Next generation kinematic "variables" for signal discovery and measurement

Prasanth Shyamsundar University of Florida



based on work with Prof. Konstantin Matchev Dr. Doojin Kim [arXiv:1906.02821 (JHEP), arXiv:1910:xxxxx]

Brookhaven National Lab

High energy theory seminar

Oct 16, 2019

Introduction



- Data from colliders is high-dimensional.
- Even at the parton level, there are often several final state particles (each with three momentum components).
- Inferences about the underlying (possibly new) physics has to be made from the distribution of this high-dimensional data.
- In this talk we will
 - 1. Identify the kinematic features in the distribution of events from some common event topologies with missing final state particles.
 - 2. Look at a new approach to seeing/exploiting these features (especially high-dimensional features).

Part I: What are the features in data?

Part 2: How do we see them? Part I: What are the features in data?

Part 2: How do we see them?

Kinematic constraints

What controls the distribution of momenta of parton level final state particles for a given diagram?



Kinematic constraints

What controls the distribution of momenta of parton level final state particles for a given diagram?

Kinematic constraints:

- Invariant mass constraints (resonances)
- 2a. Total transverse momentum constraint (hadron collider)
- 2b. Total 4-momentum constraint (lepton collider)

Everything else:

PDFs, Spin correlations... QFT, detector effects



Hypothesis for underlying event topology

Kinematic constraints

Given a sample of signal events from some diagram (possibly with invisible final state particles)

Assumptions

- Unknown intermediate particle masses
- Unknown invisible final state particle masses
- Perfect detectors
- No width effects. All particles on-shell
- No combinatorial ambiguities

Question: What features do kinematic constraints introduce?

The assumptions about the data are far from realistic. But invariant mass as an event variable is derived from this kind of consideration.



I. Constraints add features

We'll build intuition with cartoons, using two or three dimensional toy examples as proxy for our high dimensional data.



Each constraint reduces the dimensionality of the space our data lies on by 1.

I. Constraints add features

Examples



$$(p_1 + p_2)^2 = M^2$$

5-d feature in 6-d space

$$(p_1 + p_2)^2 = M_0^2$$

 $(p_1 + p_2 + p_2)^2 = M_1^2$

7-d feature in 9-d space

The shape of the allowed subspace or "feature" is parametrized by the unknown masses 2. Missing particles & dimentionality reduction introduce projections



- A particle being invisible means that the data available to us is a projection of the full phase-space onto the visible subspace.
- If we construct a low dimensional event variable (say, invariant mass and transverse momentum) and only analyze their distribution, we are looking at the projection of the visible data on that subspace.

3. Projections can kill features

To see a feature we need to be on a space at least one dimension higher Otherwise we'll say a feature is "lost"



- One of the goals of our work is to understand what features survive the projection caused by some final state particles being invisible.
- Another goal is to depict the surviving kinematic features by appropriate representation of data.

3a. Symmetries: Projections that preserve features 2-d feature onto 2-d 2-d feature onto 2-d



- If a projection or transformation preserves all the constraints, we can perform it without losing a feature.
- Example: $(x x_c)^2 + (y y_c)^2 = R^2$ is preserved by $z \to 0$
- Note: These transformations should be independent of the unknown masses/parameters

3B. Jacobian features

- When projecting an N-dimensional feature onto an N-dimensional subspace, the distribution of data has singularities.
- Singularities in visible subspace occur where the solutions for invisible momenta become degenerate (number of solutions changes).



Note: If we project an N dim feature onto < N dim subspace, the Jacobian feature doesn't show up.

3B. Jacobian features



Our papers so far focus exclusively on Jacobian features

3B. Jacobian features

Jacobian features don't have to be at the boundary



Physics examples

Let's put all these ideas together to analyze some common event topologies.



