Understanding QCD at the Large Hadron Collider, Part 1

Andrew Larkoski
Reed College

New Physics from Precision at High Energies, KITP, April 15 2021

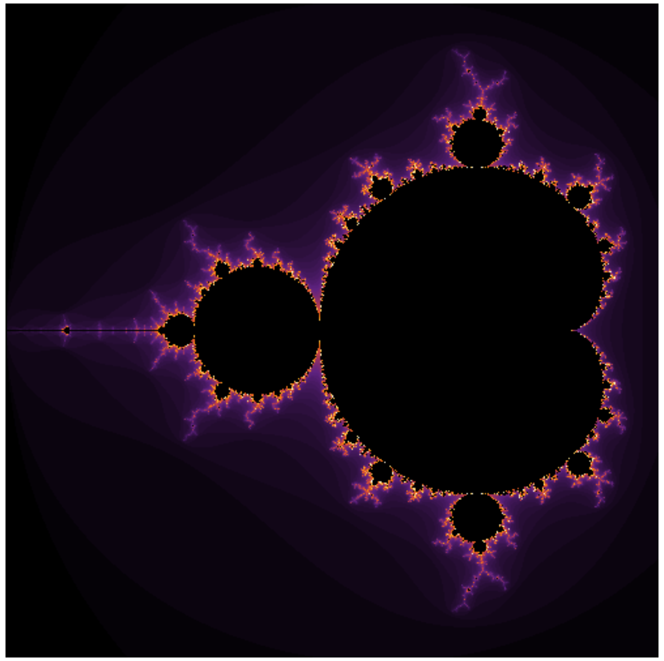
Human Learning from Thinking Like a Machine

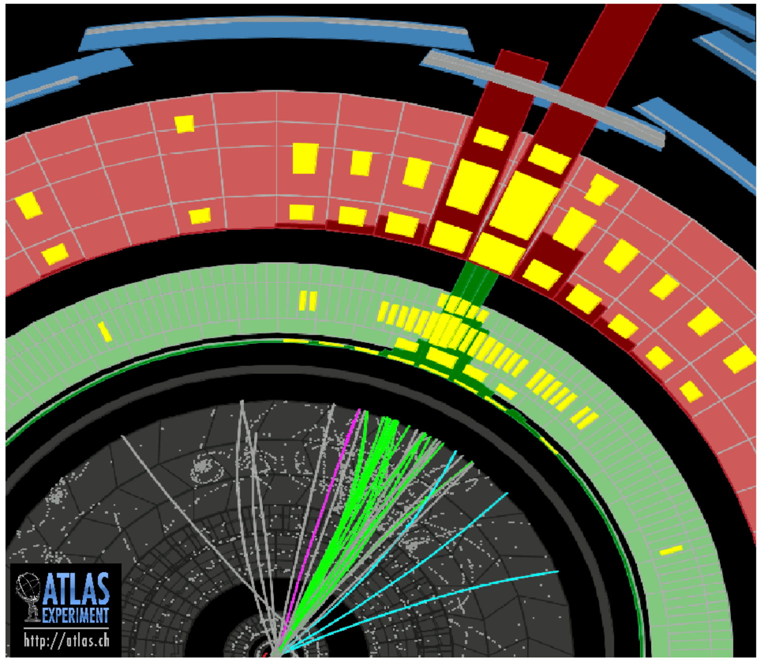
Andrew Larkoski
Reed College

New Physics from Precision at High Energies, KITP, April 15 2021

Losing the Particle Physics Forest for the ML/AI/CS Trees

Information content of a particle physics event or jet is small





arXiv:1112.6426

Kolmogorov complexity of Mandelbrot set is small

(easy to write a short computer program)

This image: 100,000s of pixels

Program to generate it: 2 lines of code

One jet with 30ish particles; ~3000 total bits

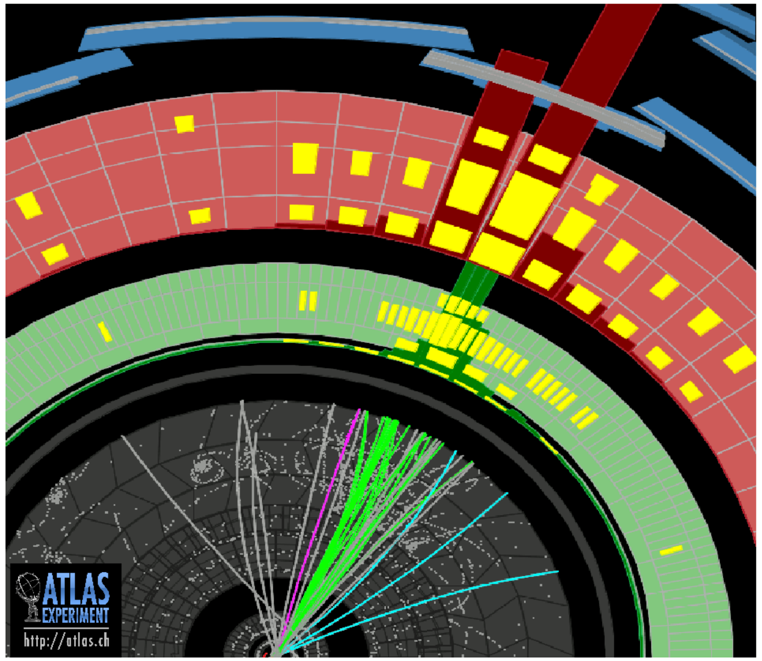
LHC has recorded > 10 billion events

Pythia + MadGraph code is about 50 MBytes

Information in physics <<<<< information of data

Losing the Particle Physics Forest for the ML/AI/CS Trees

We know the rules of particle physics and the space in which the data lives



arXiv:1112.6426

What is the space of pixels in an image of Brad Pitt?

