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 (and upcoming)

LPC Physics Forum
Fermilab
Aug 15, 2019

Before we begin...

- This talk is about constructing variables to capture kinematic features in the distribution of data (process dependent, model independent).
- Where would we expect more sensitivity (to presence of signal or value of parameter) to come from?
  1. Higher luminosity, higher energy, other experiments...
  2. Machine learning–in jet clustering and tagging, for example
  3. Better event selection/categorization (possibly using machine learning)
  4. Other novel analysis techniques.

Before we begin...

- This talk is about constructing variables to capture kinematic features in the distribution of data (process dependent, model independent).
- Where would we expect more sensitivity (to presence of signal or value of parameter) to come from?
  1. Higher luminosity, higher energy, other experiments...
  2. Machine learning–in jet clustering and tagging, for example
  3. Better event selection/categorization (possibly using machine learning)
  4. Other novel analysis techniques.
- Not kinematics... (kinematics may help identify some interesting phase space region, or design relevant trigger criteria. But not new kinematic variables) Mainly because we've had a lot of time to figure out the kinematics. What more could be left?

Before we begin...

- This talk is about constructing variables to capture kinematic features in the distribution of data (process dependent, model independent).
- Where would we expect more sensitivity (to presence of signal or value of parameter) to come from?
  1. Higher luminosity, higher energy, other experiments...
  2. Machine learning–in jet clustering and tagging, for example
  3. Better event selection/categorization (possibly using machine learning)
  4. Other novel analysis techniques.
- Not kinematics... (kinematics may help identify some interesting phase space region, or design relevant trigger criteria. But not new kinematic variables) Mainly because we've had a lot of time to figure out the kinematics. What more could be left?
- Goal: Try and convince you that what I've been working on isn't a complete waste of time (especially considering the click-baity title)

Introduction

- Collider data is high dimensional–even at the parton level
- We would love to analyze the distribution of events in the full phase space, but...
- Curse of dimensionality:
  1. The computing power needed to scan the full phase space grows exponentially with number of dimensions
  2. The amount of data (real and MC) needed to populate this space grows exponentially
  3. MC validation in full space could be a problem
- One way to deal with the curse of dimensionality is to reduce it–construct event variables like invariant masses of sets of particles, pseudorapidity, p_T, etc.

Introduction

- When reducing dimensionality, different regions of phase space get mixed.
- Loss of information ↔ loss of sensitivity (to presence of signal or value of parameter)
- Many of our fancy analysis techniques can be thought of as attempts to mitigate this loss of information
  - Event selection and categorization
  - Weighted histograms
  - Analyzing multiple 1D distributions simultaneously
  - MEKD, MELA
- These could make plots harder to interpret–weighted histograms, or 10s of categories of signal regions, for example (but that's okay)
- The more information we can retain with our event variables (in an interpretable way), the less we have to recover. This is the motivation for designing good kinematic variables.

Sidenote

- Dimensionality reduction typically cascades from
  detector level information → particle level information
  → parton level information → event variables
- The rationale is that the different stages of the evolution of an event are (approximately) decoupled. The parton level theory talks to the detector through parton showers **followed by** hadronization **followed by** interaction with detectors. So, we can backtrack in our analysis.
- Efforts are being made to minimize the information loss at every step (for example jet clustering and tagging using detector level information).
- Our work focuses on the "parton level information → event variables" part.

Kinematic constraints

- To leading order what controls the distribution of momenta of parton level final state particles (for a given diagram)?

1. Invariant mass constraints (resonance)

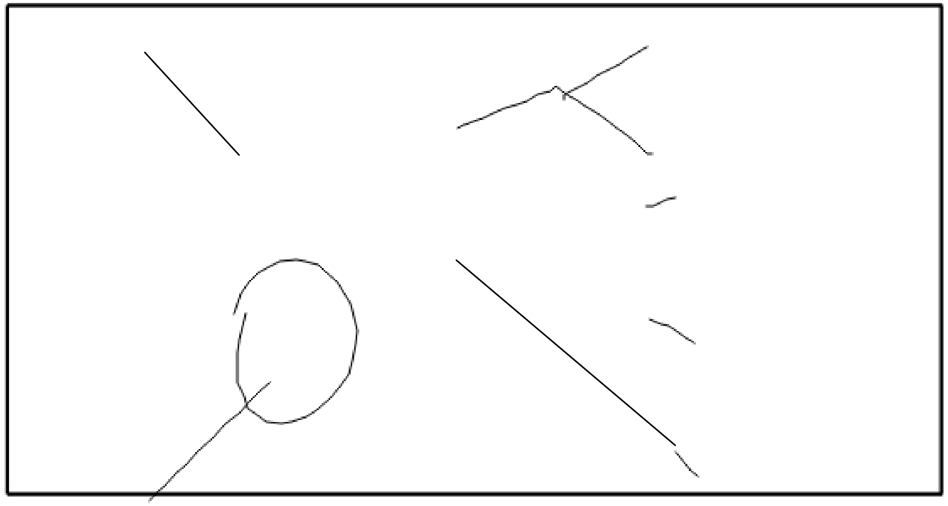
2a. Total transverse momentum (hadron collider)

2b. Total 4-momentum (lepton collider)

3. PDFs

4. Spin correlations

5. Every thing not mentioned here

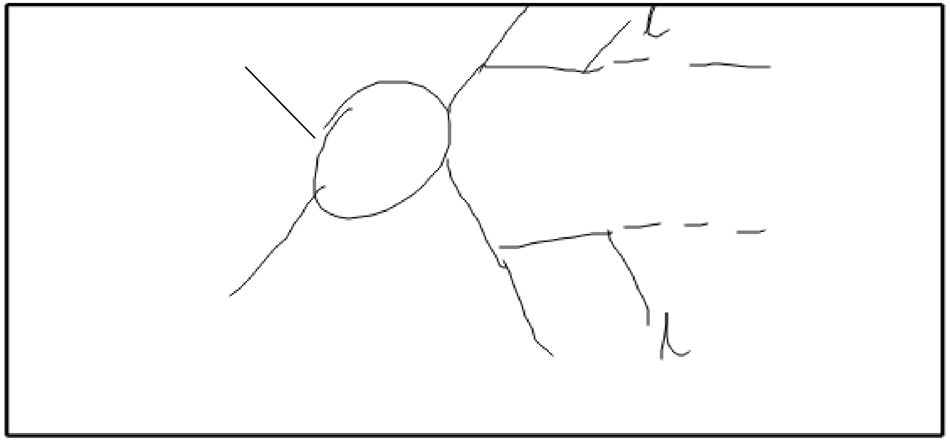


- We'll work with 1 and 2. They give (approximate) equality constraints on the final state momenta.
- PDFs can be incorporated as equality constraints in the high mass regions. High mass $\Rightarrow$ approximately produced at threshold (not relevant for this talk).

The questions

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

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



Questions: What are the features in the distribution of (visible) data? How do we capture them?

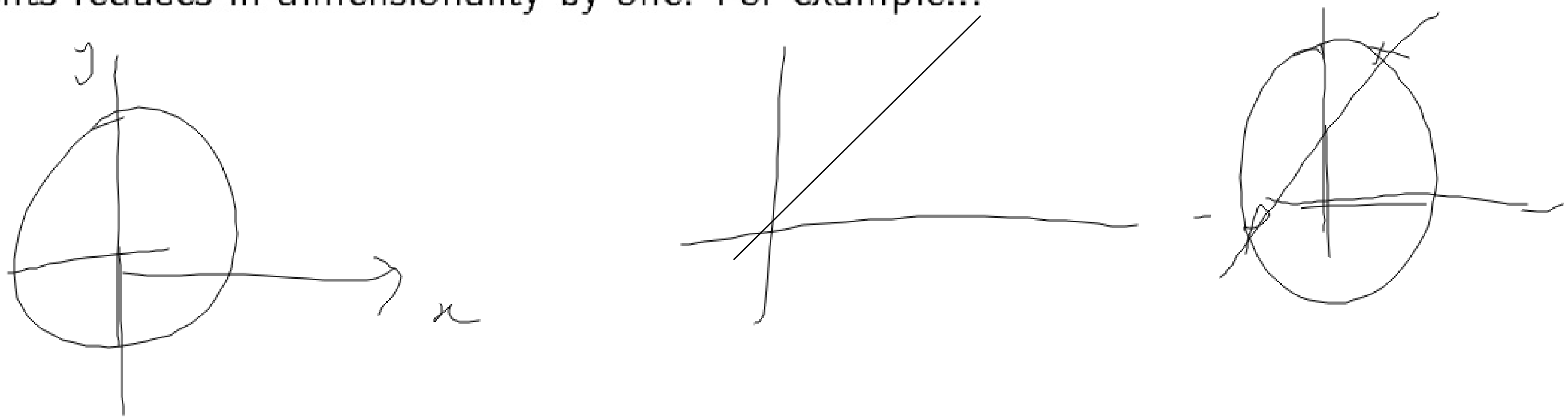
The assumptions about the data are far from realistic.
But invariant mass as an event variable is derived from this kind of consideration.

Constraints reduce allowed phase space

Let's consider all final state particle momenta (visible and invisible), say of dimensionality N.

When unconstrained, events can fully populate in this N dimensional space.

For each independent constraint (say C total) on momenta, the allowed phase space of events reduces in dimensionality by one. For example...



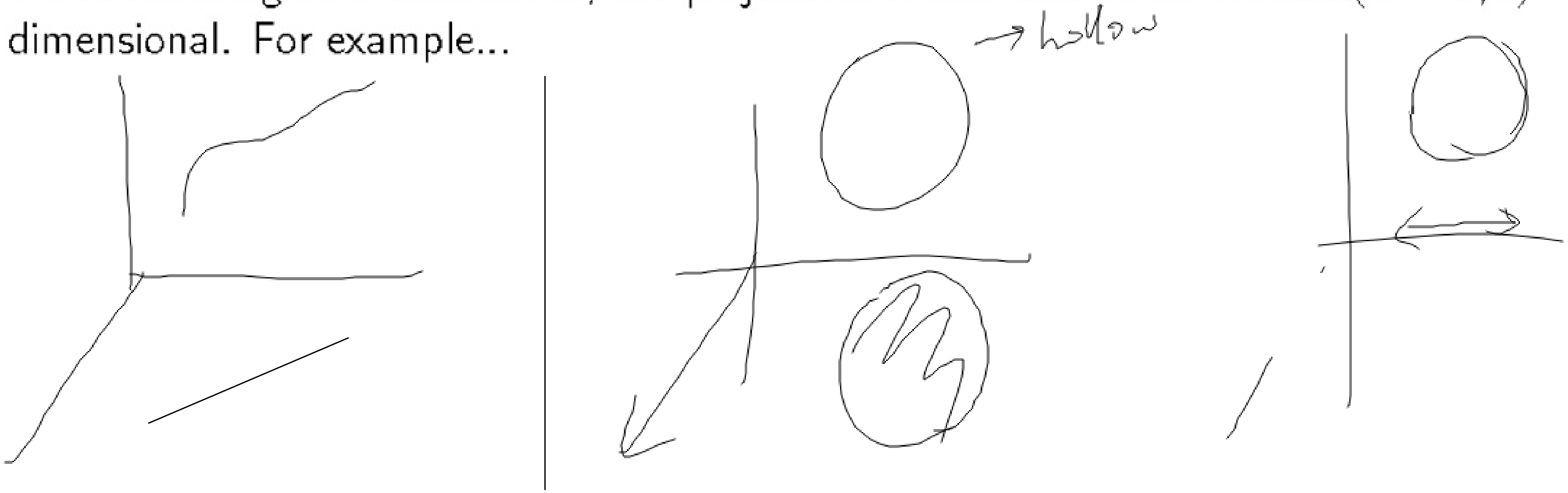
We'll call this an $N - C$ dimensional feature in an N dimensional space. The distribution of events on the allowed phase space is expected to be smooth.

Dimensionality reduction ↔ projection

Dimensionality reduction can be thought of as projecting the N dimensional space onto a lower dimensional subspace (after appropriate reparametrization).

The $N - C$ dimensional feature will also get projected.

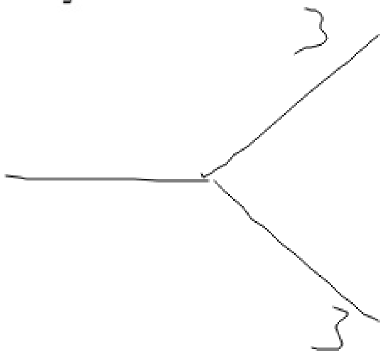
When reducing to n dimensions, the projection of the feature will be $\min(N - C, n)$ dimensional. For example...



To see a feature, we need need to live on a space at least one dimension higher.

An example

Let's try this out for a simple example (fully visible 2-body decay)



N = 6

C = 1

5

An example (oops!)

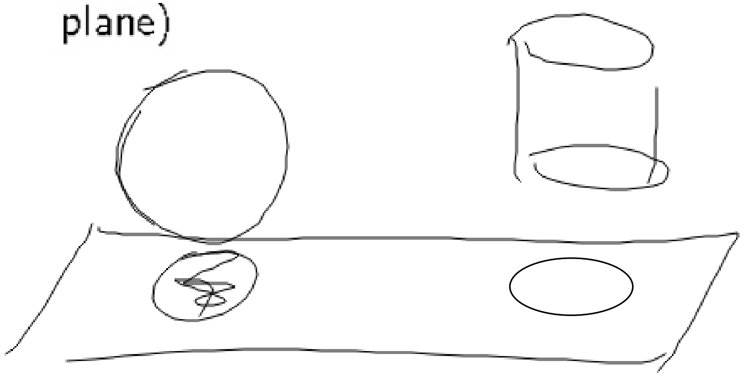
Let's try this out for a simple example (fully visible 2-body decay)

The 5d feature can only be seen from 6 dimensions(?)

Clearly this is wrong! We can construct the invariant mass of the two particles, a 1d variable, and see the feature.

Symmetries

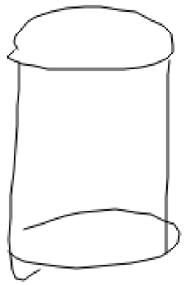
hollow sphere vs hollow cylinder (onto a plane)



Sphere (known vs unknown center)

Symmetries

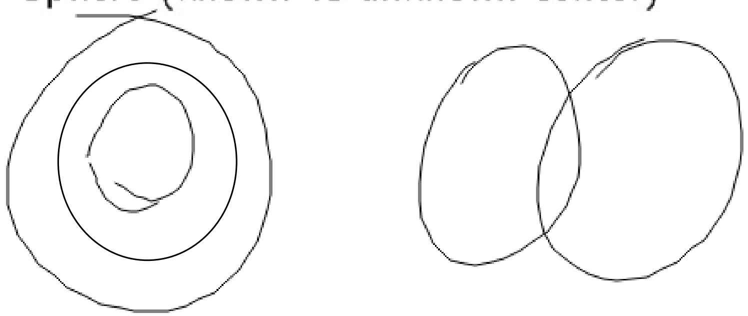
hollow sphere vs hollow cylinder (onto a plane)



Some projections can be done without losing constraints. We'll call these symmetries.