Physics examples

Blueprint:

- 1. Start with all momentum components visible and invisible, say N_{tot} .
- 2. Count constraints, say C in number. Find the dimensionality of the feature in the full (vis+invis) space, say D, as $N_{tot} C$.
- 3. Count number of visible momenta, say N_{vis} . If $N_{vis} > D$, we have a delta function feature. If $N_{vis} = D$, we have a Jacobian feature. (We will not deal with the $N_{vis} < D$ case in this talk)
- Look for symmetries to exploit. These will let us retain the feature in a low dimensional space, reducing N_{vis} and D equally.

Warmup 1: Fully visible decay topologies



Invariant mass variables capture all kinematic features in fully visible topologies

Warmup 2: Add an invisible particle (staying in low dimensions)



- This 1-d variable is p_T of the lepton, which is invariant under rotation about & boost along beam axis.
- *p_T* of the lepton exhibits a singularity at its endpoint.

To speed up our process, note that $D = N_{vis} \Leftrightarrow C = N_{invis}$.

Moving up in dimensions...



 $\vec{p}_{lT} + \vec{p}_{vT} + \vec{p}_{T}^{isr} = 0$

$$N_{tot} = 3 + 4 + 3 = 10$$

 $C = 2 + 2 = 4$
 $D = 10 - 4 = 6$
 $N_{vis} = 6$

Jacobian feature in 6-d vis space + Symmetries: 3 (rotation about beam axis + 2 independent boosts along beam axis) ↓ Jacobian feature in 3-d vis space

- > This special 3-d space should capture $\vec{p}_{l,T}$ and \vec{p}_T^{isr} (minus one rotation)
- We use $|\vec{p}_T^{isr}|$, $\vec{p}_{T\parallel}$ and $\vec{p}_{T\perp}$ (components of $\vec{p}_{l,T}$ along and perpendicular to \vec{p}_T^{isr}).

Does the Jacobian feature actually exist?

Yes it does!



Antler diagram in lepton collider (optional)



$$N_{tot} = 6 + 8 = 14$$

$$C = 4 + 4 = 8$$

$$D = 14 - 8 = 6$$

$$N_{vis} = 6$$

Jacobian feature in 6-d vis space + Symmetries: 3 (rotations) ↓ Jacobian feature in 3-d vis space

$$p_{\nu i}^{2} = M_{\nu}^{2}$$

$$(p_{\nu i} + p_{li})^{2} = M_{W}^{2}$$

$$\vec{p}_{l1} + \vec{p}_{\nu 1} + \vec{p}_{l2} + \vec{p}_{\nu 2} = 0$$

$$E_{l1} + E_{\nu 1} + E_{l2} + E_{\nu 2} = E_{\text{collider}}$$

Antler diagram in lepton collider (optional)



tt-like topology



$$N_{tot} = 12 + 8 = 20$$

$$C = 6 + 2 = 8$$

$$D = 20 - 8 = 12$$

$$N_{vis} = 12$$

Jacobian feature in 12-d vis space + Symmetries: 3 (1 rotation, 2 boosts) ↓ Jacobian feature in 9-d vis space

How do we see/exploit this feature? (part 2)

Summary for part 1

- We have a constraint counting prescription that characterizes the kind of kinematic feature that exists in the distribution of (visible) data.
- ► If $C > N_{invis}$, delta function feature If $C = N_{invis}$, Jacobian feature

If $C < N_{invis}$, neither of these (future publication)

- Based on symmetries, we have a prescription to find the least dimensionality of data that can retain these kinematic features.
- In some examples, that least dimensionality is rather high. (9 for *tt*-like topology).

Part I: What are the features in data?

Part 2: How do we see them? Part I: What are the features in data?

Part 2: How do we see them?

The problem

- We have 9 dim data with a Jacobian feature in it, i.e., an 8 dimensional hypersurface where the density of signal events will be enhanced.
- If we reduce the dimensionality of data, we lose this feature.

Why not analyze the distribution of full 9 dim data, say, with 9-d histograms? Curse of dimensionality:

- Computing power needed to scan the full phase space grows exponentially with number of dimensions.
- Amount of data (real and MC) needed to populate this space grows exponentially.
- MC validation in full space could be a problem.

