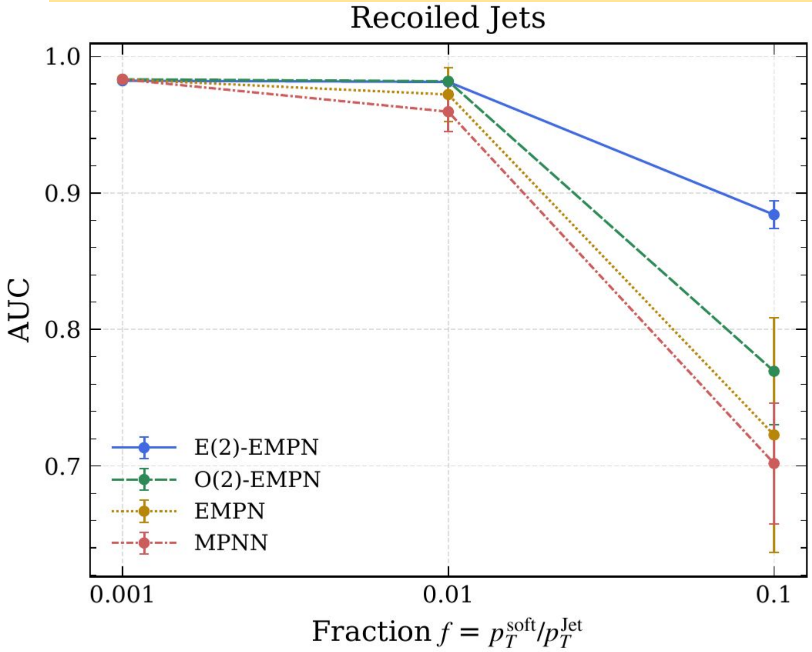
Add wide-angle soft particle to jet

AUC Variability

Quark vs Gluon (R=0.4)



Top vs QCD (R=0.8)



Recoil against a soft emission

Summary

- ***Strong Inductive Biases*** *can be viewed as an assumption of equivalence*
- *IRC safe tagging algorithms in general* ***reduce in raw performance***
- ***Gains on resilience*** *[check out* Pareto frontier studies on **arxiv:2509.19431** *and* ***arxiv:2510.06314****]*
- **Infrared and collinear safety** coupled with **E(2) invariance** shows highest stability under single particle soft recoil and emissions
  - All order effects?
  - Non Perturbative effects?