They correspond to transformations of **visible** momentum components which keep events on the allowed phase space.

Sphere (known vs unknown center)

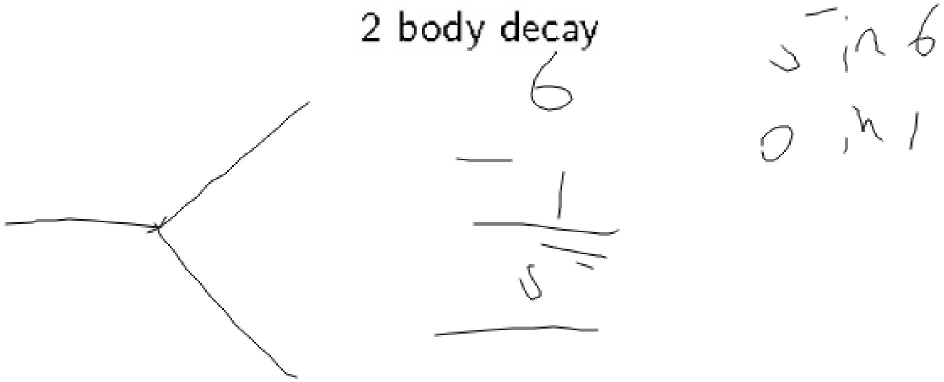


The transformations must be **independent of unknown parameters/masses.**

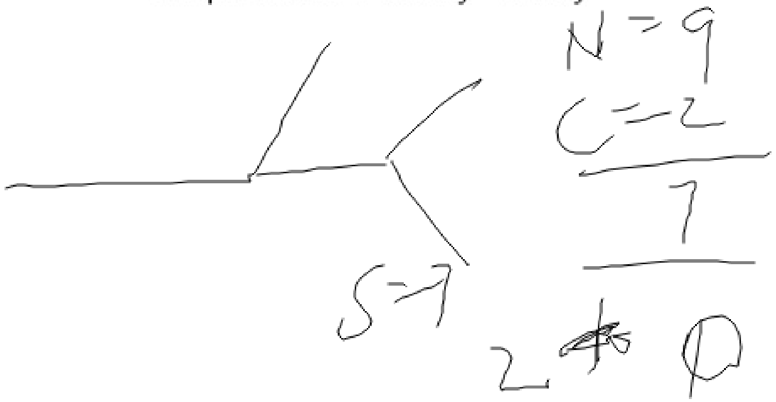
Symmetries bring down the dimensionality of the "full space" as well as the "feature".

Back to the fully visible example

2 body decay



sequential 3 body decay



High dimensions is not an issue to capture kinematic features (considered here) for **fully visible events.**

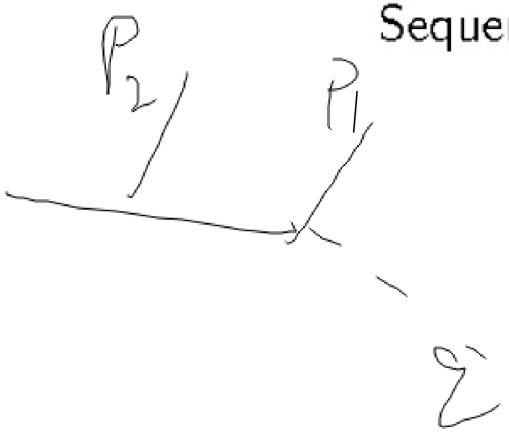
Stated differently: The information we stand to gain by working in the full phase-space is not kinematics (considered here) related.

Sidenote: In fully visible events, transverse momentum constraint is useless since both signal and background events satisfy it.

Invisible particles

- Having invisible particles in the final state can be thought of us nature doing some dimensionality reduction or projections for us.
- But not in a good or smart way
- Let's use p-s for visible momentum components and q-s for invisible.

Sequential 3 body decay (two step) in hadron collider



Only mass constraints

$N=10$ $n=6$

$C=3$

7

Mass and transverse momentum constraints

$N=10$ $n=6$

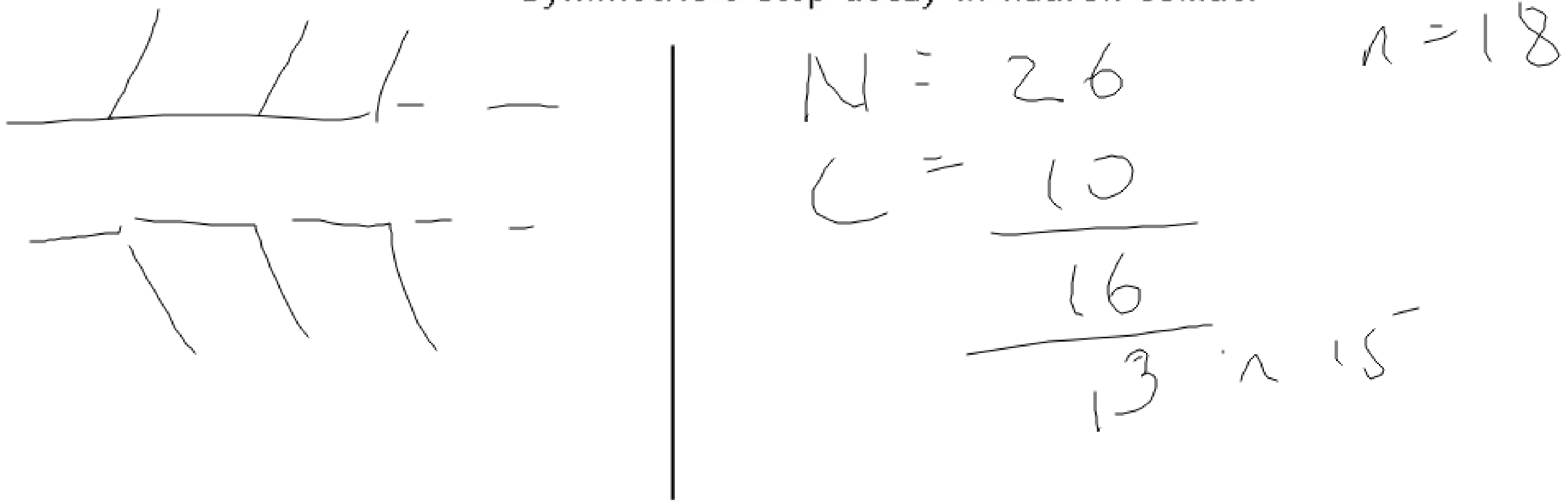
$C=5$

5

3 in 4

Another example

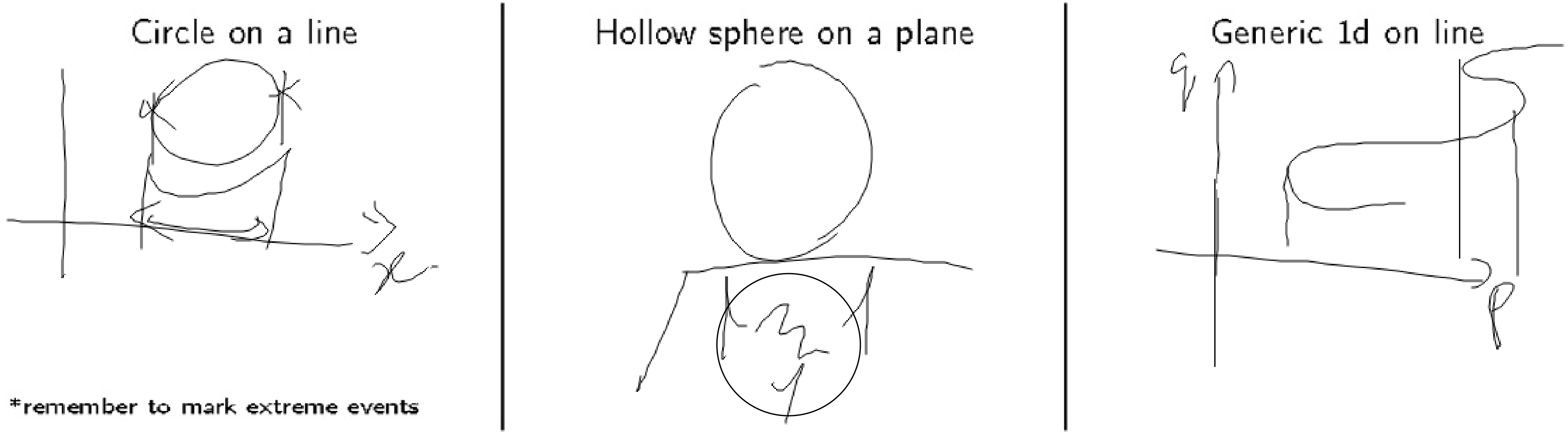
Symmetric 3 step decay in hadron collider



- Note that to get these features we need more constraints than unknown momenta. (The excess constraints can be thought of as providing constraints on the masses.)
- We've seen a 3d feature in a 4d space, and a 13d feature in a 15d space.
- There is no way of reducing the dimensionality without killing the feature (or assuming the masses).
- How to exploit these features? Let us put a pin on that for now.

Jacobian peaks - enhancement at degenerate points

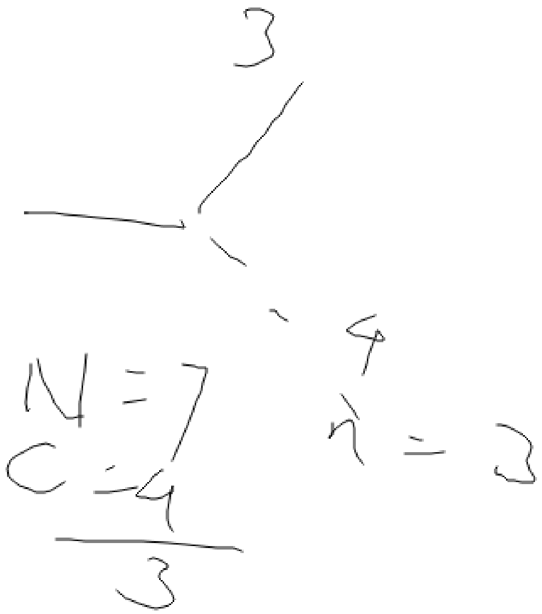
I mentioned that we lose the feature completely if we don't live in a dimension **higher** than the feature. This is not entirely true.



- Only works when feature dim. = visible subspace dim.
  Or #constraints = #invisible. The mapping is many to one, but discrete.
- Jacobian factor when transforming from full allowed phase-space to visible sub-space.

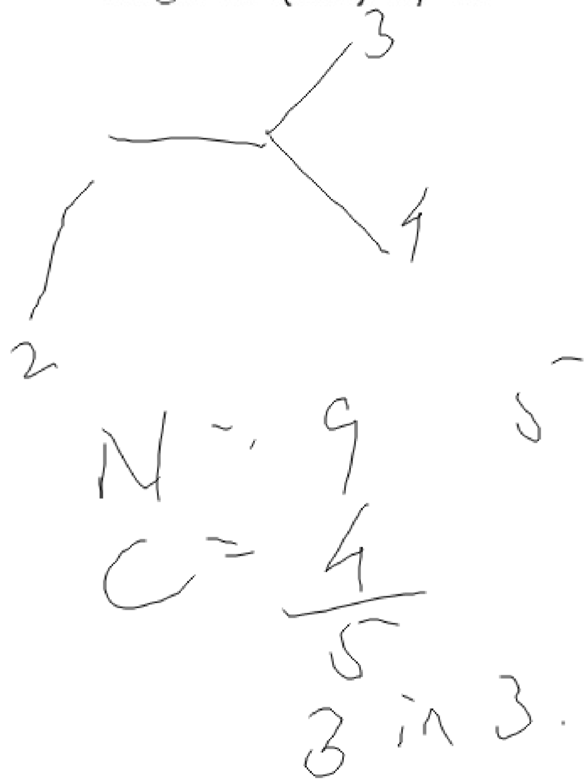
Jacobian peaks - physics examples

single W (like) w/o isr

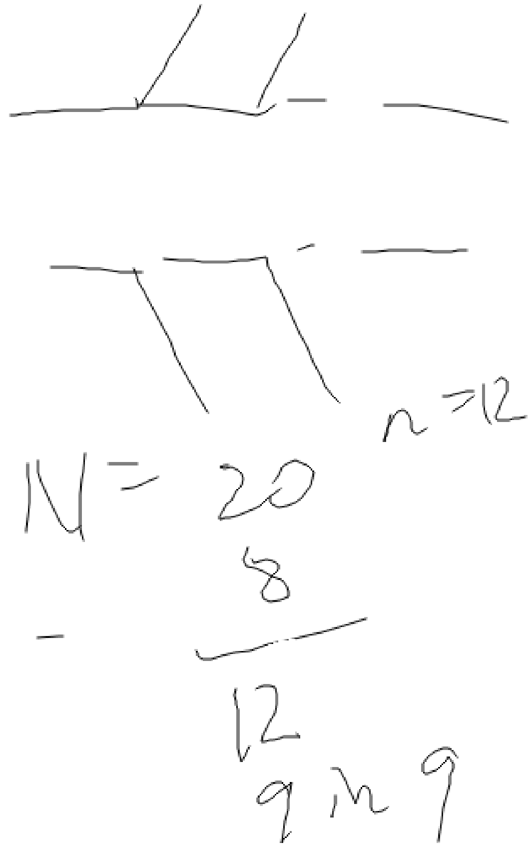


(can also be thought of as projection of $\vec{p}^{\,l}_{CM}$ onto $\vec{p}^{\,l}_{T,CM}$)