What theory governs his chiseled brow?

Whole subfields of CS are devoted to organizing data with no known structure

Fundamental theory: the QCD Lagrangian

Particle production governed by recursive equations

Resulting particles live on relativistic phase space

Recall: A Lesson for Introductory Physics

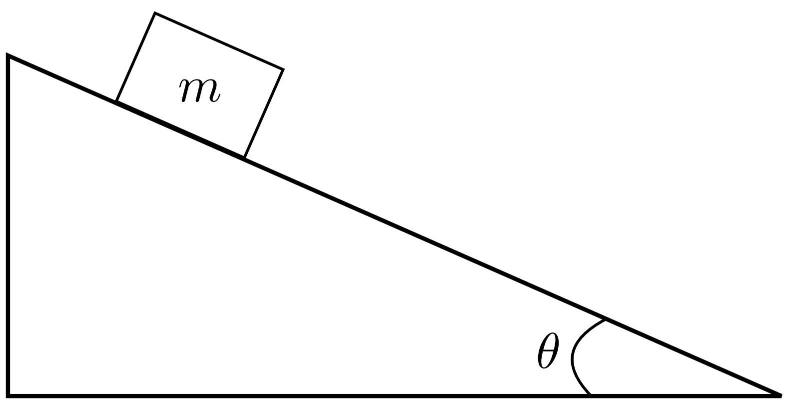
Students in introductory physics haven't honed their physical intuition

To make something habit, you have to consciously do the action 1000s of times

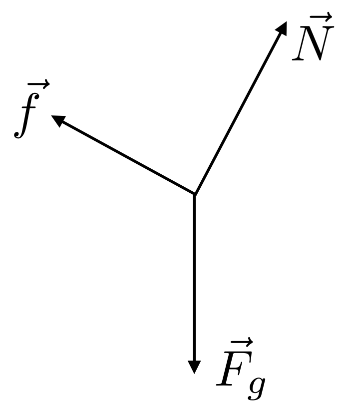
If you learn a new physics concept and are asked to solve a problem you must consciously ask:

What is the dumbest next thing that I could do?

Problem:



1st dumb thing: FBD



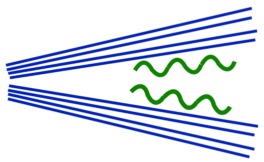
2nd dumb thing: Newton #2

$$\vec{F}_g + \vec{N} + \vec{f} = m\vec{a}$$

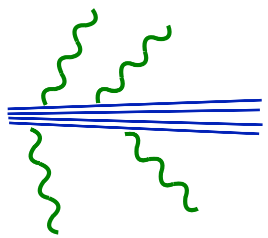
If the dumbest thing works, then it isn't so dumb, and you can completely understand everything

Example 1: Optimal Observables for Color-Singlet Identification

Problem:

W/Z/H 

vs.

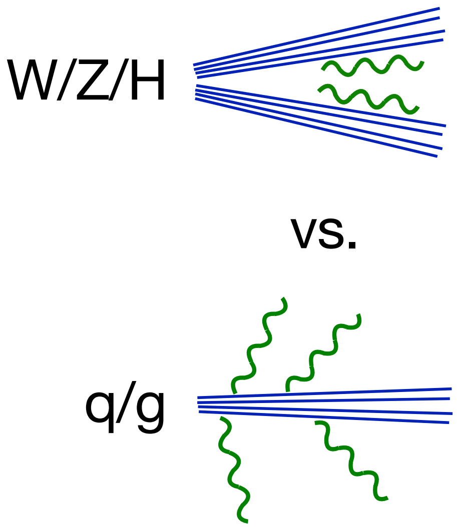
q/g 

Already assume that there is selected hard two-prong structure

arXiv:2006.10480

Example 1: Optimal Observables for Color-Singlet Identification

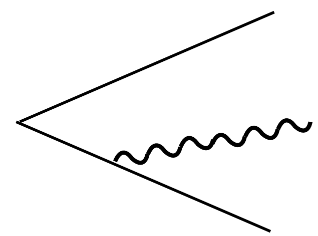
Problem:



1st dumb thing:

Jet has 3 particles,

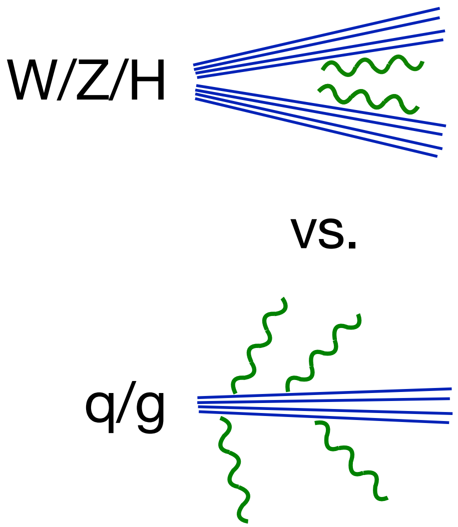
minimal for color correlations



Work in limit where third particle is
much softer than other two

Example 1: Optimal Observables for Color-Singlet Identification

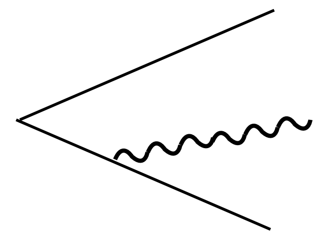
Problem:



1st dumb thing:

Jet has 3 particles,

minimal for color correlations



2nd dumb thing:

Neyman-Pearson:

Likelihood ratio is optimal

Neyman, Pearson 1933

$$\mathcal{O} = \frac{|\mathcal{M}_{\mathrm{QCD}}|^2}{|\mathcal{M}_{\mathrm{singlet}}|^2}$$

Best discriminant is some function of the location of the soft particle

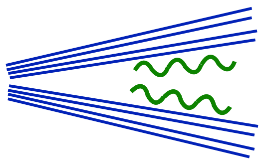
If the dumbest thing works, then it isn't so dumb, and you can completely understand everything

arXiv:2006.10480

Example 1: Optimal Observables for Color-Singlet Identification

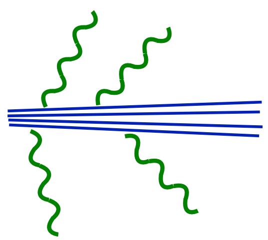
Problem:

W/Z/H



vs.

q/g



Consequences:

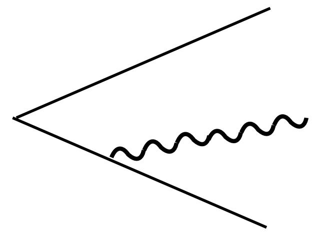
Observables “designed” to do this are suboptimal

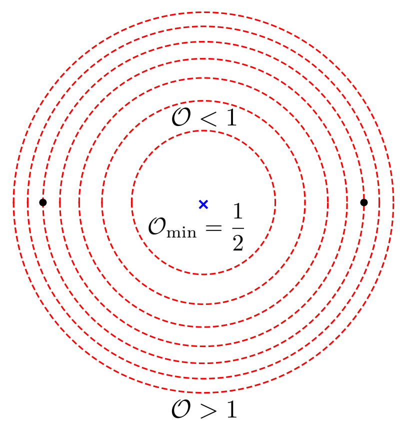
See also arXiv:2010.11998

1st dumb thing:

Jet has 3 particles,

minimal for color correlations





Contours of constant optimal observable

“jet color ring”

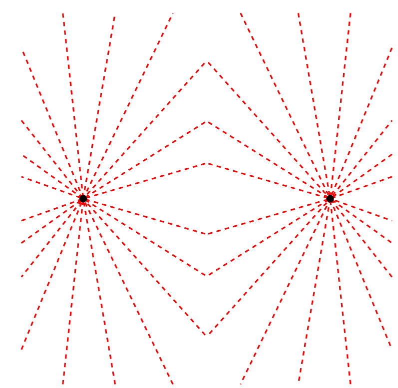
2nd dumb thing:

Neyman-Pearson:

Likelihood ratio is optimal

Neyman, Pearson 1933

$$\mathcal{O} = \frac{|\mathcal{M}_{\text{QCD}}|^2}{|\mathcal{M}_{\text{singlet}}|^2}$$



Contours of constant pull angle

arXiv:1001.5027

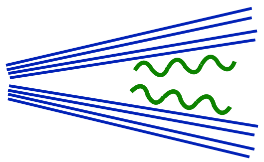
If the dumbest thing works, then it isn’t so dumb, and you can completely understand everything

arXiv:2006.10480

Example 1: Optimal Observables for Color-Singlet Identification

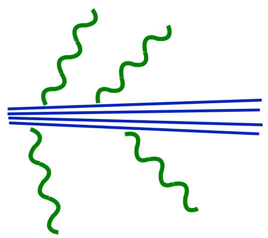
Problem:

W/Z/H



vs.

q/g



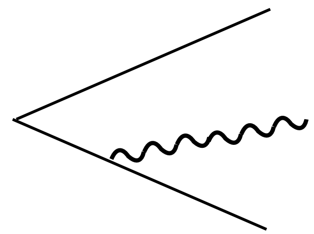
Consequences:

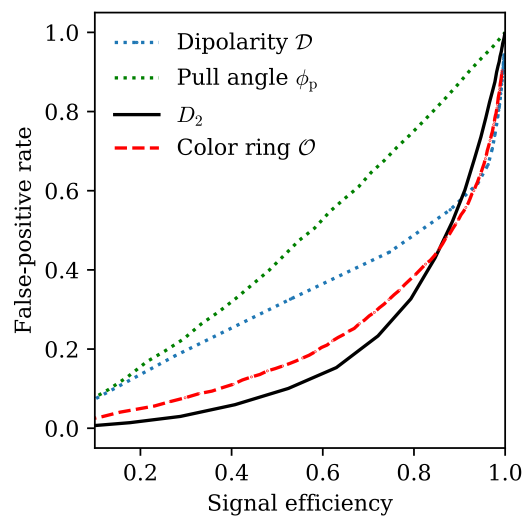
Discrimination works in simulation

1st dumb thing:

Jet has 3 particles,

minimal for color correlations





2nd dumb thing:

Neyman-Pearson:

Likelihood ratio is optimal

Neyman, Pearson 1933

$$\mathcal{O} = \frac{|\mathcal{M}_{\rm QCD}|^2}{|\mathcal{M}_{\rm singlet}|^2}$$

D_2 still apparently has some interesting color information

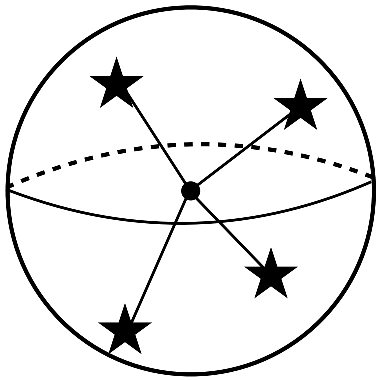
arXiv:1409.6298

arXiv:2006.10480

Example 2: The Topology of Phase Space

Problem:

What is the manifold on which particles in collision event live?

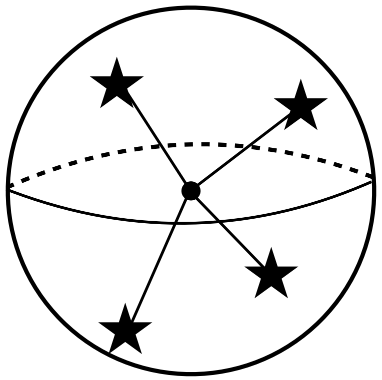


For simplicity, let's take all particles as massless

Example 2: The Topology of Phase Space

Problem:

What is the manifold on which particles in collision event live?



1st dumb thing:

Answer: Relativistic N-body phase space

Dirac 1927

$$d\Pi_N = \prod_{i=1}^{N} \left[d^4 p_i\right]_+ \delta^{(4)}\left(Q - \sum_{i=1}^{N} p_i\right)$$

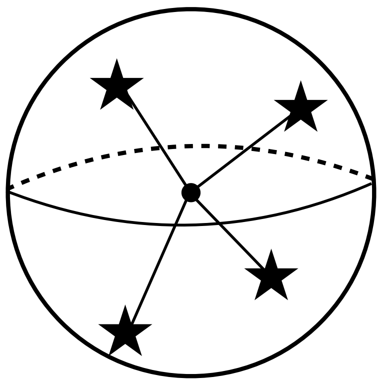
All particles are on-shell, have positive energy, and total momentum is conserved

arXiv:2008.06508

Example 2: The Topology of Phase Space

Problem:

What is the manifold on which particles in collision event live?



1st dumb thing:

Answer: Relativistic N-body phase space

Dirac 1927

$$d\Pi_N = \prod_{i=1}^{N} \left[d^4 p_i\right]_+ \, \delta^{(4)}\left(Q - \sum_{i=1}^{N} p_i\right)$$

2nd dumb thing:

Simplify and rewrite in a form that makes geometry manifest

$$\Pi_N \simeq \Delta_{N-1} \times S^{2N-3}$$

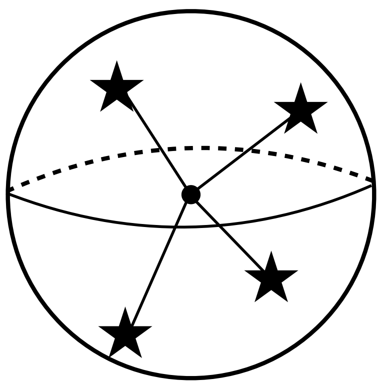
Isomorphic to the product of the *N*-1 simplex and the 2*N*-3 sphere

arXiv:2008.06508

Example 2: The Topology of Phase Space

Problem:

What is the manifold on which particles in collision event live?



Consequences:

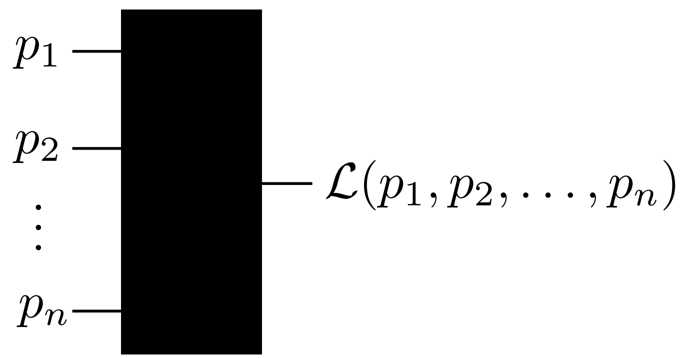
Phase space is **not** Euclidean space

1st dumb thing:

Answer: Relativistic N-body phase space

Dirac 1927

$$d\Pi_N = \prod_{i=1}^{N} \left[d^4 p_i\right]_+ \delta^{(4)}\left(Q - \sum_{i=1}^{N} p_i\right)$$



2nd dumb thing:

Simplify and rewrite in a form that makes geometry manifest

$$\Pi_N \simeq \Delta_{N-1} \times S^{2N-3}$$

ML techniques that assume flat space might fail on a curved manifold

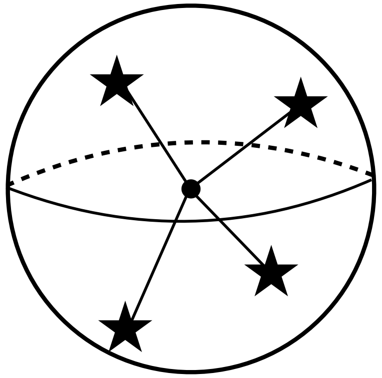
If the dumbest thing works, then it isn't so dumb, and you can completely understand everything

arXiv:2008.06508

Example 2: The Topology of Phase Space

Problem:

What is the manifold on which particles in collision event live?