How does a computer see a circle?

Again, I'll use a toy example to illustrate our method.

- Let's say we have a picture with some points (x_i, y_i).
 We want to know if there is a circular signal in the picture.
- If we are guaranteed that "if a circular signal exists, it will be centered at the origin..."

Histogram the distance of all points from the origin. If there is a circle, there'll be a peak in the histogram at the correct radius.

What if we don't know the x-coordinate of the center...

- Assume that each point came from a circle. Let each point vote for the parameters of the circle it could be a part of.
- If there is indeed a circle, the corresponding parameters will receive a lot of votes. Look for peaks.



- Assume that each point came from a circle. Let each point vote for the parameters of the circle it could be a part of.
- If there is indeed a circle, the corresponding parameters will receive a lot of votes. Look for peaks.



- Assume that each point came from a circle. Let each point vote for the parameters of the circle it could be a part of.
- If there is indeed a circle, the corresponding parameters will receive a lot of votes. Look for peaks.



- Assume that each point came from a circle. Let each point vote for the parameters of the circle it could be a part of.
- If there is indeed a circle, the corresponding parameters will receive a lot of votes. Look for peaks.



- Assume that each point came from a circle. Let each point vote for the parameters of the circle it could be a part of.
- If there is indeed a circle, the corresponding parameters will receive a lot of votes. Look for peaks.



- Assume that each point came from a circle. Let each point vote for the parameters of the circle it could be a part of.
- If there is indeed a circle, the corresponding parameters will receive a lot of votes. Look for peaks.



- Assume that each point came from a circle. Let each point vote for the parameters of the circle it could be a part of.
- If there is indeed a circle, the corresponding parameters will receive a lot of votes. Look for peaks.



- Assume that each point came from a circle. Let each point vote for the parameters of the circle it could be a part of.
- If there is indeed a circle, the corresponding parameters will receive a lot of votes. Look for peaks.



- Assume that each point came from a circle. Let each point vote for the parameters of the circle it could be a part of.
- If there is indeed a circle, the corresponding parameters will receive a lot of votes. Look for peaks.



- Assume that each point came from a circle. Let each point vote for the parameters of the circle it could be a part of.
- If there is indeed a circle, the corresponding parameters will receive a lot of votes. Look for peaks.





"The unifying feature of points in a circle (red) is the fact that they lie on the same circle"

Hough transform

- What if we didn't know the y-coordinate either? Now each point/event will vote for a 2d surface of parameters in the (x_c, y_c, r) space.
- What is the advantage of mapping a point in a 2d space to a 2d surface in 3d space?
- Now the computer doesn't have to "see a circle" in the picture. It just needs to look for a peak.

Port the same idea to physics processes.

- Delta features: Assume that each event came from a given diagram. Vote for all possible mass parameters that allow this.
- Jacobian features: Assume that every event is an extreme event "extreme event" (degenerate solutions for invisible momenta).

Vote for all possible mass parameters that support this assumption.

We call this the focus points method.



Assuming symmetric decay, we have 3 mass parameters (700 GeV, 800 GeV, 1000 GeV)

1e3



Assuming symmetric decay, we have 3 mass parameters (700 GeV, 800 GeV, 1000 GeV)

1e3



Assuming symmetric decay, we have 3 mass parameters (700 GeV, 800 GeV, 1000 GeV)

1e3



Assuming symmetric decay, we have 3 mass parameters (700 GeV, 800 GeV, 1000 GeV)

1e3



Assuming symmetric decay, we have 3 mass parameters (700 GeV, 800 GeV, 1000 GeV)

1e3

 $\tilde{m}_W(GeV)$

tt-like It works!



Assuming symmetric decay, we have 3 mass parameters (700 GeV, 800 GeV, 1000 GeV)

1e3

 $\tilde{m}_W(GeV)$

tt-like It works!