single W (like) w/ isr



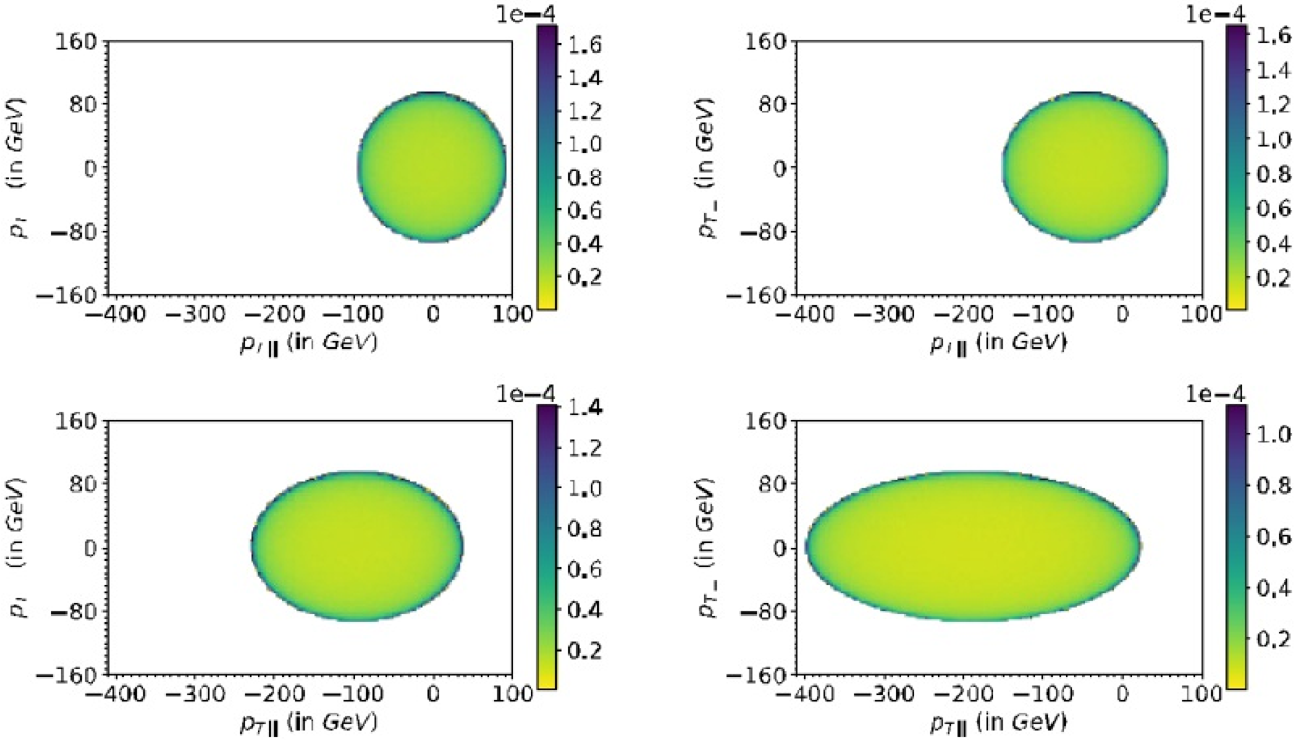
leptonic ttbar (like)

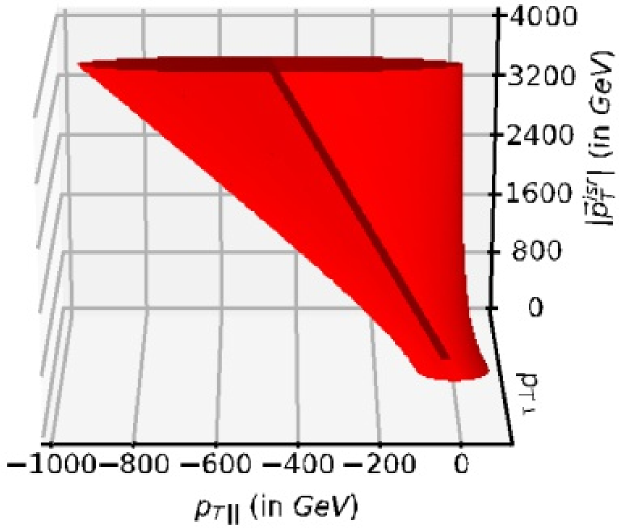


Jacobian peak - in pictures

<u>single W (like) w/ isr</u>

The three dimensional space which shows the Jacobian peak is $(p^{l}_{T\parallel}\ ,\ p^{l}_{T\perp}\ ,\ p^{isr}_{T})$

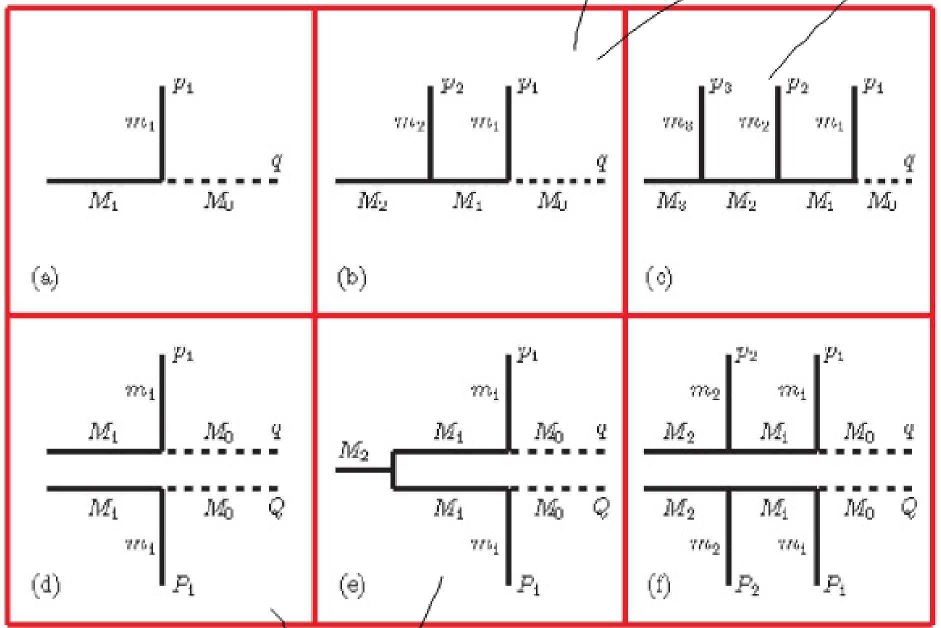




Can only be seen in 3d
(What about m_T?)

Jacobian peak

Jacobian peak features show up in a few other diagrams too:



Quick summary

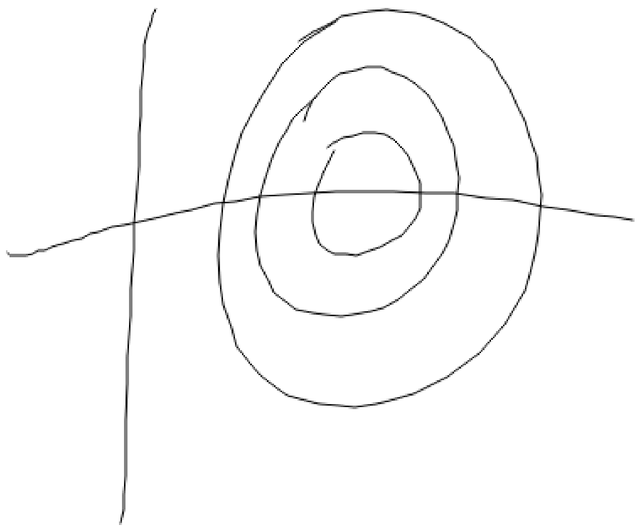
- Depending on the number of constraints and number of invisible momenta, there may be a delta function feature (let's call them that), or a Jacobian feature in the visible phase space.
- These features lie in high dimensions cannot be retained in reduced dimensions (no more lies!)

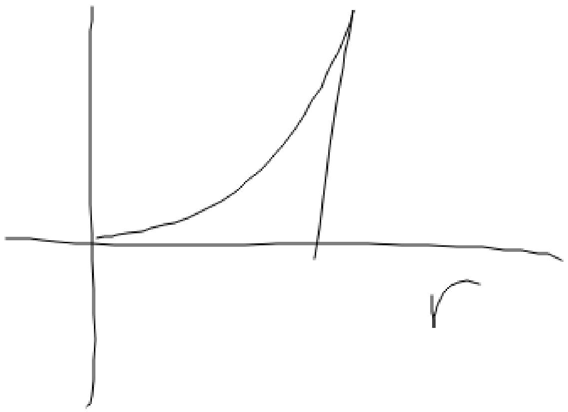
How to capture/exploit these features?

- One possibility is to construct an implicit variable $f(\vec{p}; \vec{m}_{\text{unknown}})$ that captures how far the event $\vec{p}$ is from the feature for a given set of mass values $\vec{m}_{\text{unknown}}$.
- Then we can study the distribution of $f(\vec{p})$ at different test masses. These are called singularity variables (we also do this in an upcoming paper).
- This is not what I'll talk about today...

How does a computer see a circle?

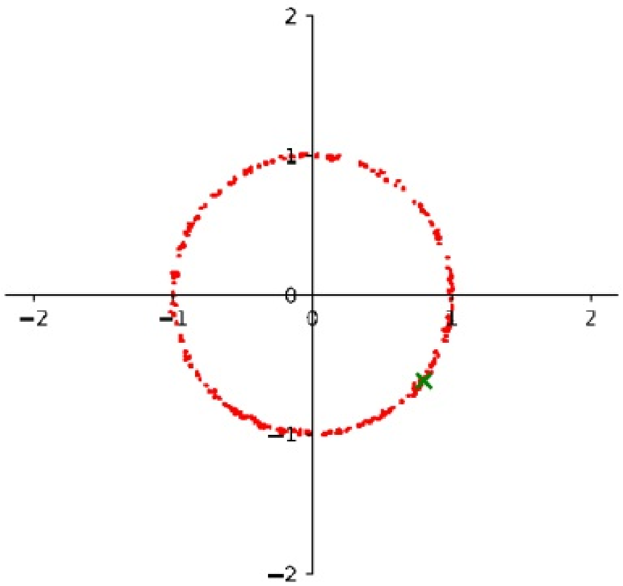
- Let's say we have a picture with some points (x_i, y_i). There may or may not be a circular signal in the picture. We want to know if there is
- If we know where the center of the (potential) circle will be (but not the radius) we can histogram the distance of all points from the center.
  If there is a circle, there'll be a peak in the histogram.
- What if we don't know the x-coordinate of the center...

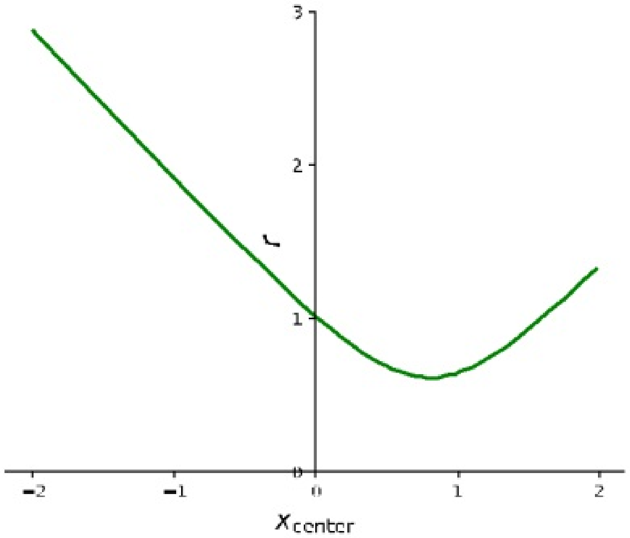




Hough transform for a circle

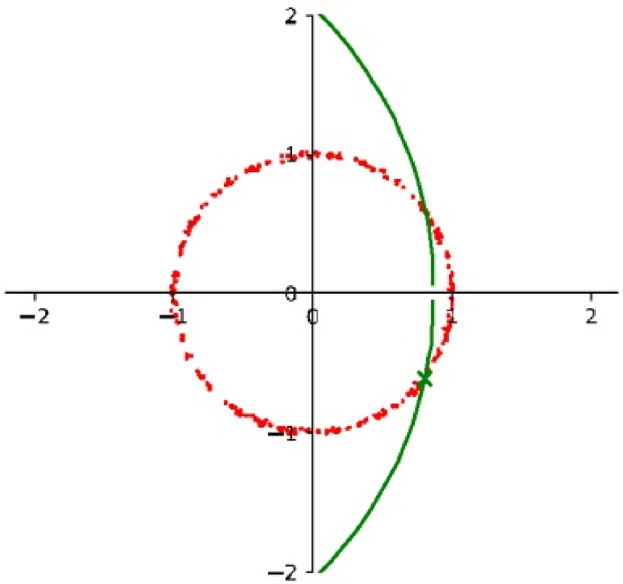
- Assume that each point came from a circle. Let each point vote for the parameters of the circle it could be a part of.
- If there is indeed a circle, the corresponding parameters will receive a lot of votes. Look for peaks.

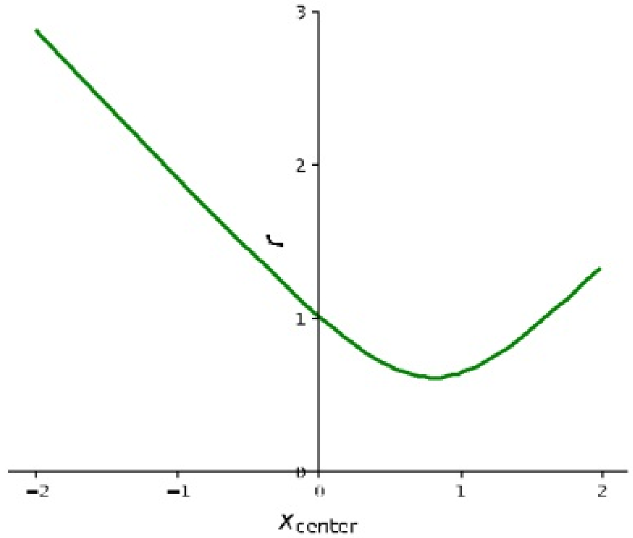




Hough transform for a circle

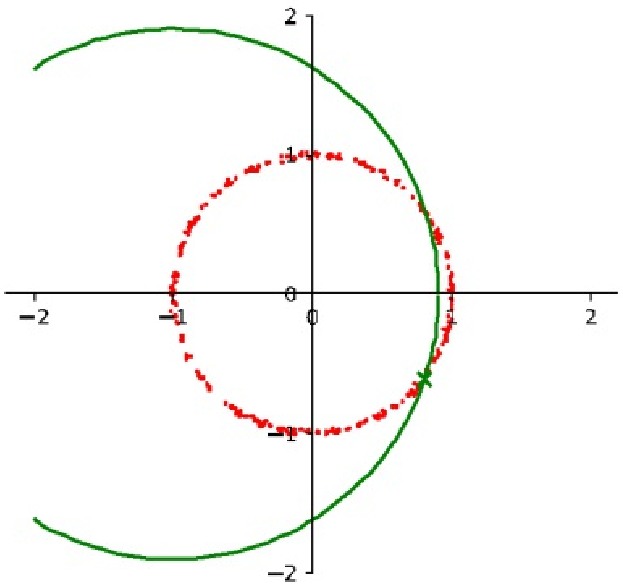
- Assume that each point came from a circle. Let each point vote for the parameters of the circle it could be a part of.
- If there is indeed a circle, the corresponding parameters will receive a lot of votes. Look for peaks.