Consequences:

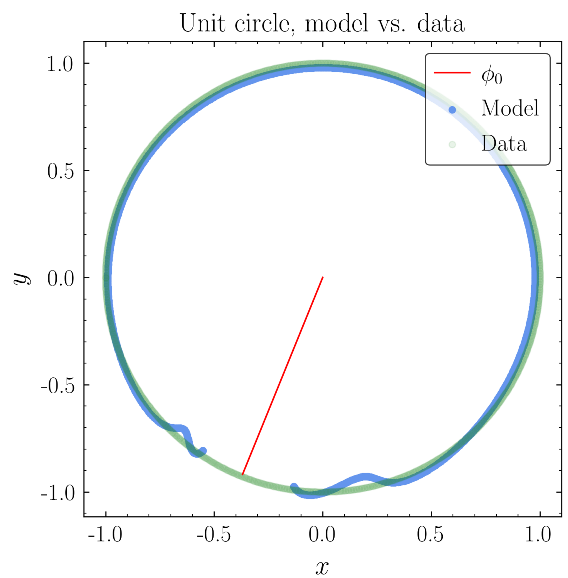
Problems with autoencoders on non-trivial manifolds

1st dumb thing:

Answer: Relativistic N-body phase space

Dirac 1927

$$d\Pi_N = \prod_{i=1}^{N} \left[d^4 p_i\right]_+ \delta^{(4)}\left(Q - \sum_{i=1}^{N} p_i\right)$$



2nd dumb thing:

Simplify and rewrite in a form that makes geometry manifest

$$\Pi_N \simeq \Delta_{N-1} \times S^{2N-3}$$

Textbook example of spontaneous symmetry breaking

Autoencoder "tears" manifold at random point

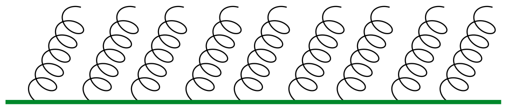
arXiv:2102.08380

If the dumbest thing works, then it isn't so dumb, and you can completely understand everything

arXiv:2008.06508

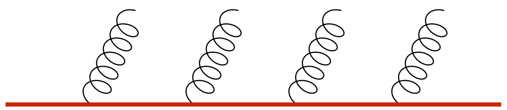
Example 3: The Likelihood for Quark vs. Gluon Discrimination

Problem:



gluon jet

vs.



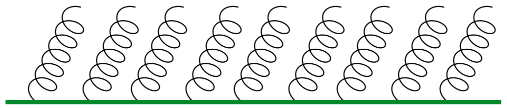
quark jet

Just assume that the only difference is the total color

arXiv:1906.01639, 2008.09673

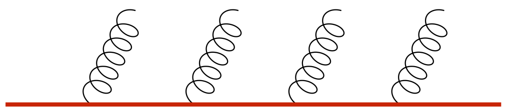
Example 3: The Likelihood for Quark vs. Gluon Discrimination

Problem:



gluon jet

vs.



quark jet

1st dumb thing:

Most naive results from stats/ML

Universal Approx. Theorem
Cybenko 1989, etc

$$\mathcal{L}(\{p_i\}) = \mathcal{L}(\{x_i\})$$

Neyman-Pearson Lemma
Neyman, Pearson 1933

$$\mathcal{L}(\{x_i\}) = \frac{p_g(\{x_i\})}{p_q(\{x_i\})}$$

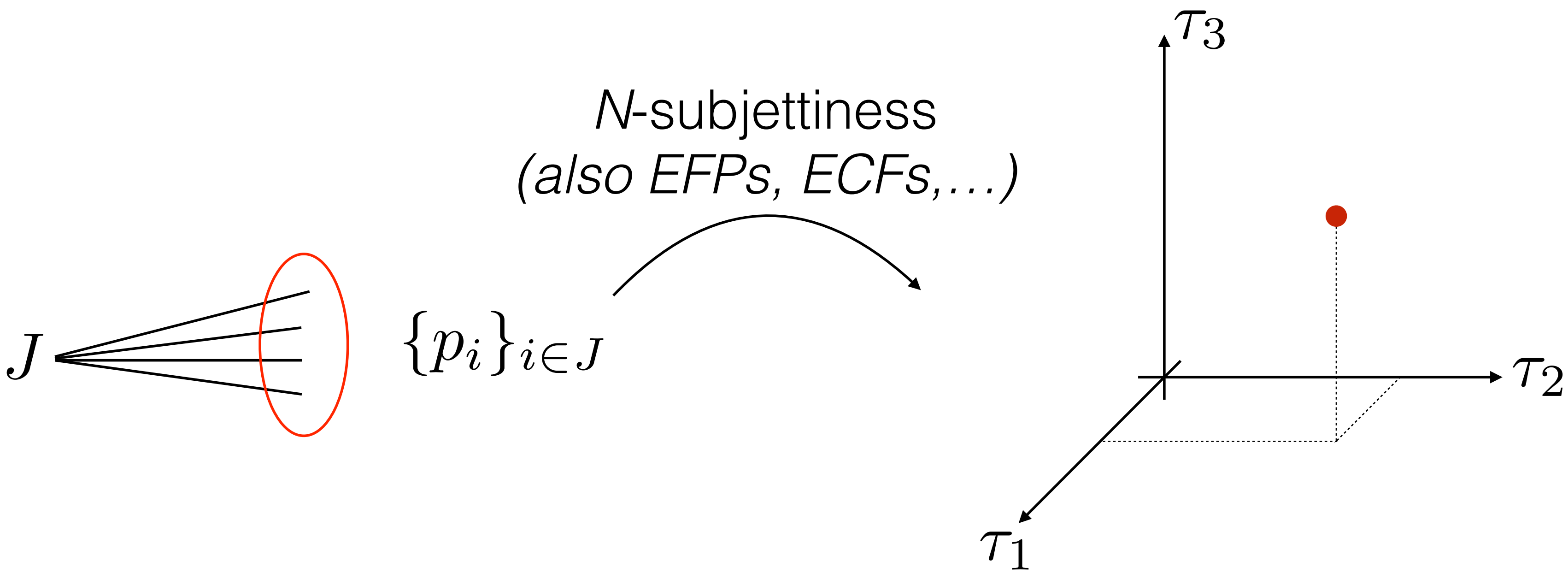
Choose a basis of IRC safe observables
to express particle momenta

If the dumbest thing works, then it isn't so dumb, and you can completely understand everything

arXiv:1906.01639, 2008.09673

Example 3: The Likelihood for Quark vs. Gluon Discrimination

N-subjettinesses and related observables accomplish this



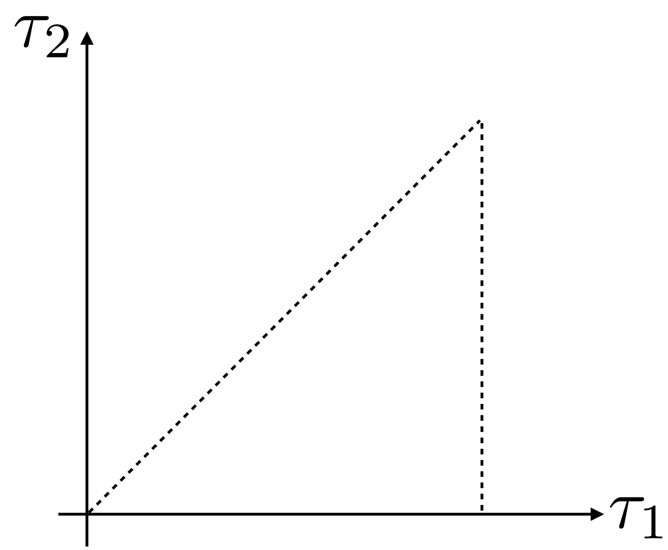
arXiv:1704.08249

history:
arXiv:1004.2489,1011.2268,1108.2701
Brandt, Dahmen 1979
Wu, Zobernig 1979
Nachtmann, Reiter 1982

$$\tau_N^{(\beta)} = \frac{1}{p_{TJ}} \sum_{i \in J} p_{Ti} \min \left\{ R_{1i}^{\beta}, R_{2i}^{\beta}, \ldots, R_{Ni}^{\beta} \right\}$$

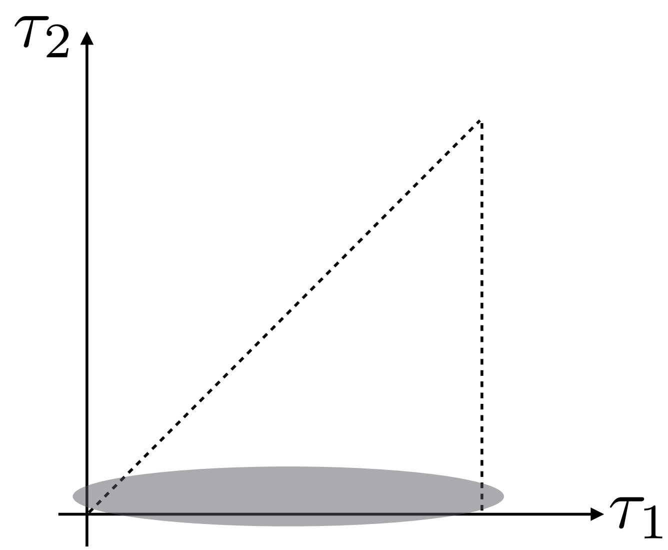
Example 3: The Likelihood for Quark vs. Gluon Discrimination

For visualization simplicity, just consider (τ_1,τ_2)



Example 3: The Likelihood for Quark vs. Gluon Discrimination

For visualization simplicity, just consider (τ_1,τ_2)