Assuming symmetric decay, we have 3 mass parameters (700 GeV, 800 GeV, 1000 GeV)

1e3

 $\tilde{m}_W(GeV)$

 $t\bar{t}$ -like It works!

arXiv:1906.02821



Assuming symmetric decay, we have 3 mass parameters (700 GeV, 800 GeV, 1000 GeV)



66% of signal events pass through a 10 GeV × 10 GeV square near the true parameters!

ttbar 3d heatmaps



Brings the possibility of bump hunt to a diagram with invisible final state particles!

Where does the power come from?

Q: How did we retain the feature after reducing dimensionality?

A: We did not reduce the dimensionality! These curves are parametrized by 9 components of the visible data (out of 12).

Q: If we did not reduce the dimensionality, then how did we overcome the curse of dimensionality?

A: We are not looking for features. We are looking for metaphorical circles and are blind to metaphorical squares.

Each diagram will have a different Hough transform.

Other examples...

What we have is a generic way for capturing delta and Jacobian features in data. There's no reason why it should be used only for high dimensional features...

Let's use it for a fully visible 2 body decay...



- Assume the particles came from this diagram. Vote for the parent mass that could've produced it.
- This is precisely the invariant mass!

Other examples...



- Assume that the lepton came from this diagram. Vote for the masses of W and v for which this event would be extreme.
- This is precisely the transverse mass variable M_T!

 M_T is an implicit variable. It depends on M_{ν} .



The focus points technique is a generalization of invariant mass and transverse mass variables.

Other examples...



- Assume that the lepton came from this diagram. Vote for the masses of W and v for which this event would be extreme.
- This is precisely the transverse mass variable M_T!

 M_T is an implicit variable. It depends on M_{ν} .



The focus points technique is a generalization of invariant mass and transverse mass variables.

Practical effectiveness...

Q: Once we include detector effects and combinatorial background, the plots will not be as impressive? Will it still be worth doing all this in real life.





Practical effectiveness...

Q: Once we include detector effects and combinatorial background, the plots will not be as impressive? Will it still be worth doing all this in real life.

Instead of punching a strawman here...

A1: The precedent of success...

The distributions of invariant mass and M_T also become less impressive after including combinatorics and detector resolution. Yet our choice is clear between using them and not using them.

Practical effectiveness...

Q: Once we include detector effects and combinatorial background, the plots will not be as impressive? Will it still be worth doing all this in real life.

A2:

- Empirically, chopping up the phase-space into several categories leads to significant improvements in sensitivity (even in the presence of detector effects).
- This is a sign that there are features in the data in high dimensions completely missed after some projections.

Our idea tries to extract sensitivity by tapping into the kinematics (key) aspect of those features.
 (So far our only portals to those features have been event categorization and machine learning)

Summary and perspective...

- We have a novel technique to capture kinematic features that simply cannot be retained in traditional (low-dim) variables.
- Our "variables" are curves in appropriate spaces. The technique works for a number of diagrams.
- So far our only portals to this high dimensional information has been machine learning techniques, or chopping up the phase-space into 10s or 100s of signal regions. These suffer from poor interpretability.
- There's no mistaking the meaning of a peak on a focus points plot.

Future:

- There's more to the Hough transform method than a peak in the density of curves.
- Embracing them as representations of high dimensional data opens up a world of possibilities for HE data analysis.
- Happy to discuss these after the talk.

Summary and perspective...

- We have a novel technique to capture kinematic features that simply cannot be retained in traditional (low-dim) variables.
- Our "variables" are curves in appropriate spaces. The technique works for a number of diagrams.
- So far our only portals to this high dimensional information has been machine learning techniques, or chopping up the phase-space into 10s or 100s of signal regions. These suffer from poor interpretability.
- There's no mistaking the meaning of a peak on a focus points plot.

Future:

- There's more to the Hough transform method than a peak in the density of curves.
- Embracing them as representations of high dimensional data opens up a world of possibilities for HE data analysis.
- Happy to discuss these after the talk.

Thank you!