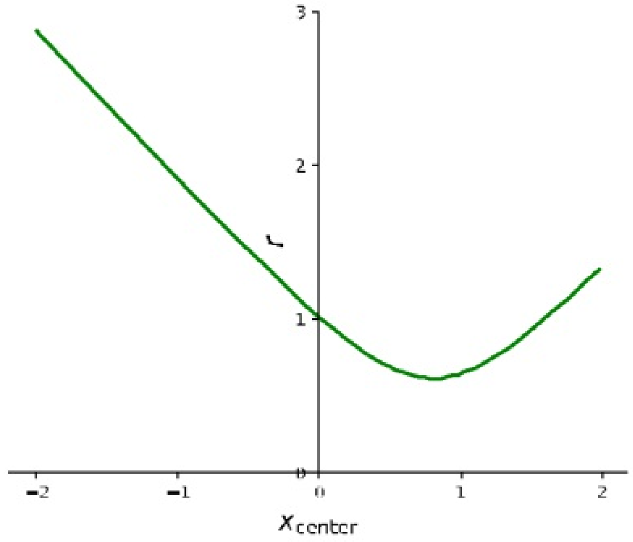




Hough transform for a circle

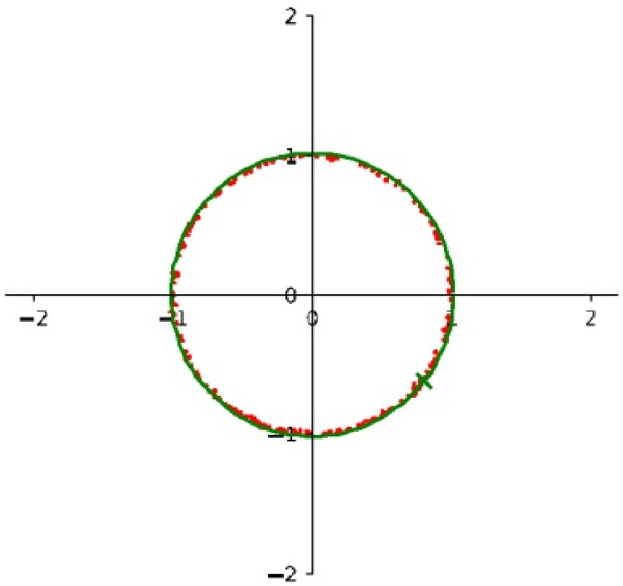
- Assume that each point came from a circle. Let each point vote for the parameters of the circle it could be a part of.
- If there is indeed a circle, the corresponding parameters will receive a lot of votes. Look for peaks.

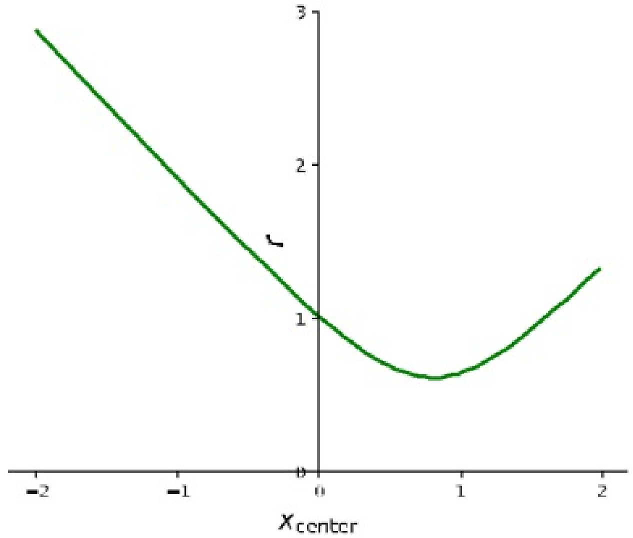




Hough transform for a circle

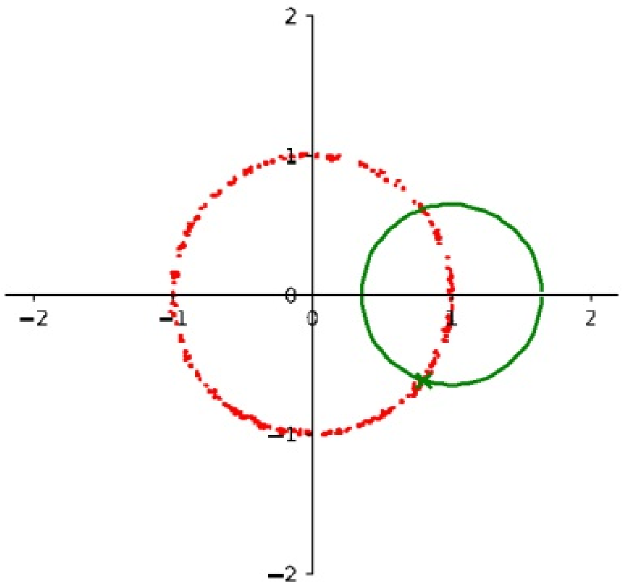
- Assume that each point came from a circle. Let each point vote for the parameters of the circle it could be a part of.
- If there is indeed a circle, the corresponding parameters will receive a lot of votes. Look for peaks.

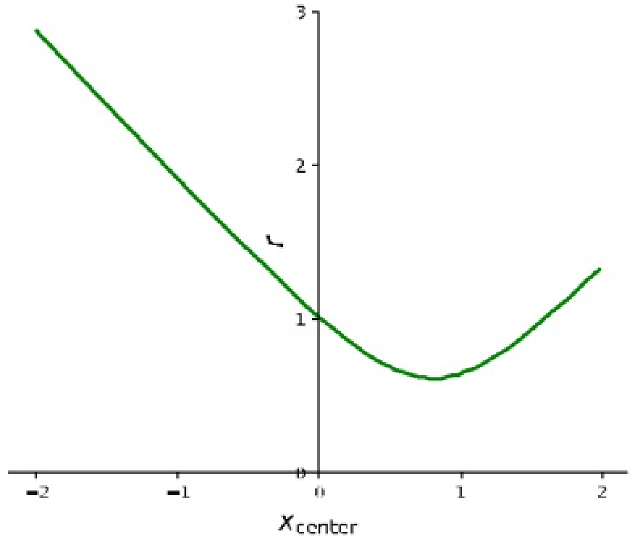




Hough transform for a circle

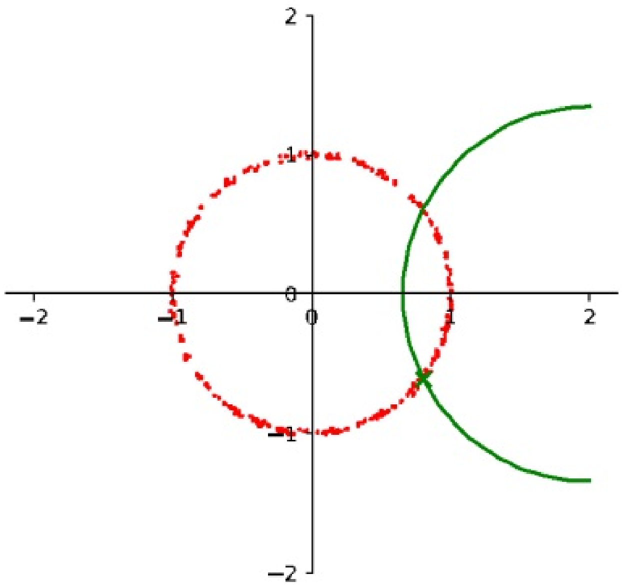
- Assume that each point came from a circle. Let each point vote for the parameters of the circle it could be a part of.
- If there is indeed a circle, the corresponding parameters will receive a lot of votes. Look for peaks.

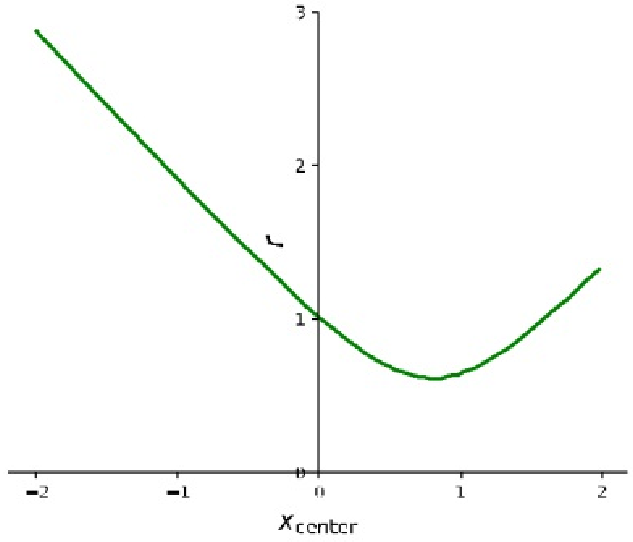




Hough transform for a circle

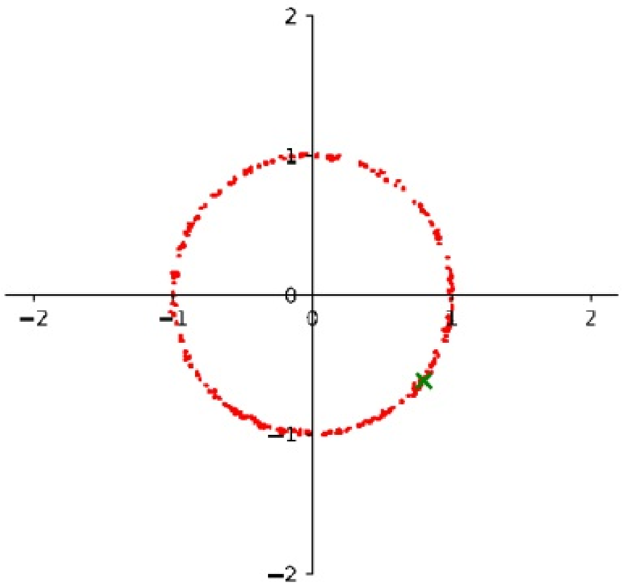
- Assume that each point came from a circle. Let each point vote for the parameters of the circle it could be a part of.
- If there is indeed a circle, the corresponding parameters will receive a lot of votes. Look for peaks.

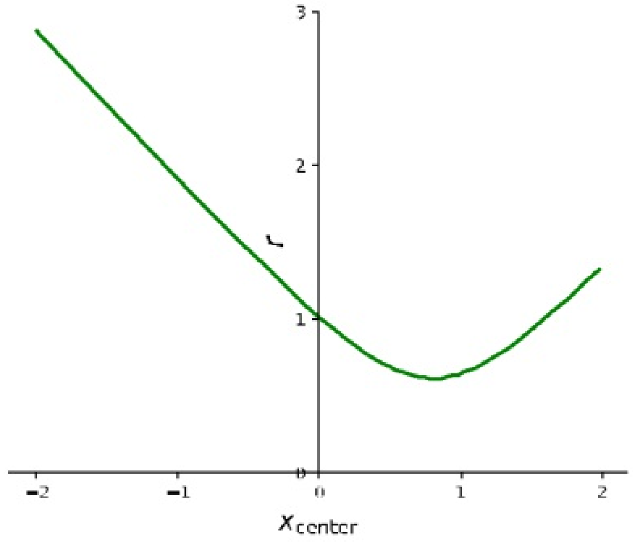




Hough transform for a circle

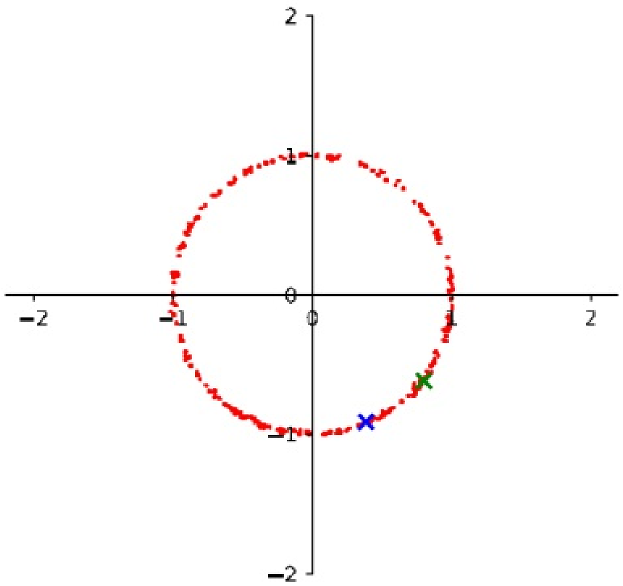
- Assume that each point came from a circle. Let each point vote for the parameters of the circle it could be a part of.
- If there is indeed a circle, the corresponding parameters will receive a lot of votes. Look for peaks.

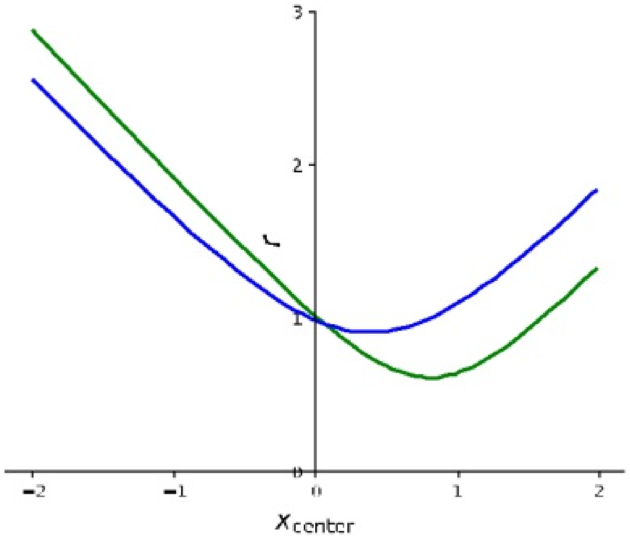




Hough transform for a circle

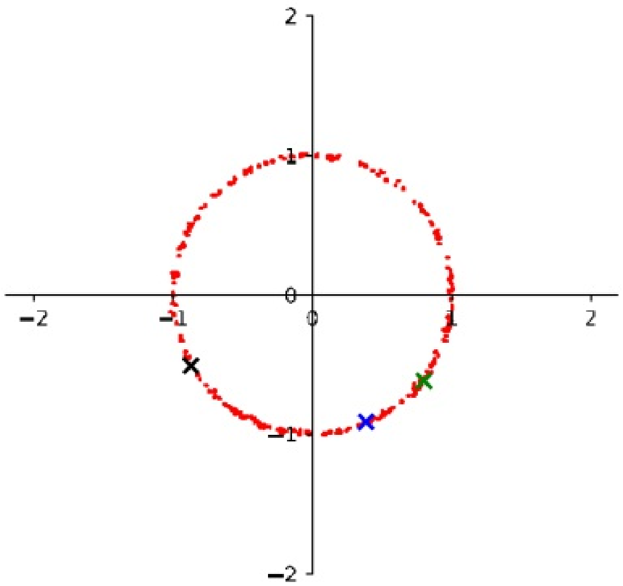
- Assume that each point came from a circle. Let each point vote for the parameters of the circle it could be a part of.
- If there is indeed a circle, the corresponding parameters will receive a lot of votes. Look for peaks.