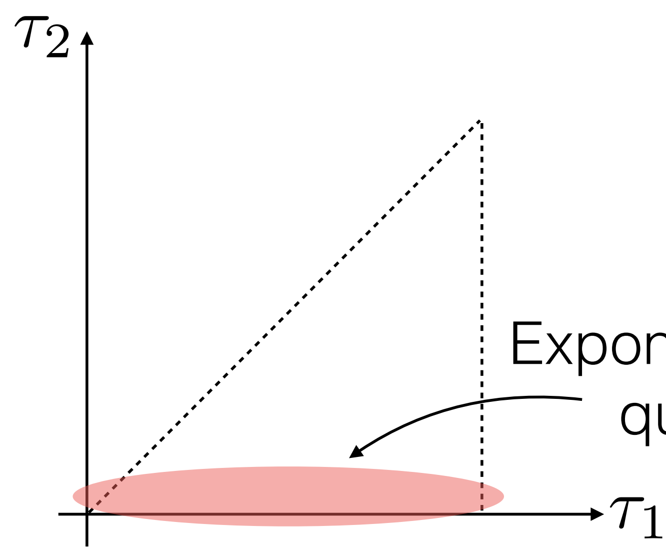
Particle production as Poisson process

IRC safety + additivity = exponential suppression

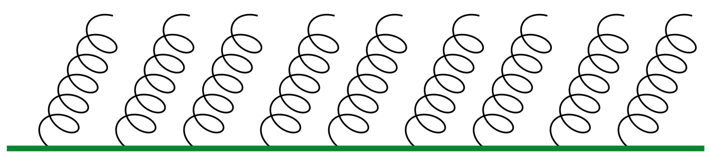
Exponentially small probability in regions where $\tau_N \to 0$

Example 3: The Likelihood for Quark vs. Gluon Discrimination

For visualization simplicity, just consider (τ_1, τ_2)

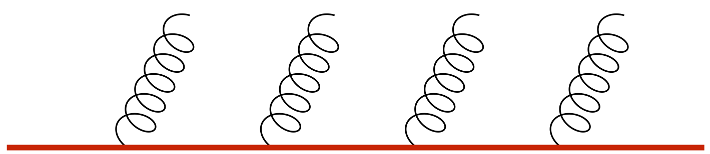


Exponentially more likely to be quark than gluon here



$\propto C_A$

gluon jet



$\propto C_F < C_A$

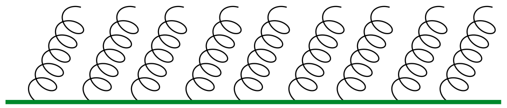
quark jet

Additivity then implies

$$\tau_N^{\text{gluon}} > \tau_N^{\text{quark}}$$

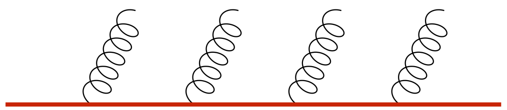
Example 3: The Likelihood for Quark vs. Gluon Discrimination

Problem:



gluon jet

vs.



quark jet

1st dumb thing:

Most naive results from stats/ML

Universal Approx. Theorem
Cybenko 1989, etc

$$\mathcal{L}(\{p_i\}) = \mathcal{L}(\{x_i\})$$

Neyman-Pearson Lemma
Neyman, Pearson 1933

$$\mathcal{L}(\{x_i\}) = \frac{p_g(\{x_i\})}{p_q(\{x_i\})}$$

2nd dumb thing:

IRC safe observables have a Sudakov form factor

$$p_g(\{x_i\}) \sim e^{-C_A f(\{x_i\})}$$

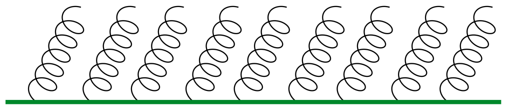
$$p_q(\{x_i\}) \sim e^{-C_F f(\{x_i\})}$$

Vanishes exponentially in the soft/collinear limits; $C_A > C_F$

arXiv:1906.01639, 2008.09673

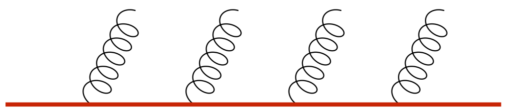
Example 3: The Likelihood for Quark vs. Gluon Discrimination

Problem:



gluon jet

vs.



quark jet

Consequences:

Likelihood is IRC safe

1st dumb thing:

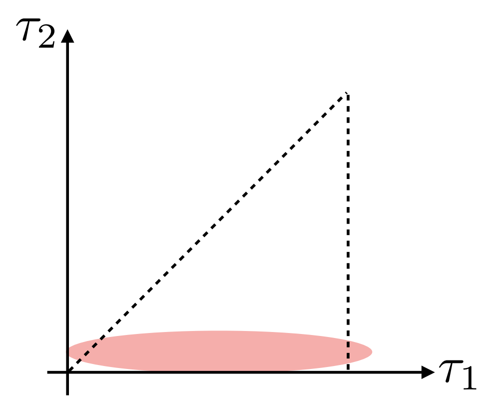
Most naive results from stats/ML

Universal Approx. Theorem
Cybenko 1989, etc

$$\mathcal{L}(\{p_i\}) = \mathcal{L}(\{x_i\})$$

Neyman-Pearson Lemma
Neyman, Pearson 1933

$$\mathcal{L}(\{x_i\}) = \frac{p_g(\{x_i\})}{p_q(\{x_i\})}$$



2nd dumb thing:

IRC safe observables have a Sudakov form factor

$$p_g(\{x_i\}) \sim e^{-C_A f(\{x_i\})}$$

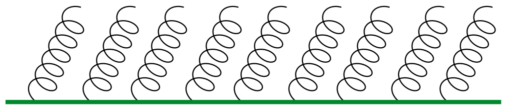
$$p_q(\{x_i\}) \sim e^{-C_F f(\{x_i\})}$$

Entire singular region is mapped to the unique value $L = 0$

arXiv:1906.01639, 2008.09673

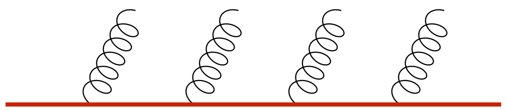
Example 3: The Likelihood for Quark vs. Gluon Discrimination

Problem:



gluon jet

vs.



quark jet

Consequences:

Explains why IRC safe observables are good q/g discriminants

1st dumb thing:

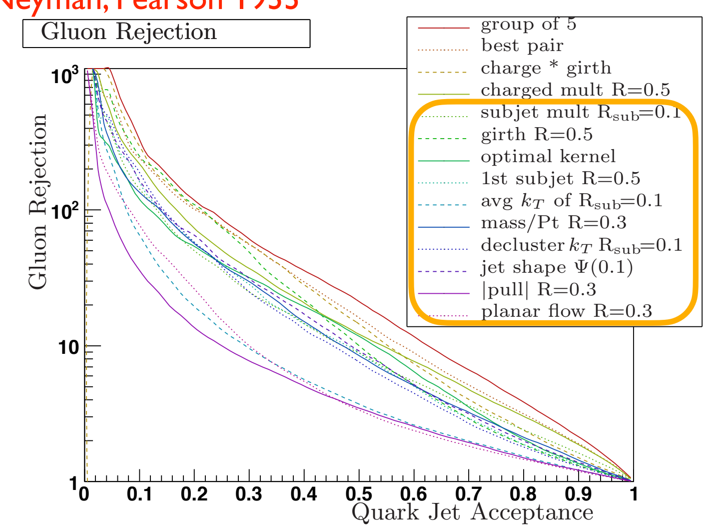
Most naive results from stats/ML

Universal Approx. Theorem
Cybenko 1989, etc

$$\mathcal{L}(\{p_i\}) = \mathcal{L}(\{x_i\})$$

Neyman-Pearson Lemma
Neyman, Pearson 1933

$$\mathcal{L}(\{x_i\}) = \frac{p_g(\{x_i\})}{p_q(\{x_i\})}$$



arXiv:1106.3076

2nd dumb thing:

IRC safe observables have a Sudakov form factor

$$p_g(\{x_i\}) \sim e^{-C_A f(\{x_i\})}$$

$$p_q(\{x_i\}) \sim e^{-C_F f(\{x_i\})}$$

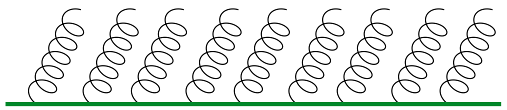
Simplify the work of an NN and give it IRC safe inputs

arXiv:1906.01639, 2008.09673

If the dumbest thing works, then it isn't so dumb, and you can completely understand everything

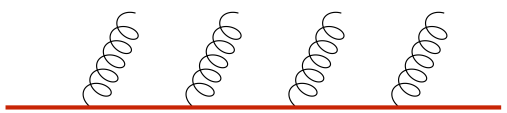
Example 3: The Likelihood for Quark vs. Gluon Discrimination

Problem:



gluon jet

vs.



quark jet

Consequences:

Explains why IRC safe observables are good q/g discriminants

1st dumb thing:

Most naive results from stats/ML

Universal Approx. Theorem
Cybenko 1989, etc

$$\mathcal{L}(\{p_i\}) = \mathcal{L}(\{x_i\})$$

Neyman-Pearson Lemma
Neyman, Pearson 1933

$$\mathcal{L}(\{x_i\}) = \frac{p_g(\{x_i\})}{p_q(\{x_i\})}$$

TABLE III. Comparison of the quark-gluon classification performance of EFN and PFN networks, via AUC, on jets with no hadronization effects included.

Jet p_T Range	EFN	PFN	Δ(PFN-EFN)
200-220 GeV	0.739 ± 0.001	0.737 ± 0.001	-0.002 ± 0.002
500-550 GeV	0.753 ± 0.001	0.750 ± 0.001	-0.003 ± 0.002
1000-1100 GeV	0.759 ± 0.001	0.758 ± 0.001	-0.001 ± 0.002

arXiv:2103.09103

2nd dumb thing:

IRC safe observables have a Sudakov form factor

$$p_g(\{x_i\}) \sim e^{-C_A f(\{x_i\})}$$

$$p_q(\{x_i\}) \sim e^{-C_F f(\{x_i\})}$$

IRC safe optimality validated in explicit NN implementation

If the dumbest thing works, then it isn't so dumb, and you can completely understand everything

arXiv:1906.01639, 2008.09673

It is our job as physicists to make sure we understand the problem well so that we can trust machine learning to output something sensible