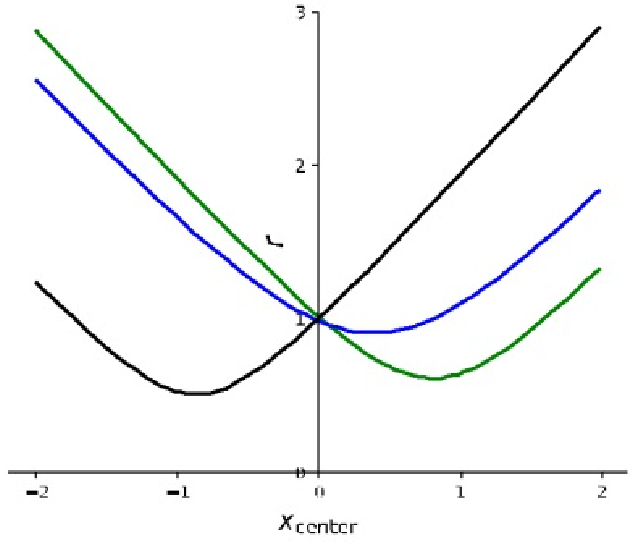




Hough transform for a circle

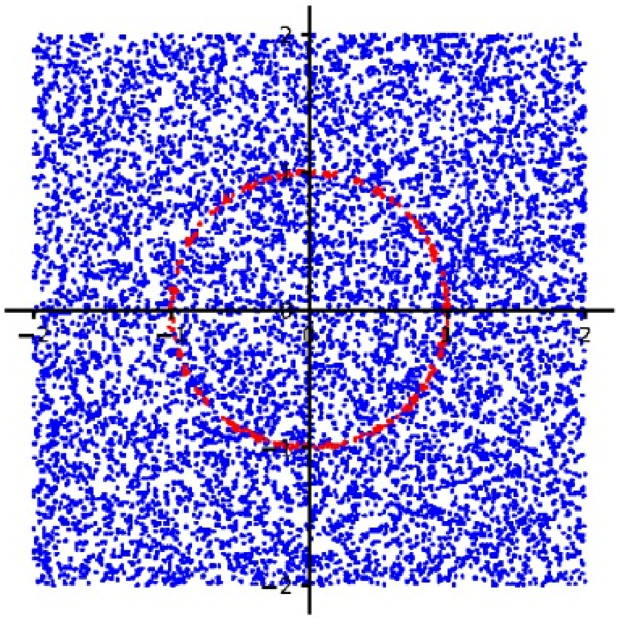
- Assume that each point came from a circle. Let each point vote for the parameters of the circle it could be a part of.
- If there is indeed a circle, the corresponding parameters will receive a lot of votes. Look for peaks.

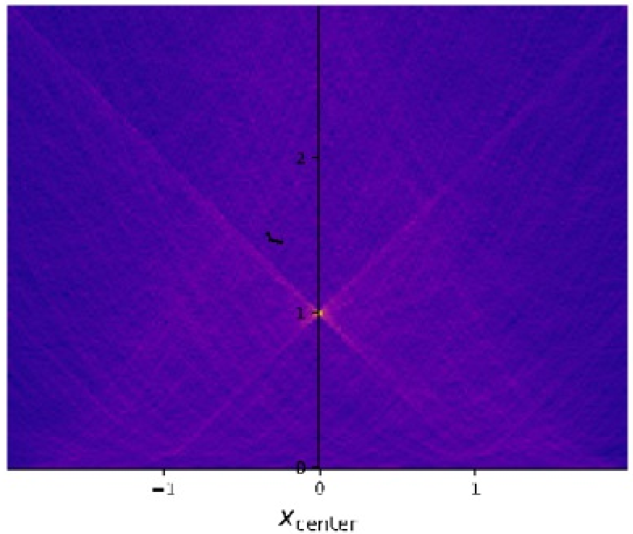




Hough transform for a circle

- 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.





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, x_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.

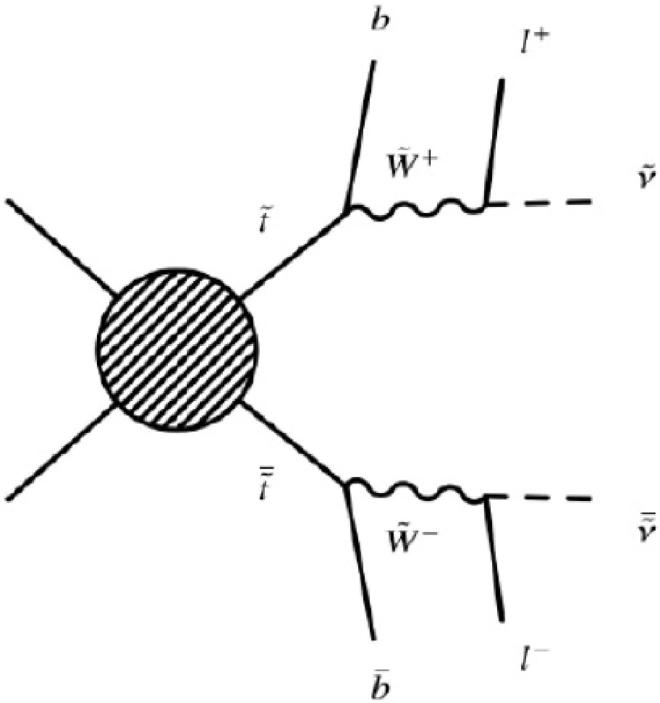
- Just like computers cannot see circles in a plane, we cannot see 8d features in a 9d space
- Port the same idea to physics processes
  - Delta features: Assume that each event came from a given diagram. Vote for all possible mass parameters that allow this
  - Jacobian features: Vote for all possible mass parameters for which the given event will be an extreme event (what I have plots for)

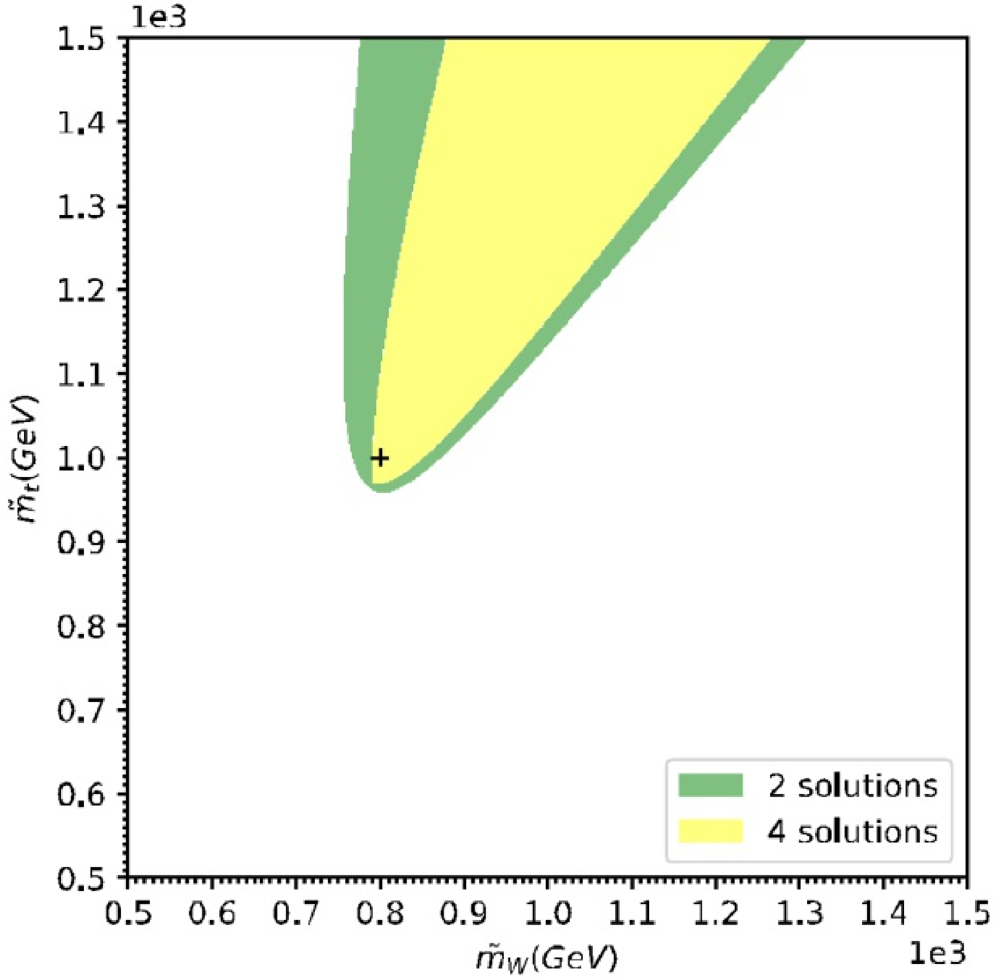
Delta features: Toy example done, physics example skip

Jacobian features: Toy example skip, physics example...

ttbar like

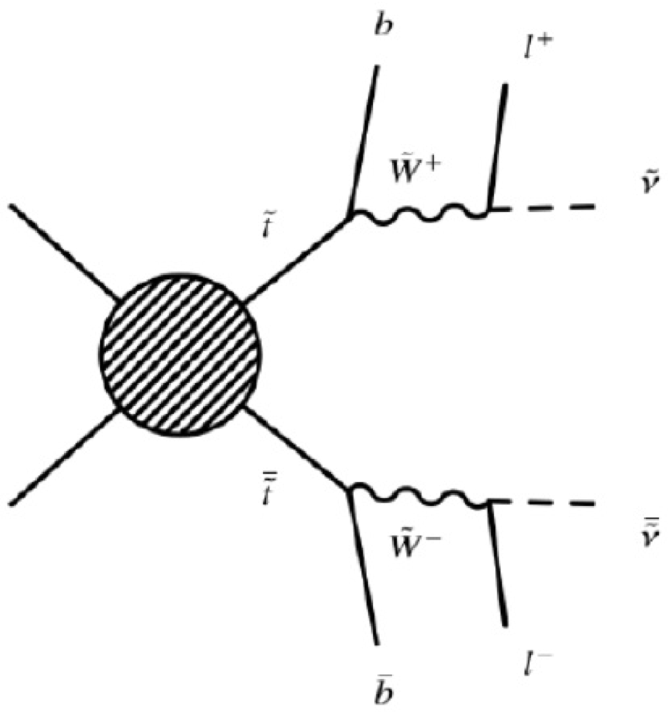
arXiv:1906.02821

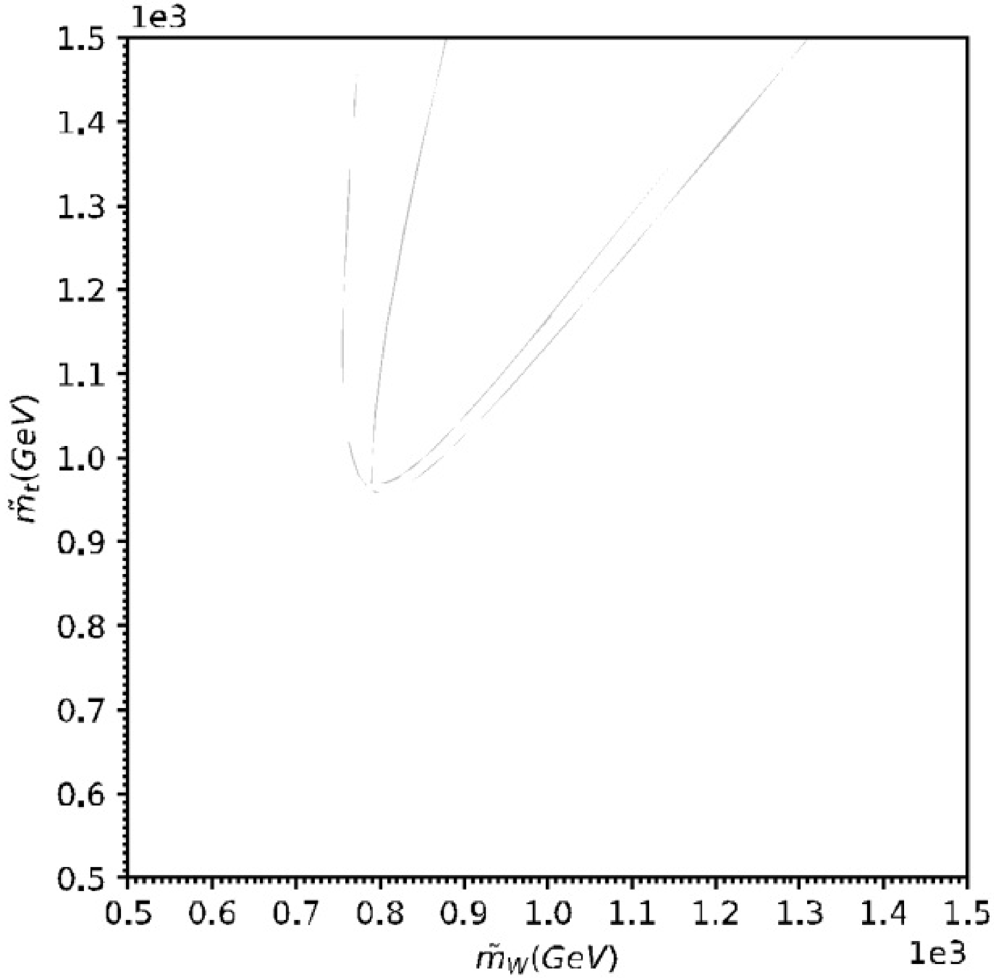




ttbar like

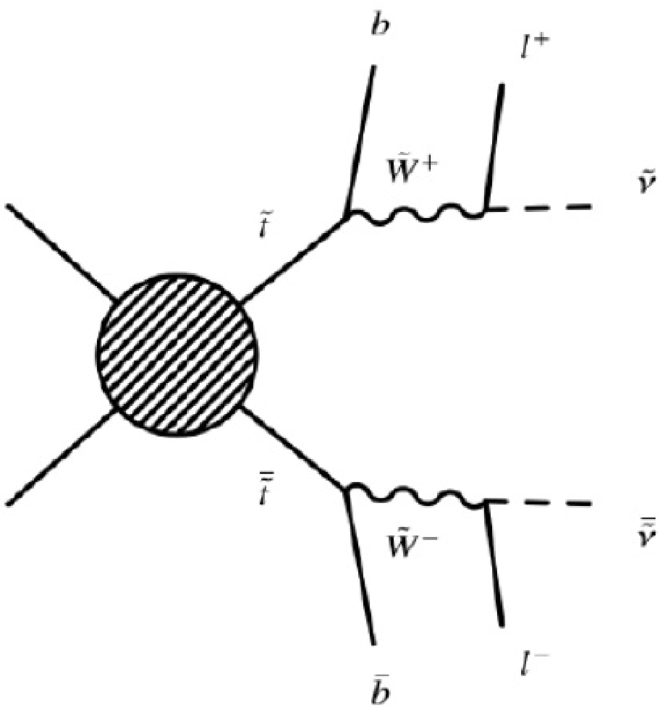
arXiv:1906.02821

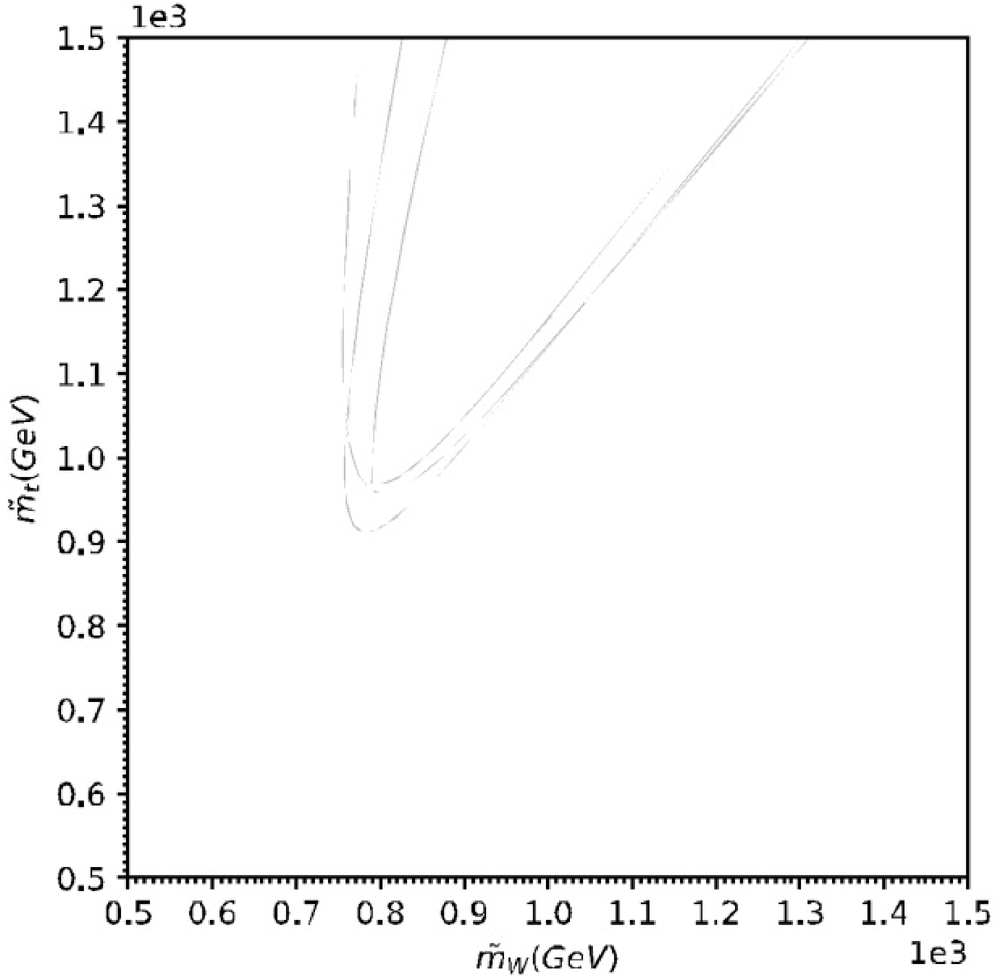




ttbar like

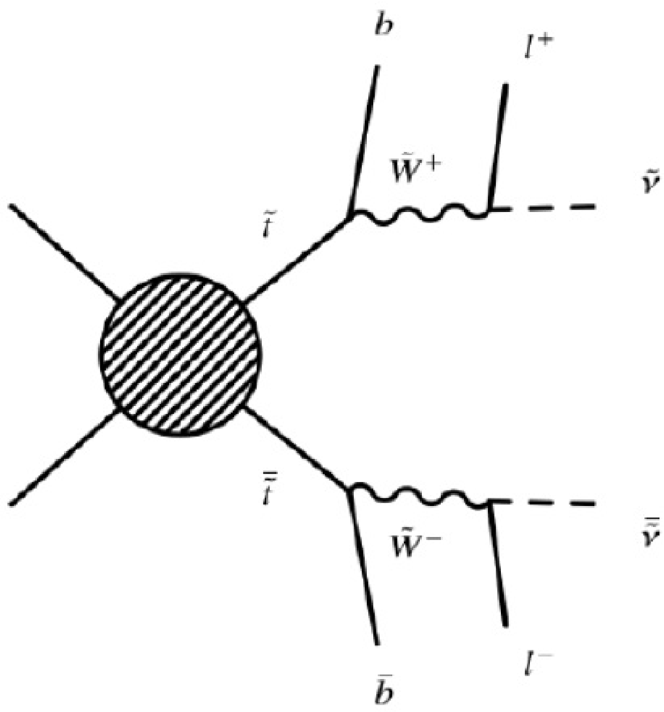
arXiv:1906.02821

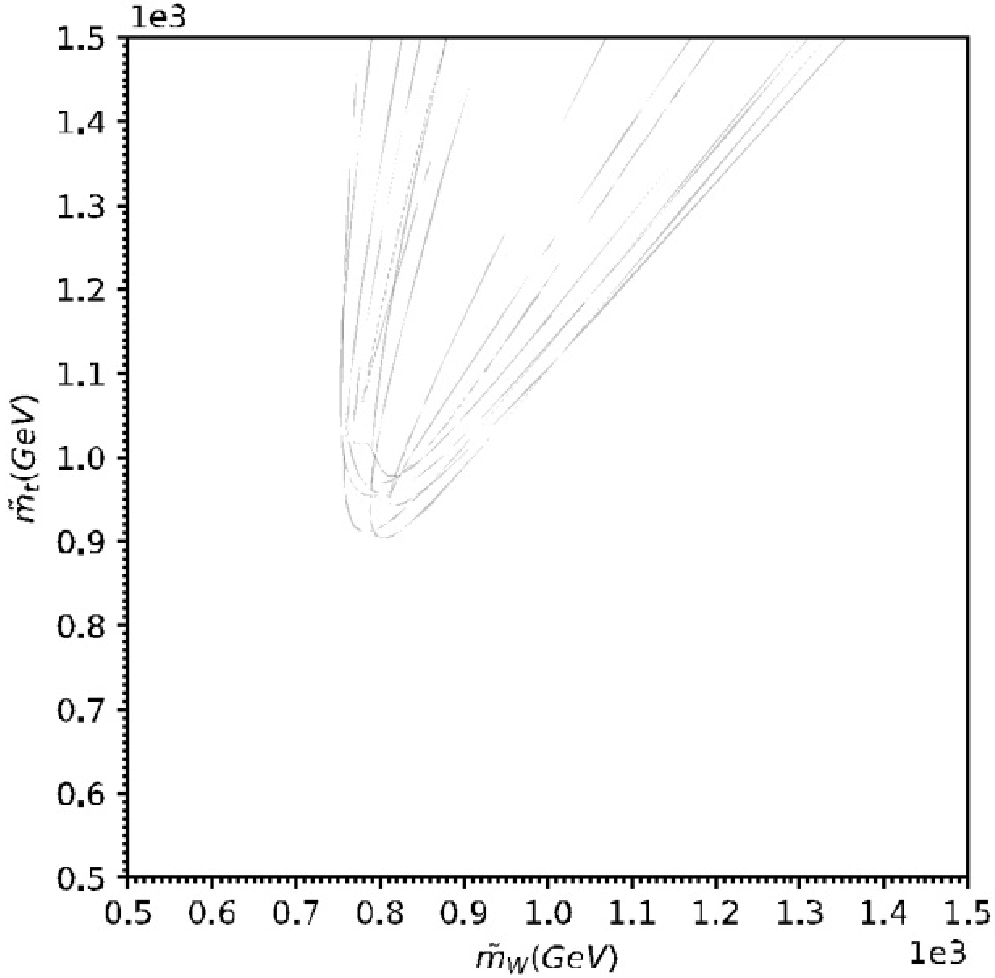




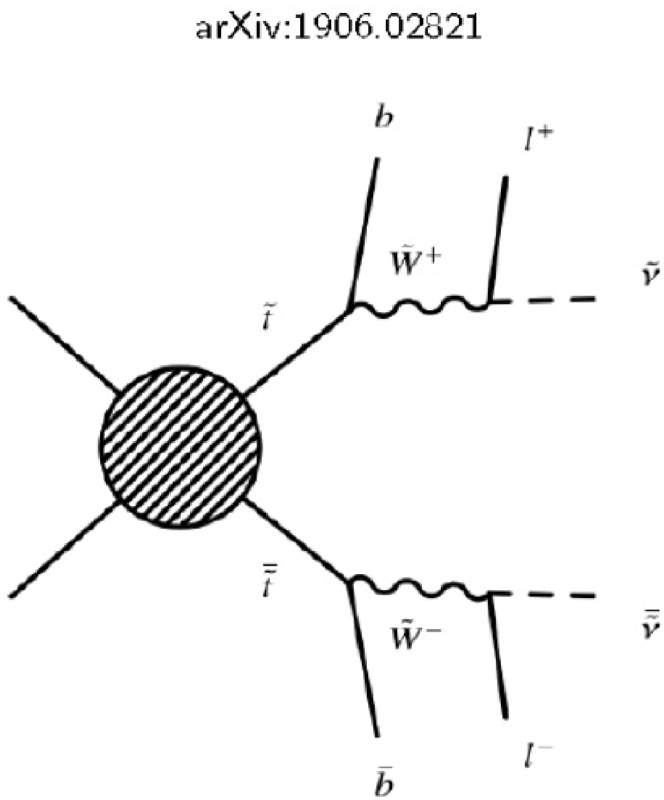
ttbar like

arXiv:1906.02821

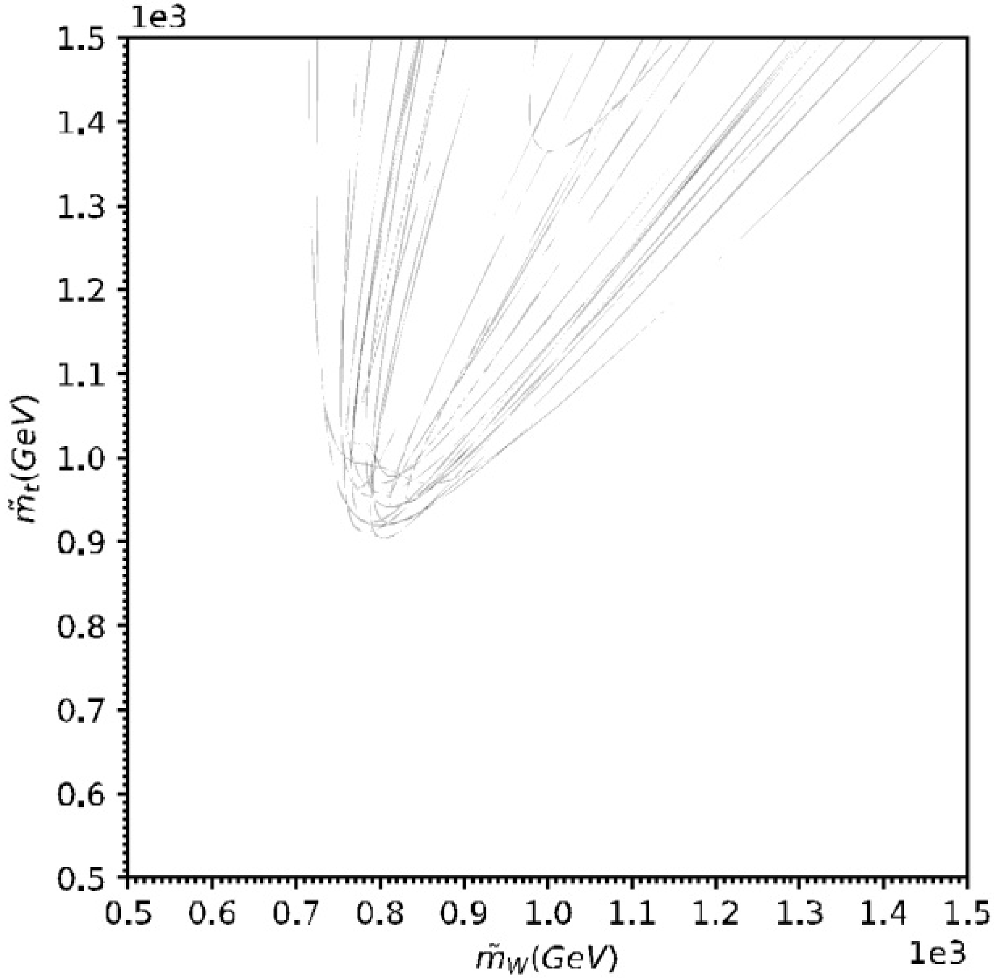




ttbar like



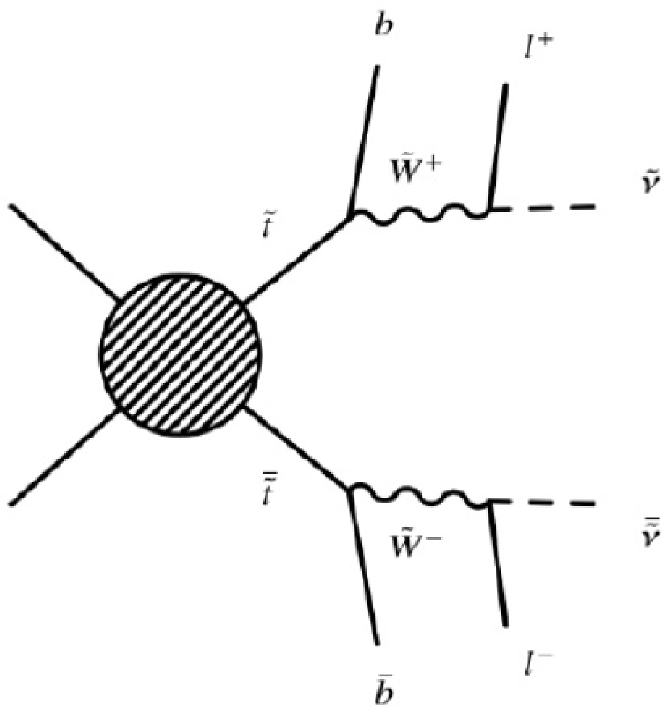
arXiv:1906.02821
b
l^+
$\tilde{W}^+$
$\tilde{\nu}$
$\tilde{t}$
$\bar{\tilde{t}}$
$\tilde{W}^-$
$\bar{\tilde{\nu}}$
$\bar{b}$
l^-

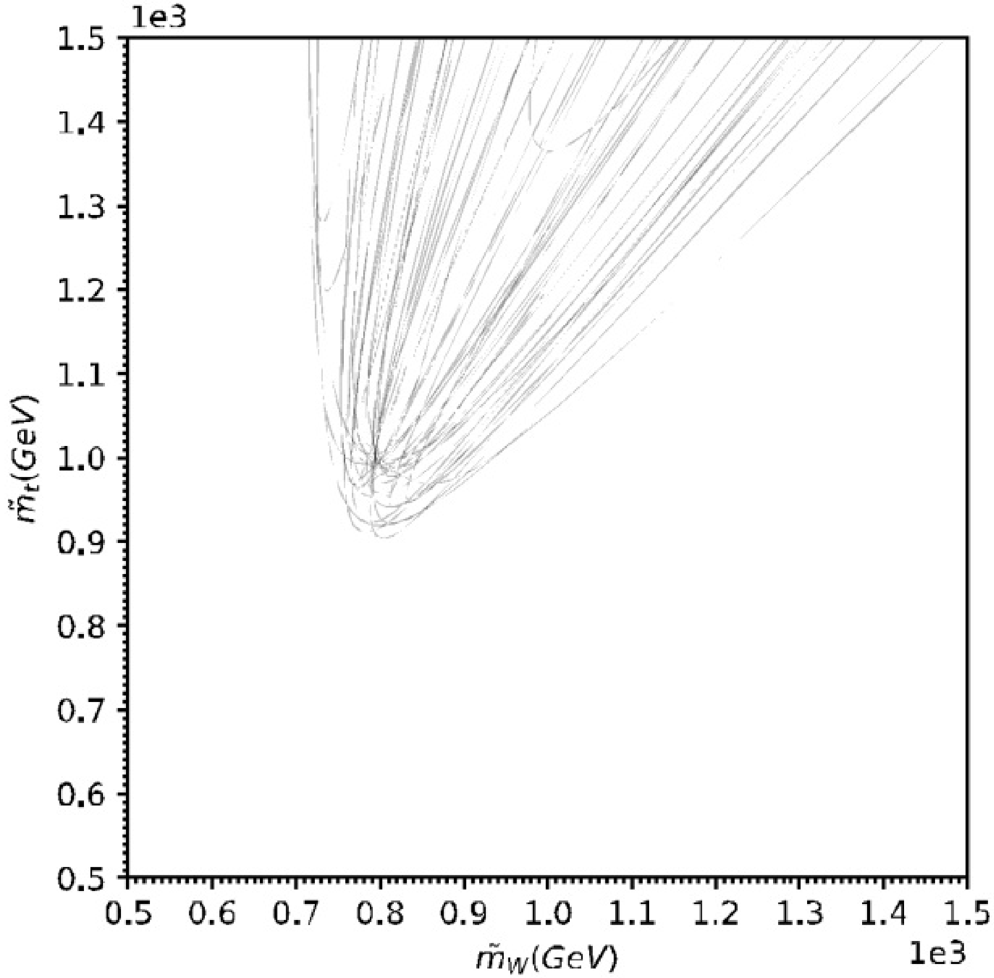


1e3
1.5
1.4
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
$\tilde{m}_t(GeV)$
0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5
$\tilde{m}_W(GeV)$
1e3

ttbar like (It works!)

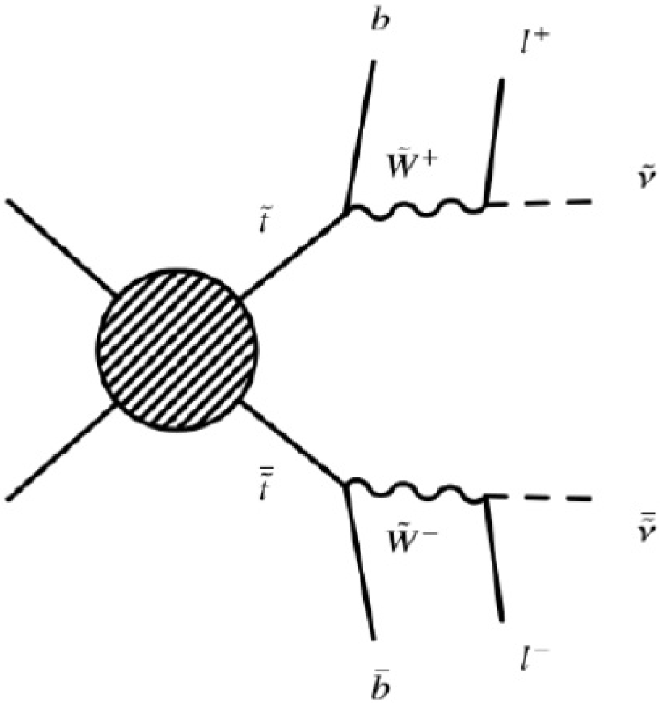
arXiv:1906.02821

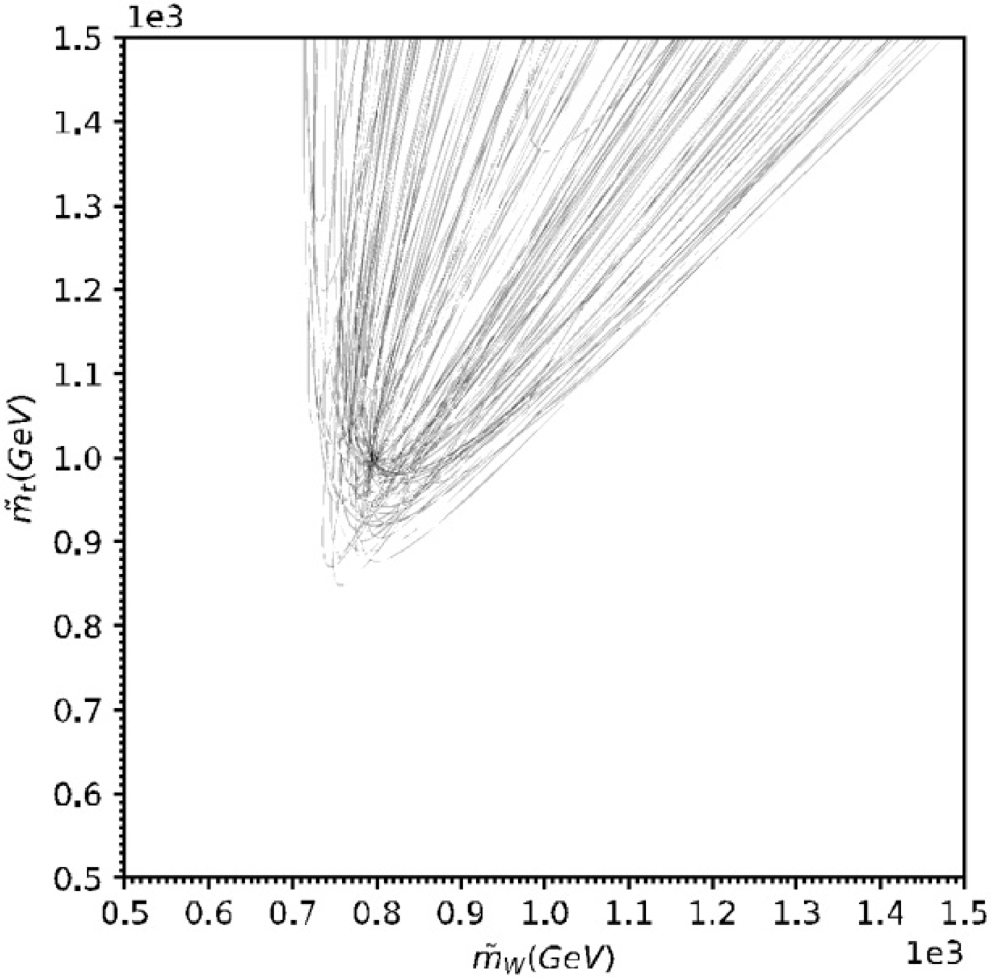




ttbar like (It works!)

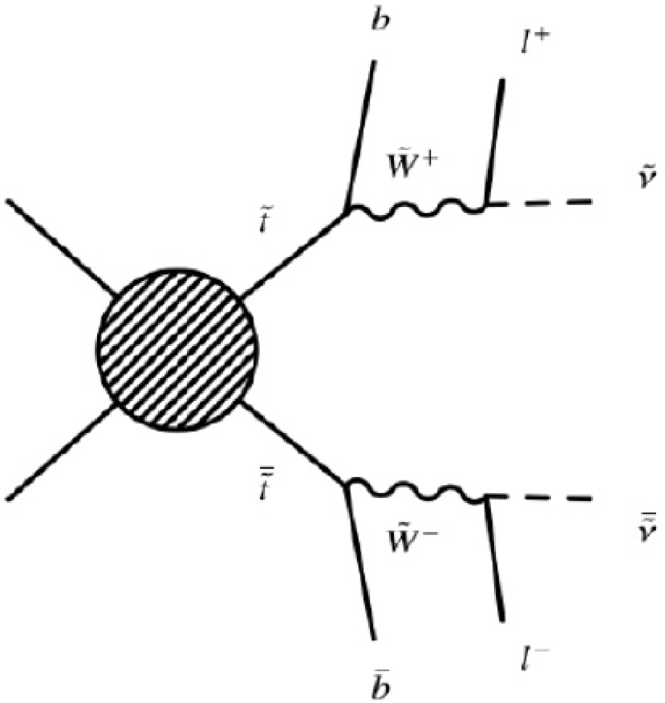
arXiv:1906.02821

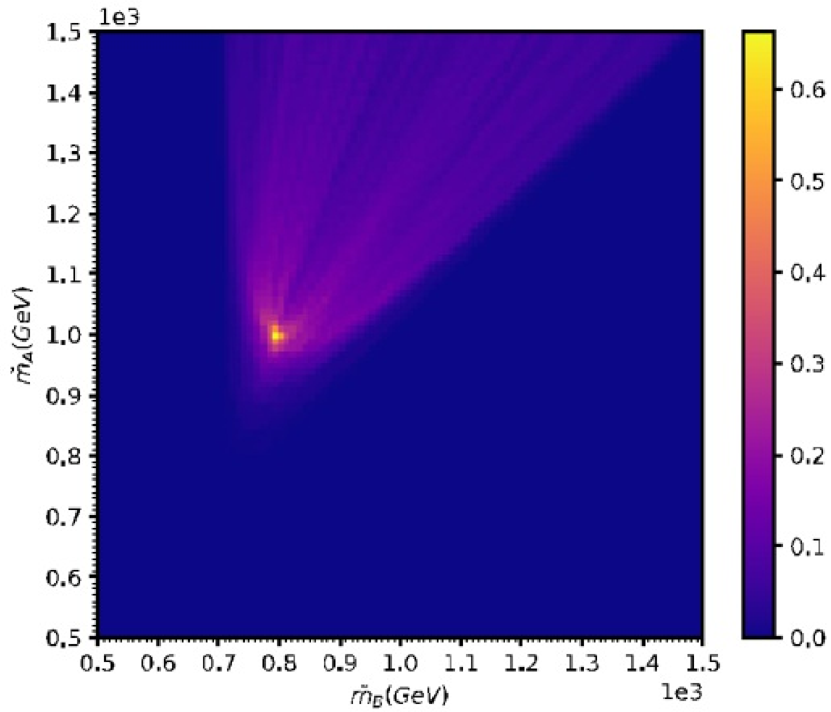




ttbar like (It works!)

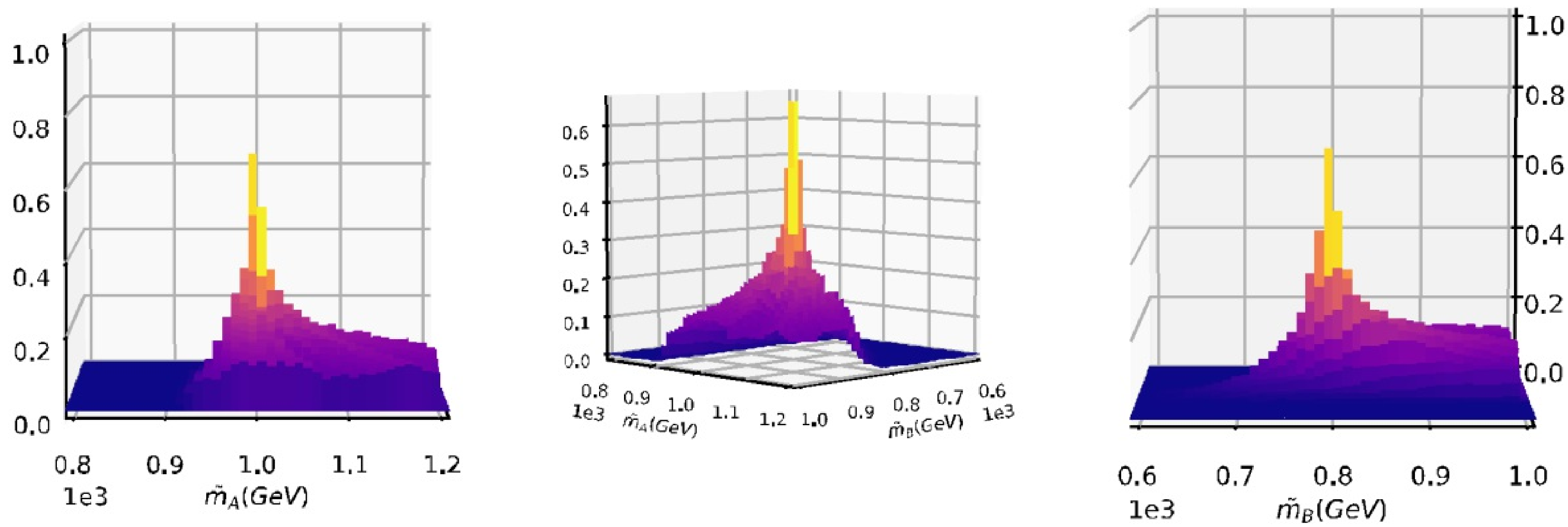
arXiv:1906.02821





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

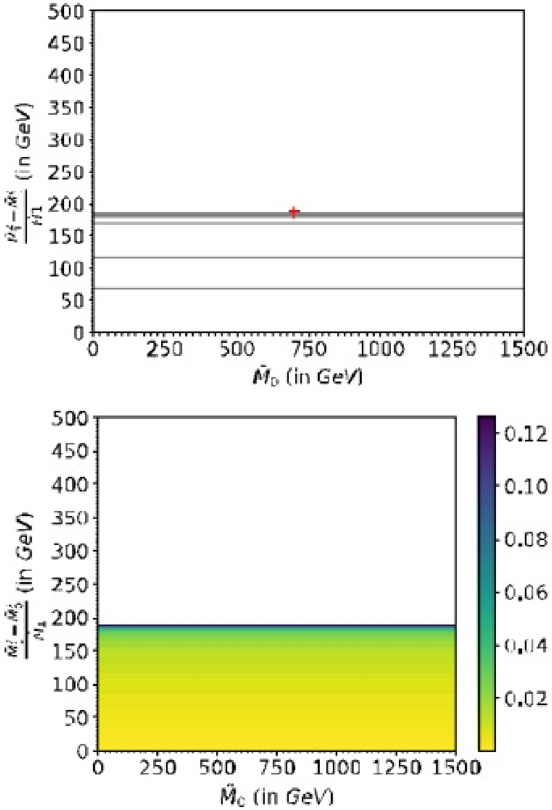
ttbar 3d heatmaps



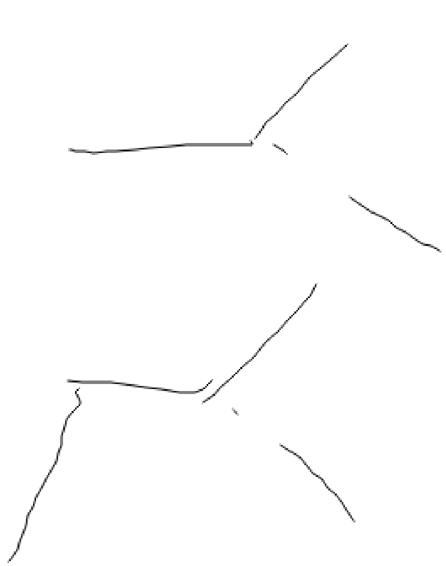
What about $m_{I'}$?

Hough transform for a fixed the neutrino mass is precisely $m_{I'}$

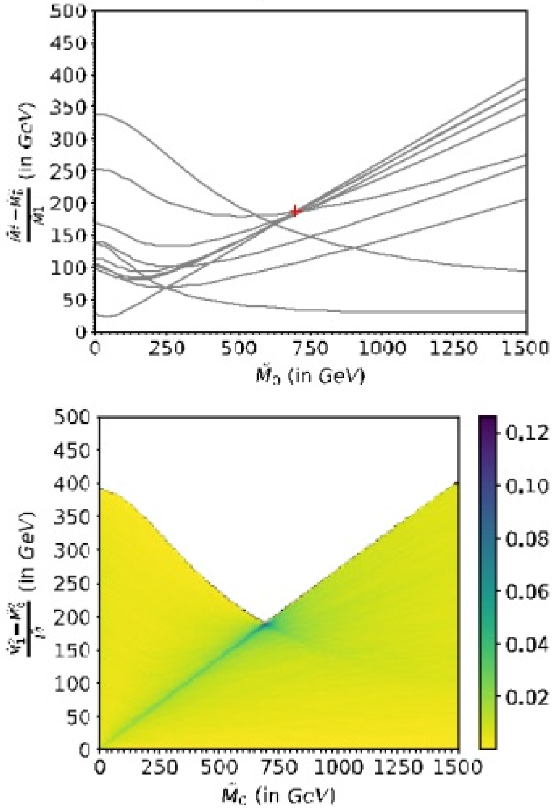
No isr



single W



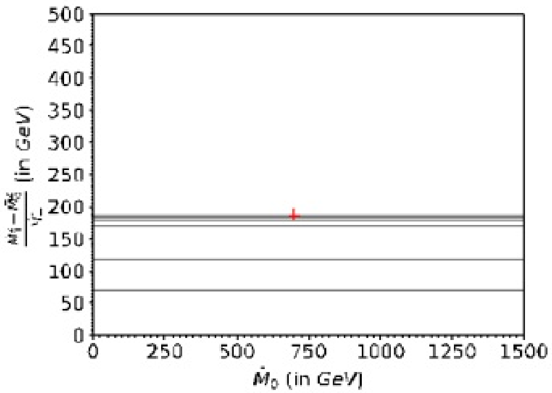
A fixed finite isr



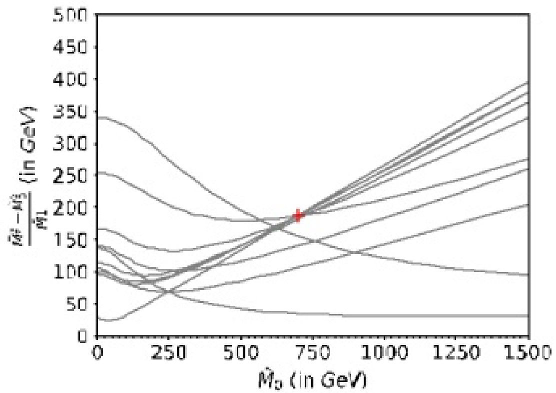
Hough transform - Where does the power come from?

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

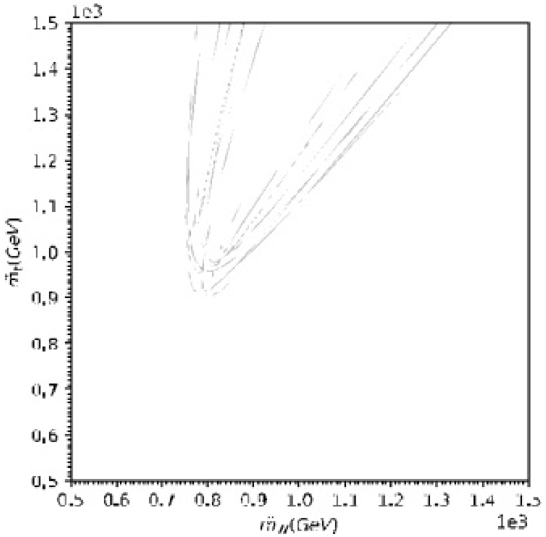
A: We did not reduce the dimensionality! These curves are parametrized by the visible momentum components.
Hough transform is a just transformation. Each curve/surface is a representation of the corresponding event. The more the parameters, the funkier the curves get!



1 parameter



2 parameters (p_T^{isr} fixed)



9 parameters (more like 7)

Hough transform - Where does the power come from?

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.

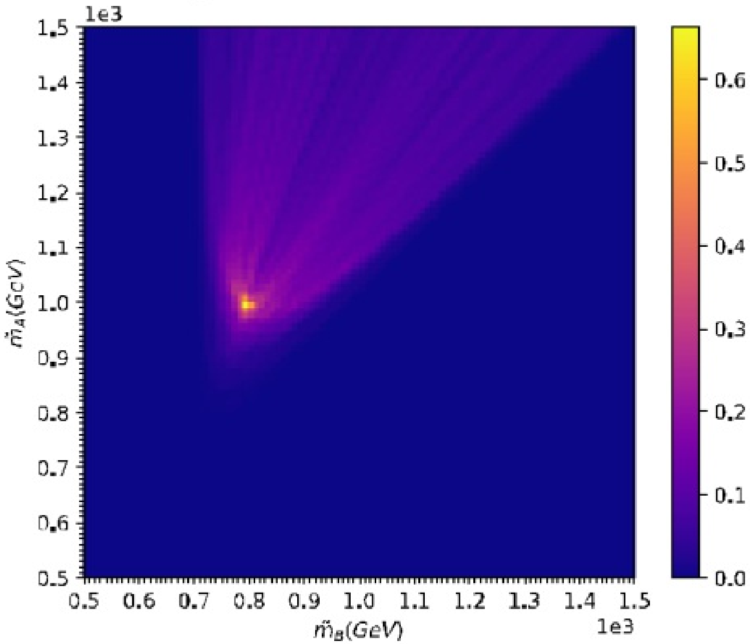
Each diagram will have a different Hough tranform.

Even for a given search, we can have multiple transforms corresponding to each background process.

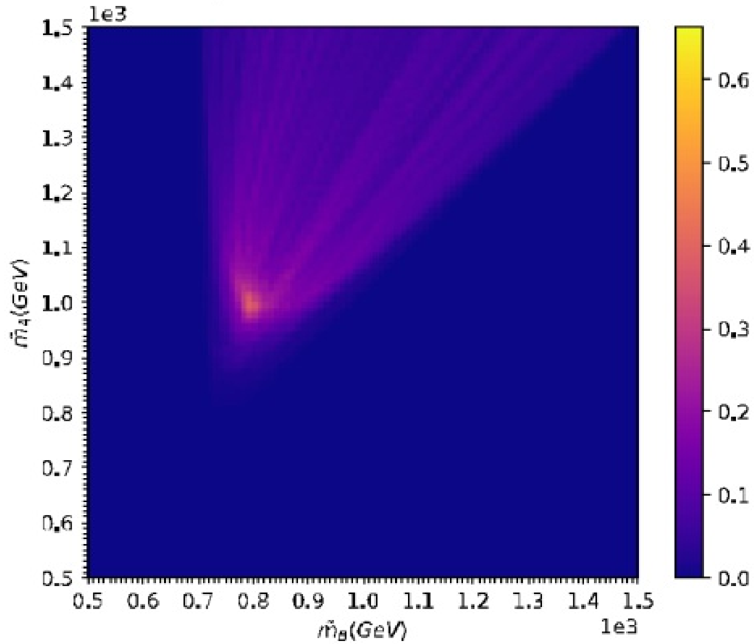
Will it be worth it?

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.

w/o detector effects



w/ detector effects



Will it be worth it?

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...

- An event voting for all the masses that could've produced it is precisely the invariant mass for fully visible decays.
- An event voting for all masses for which it would be extreme is precisely m_T for single W.
- The distributions of these variables don't look impressive after including combinatorics and detector resolution. Yet our choice is clear between using them and not using them.

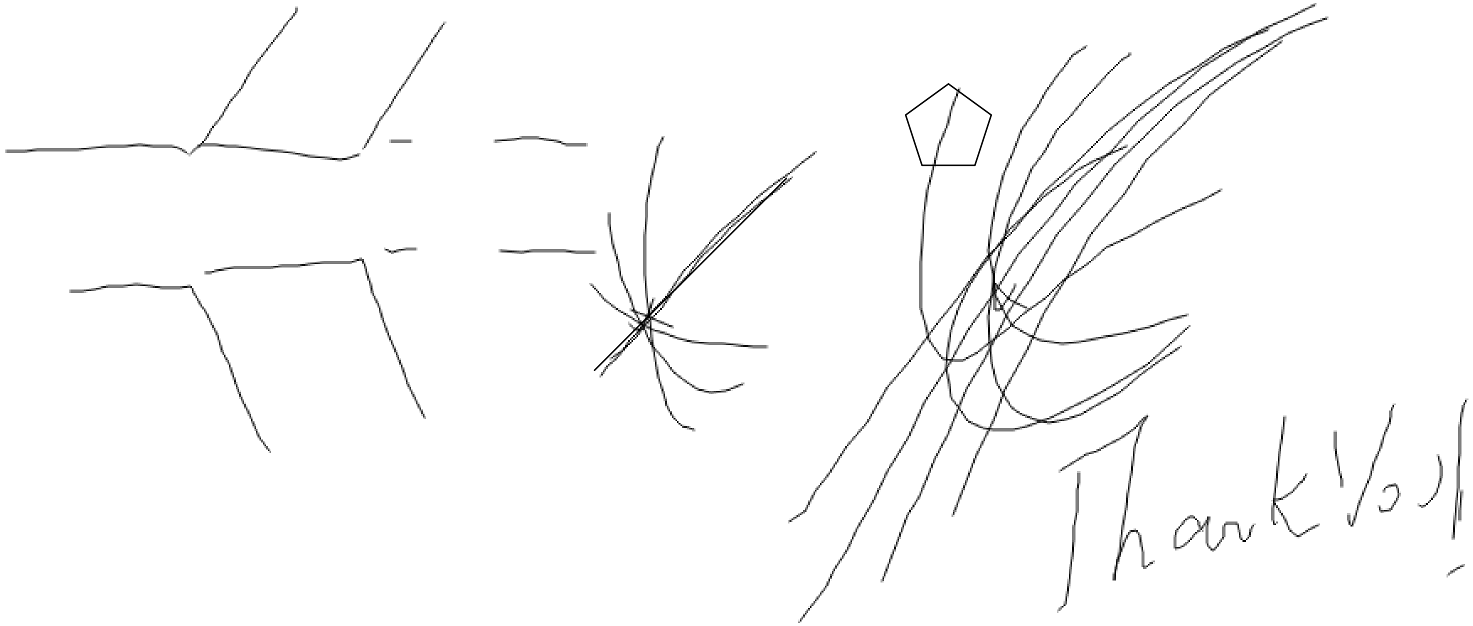
Will it be worth it?

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 tap into the kinematics (key) aspect of those features.
  (So far our only portals to those features have been event categorization and ML)
- Note that for the ttbar case, there are 11 physically relevant observables, and the tranformation retains 9 of them.

Empty slide 1 - Beyond density of curves... Statistics... ML... NEXT GEN...



Empty slide 2 - Underconstrained case...