Jesse Thaler

"Two Lectures on Jet Substructure"

P.TP Summer School
July 23-24, 2013

Outline:

- Overview of Jet Substructure
- Quark vs. Gluon Discrimination
- Boosted Objects
- Jet Grooming
Part 1: Overview of Jet Substructure

A Renaissance in Jet Physics!

→ First jet algorithm: 1970s (Sherman-Weinberg)

→ First jet algorithm that experimentalists & theorists could agree on: 2008 (anti-$k_T$)

For me, 5 amazing years learning about structure of QCD, relevance for BSM physics, realities of experimental methods.

Convergence of 3 communities

Driven by advances at the intersection of these 3 areas.
Experiment: Fantastic Performance of LHC
& Granularity of ATLAS/CMS

Compared to Tevatron:
\[ \approx 3.5-7 \] more energy
\[ \approx 10-20 \] more luminosity
\[ \approx 5 \] better segmentation

Ability to Resolve Individual Hadrons!
Increased Sophistication for Jet Studies.

QCD Theory: Automated LO (\& NLO) Jet Cross Sections
Sophisticated Monte Carlo (ME/PS matching)
New Approaches/Applications of Factorization \& Resummation.

Some Control of Non-perturbative Effects.

BSM Theory: Realization of Importance of Boosted Regime
\[ \sqrt{S} \gg m_{\text{top}}, m_{\nu_1}, m_H, m_{\text{Wino}} \]
Novel Jet Observables Customized for New Physics Searches

New Approaches to Cascade Decays & Multi-Jet Final States
These Lectures: 3 examples of Jet Substructure in Action.

→ Quark vs. Gluon
→ Boosted W/Z/H/top
→ Jet Grooming (time permitting)

Apologies: I will be a bit biased toward my own work, but I'll try to be pedagogical.

Crucial Topic Not Covered: Jet Algorithms

For my purposes: "A jet is a collimated spray of hadrons with a jet radius R."

Ignoring subtlety that quarks/gluons have color charge, but bound hadrons are color singlets.

(Fundamental Ambiguity in Defining Jets typically shows up at \( \mathcal{O}(\frac{\alpha_s}{\ln^2 \frac{p_T}{\Delta R}}) \).)

Ignoring subtle effects of jet boundaries, jet splitting/merging.

(As a community, we've defaulted to anti-kt to address these issues.)
Part 2: Quark vs. Gluon Jets

A jet is a jet is a jet?

| Light Quarks | d\bar{u} s\bar{d} s | \Rightarrow \text{Indistinguishable unless some kind of jet charge information is used} |
| Heavy Quarks | c\bar{c} b\bar{b} b | \Rightarrow \text{Can tag because of lifetime/decays of resulting } D \text{ and } B \text{ mesons} |
| Top Quark | \frac{t}{\bar{t}} \frac{\bar{t}}{t} | \Rightarrow \text{Totally different beast, see next lecture. (} t \rightarrow bW) |
| Gluon | \frac{g}{\bar{g}} \frac{\bar{g}}{g} | \Rightarrow \text{Different Color Charge from Quarks. Should give rise to different looking jets.} |

This lecture: Build up observable sensitive to $C_F$ (quarks) vs. $C_A$ (gluon)

Punchline: In the "strongly-ordered" limit, $\text{Quark efficiency} \times \frac{G_A/C_F}{x} = x^{-a/4}$
First, why do jets form?

1) Soft-Collinear Singularities in QCD

\[ p = p_1 + p_2 \quad \text{with} \quad p_1^2 = p_2^2 = 0 \]

When is \( p^2 = 0 \)?
- When \( \vec{p}_1 \parallel \vec{p}_2 \) (Collinear Limit)
- When \( |\vec{p}_2| \to 0 \) (Soft Limit)

2) At sufficiently high energies, \( \alpha_s \) is small.

\[ \alpha_s(m_Z) \approx 0.12 \]

Otherwise, many soft-collinear emissions start to look like ball (instead of cone) of radiation.
3) Color "strings" break

\[ \Rightarrow \text{Direction of } q/g \cong \text{Direction of Jet} \]

Confining Flux

Energetically Favorable to Sever Flux Tube

Otherwise, you would get excited hadron states instead of jets.

QCD in the Soft-Collinear Limit

Energy Flow of Partons \( \cong \) Energy Flow of Hadrons

Allows us to study jets at partonic level (up to some non-perturbative effects)
Key Exercise: Show that \( \left( \text{in soft } q = \text{collinear limit} \right) \)

\[
\Sigma_{\text{polar}} \int dq_{\text{jet}} dq_{\text{soft}} \left| \frac{q}{q_{\text{jet}}} \right|^{2} \quad \text{with } q \text{ and } q_{\text{jet}} \text{ in the same jet}
\]

\[
= \Sigma_{\text{polar}} \int dq_{\text{jet}} dq_{\text{soft}} \left| \frac{q}{q_{\text{jet}}} \right|^{2} \quad \text{Swap } C_{F} \rightarrow C_{A} \text{ for gluon jets}
\]

\[
\times \left( \int_{0}^{1} d\tau R \int_{0}^{2\pi} d\Theta \right) \frac{2d\tau}{\tau} \quad \text{soft/collinear singularities}
\]

Energy Fraction \( z = \frac{E_{\text{soft}}}{E_{\text{jet}}} \) \( \rightarrow 0 \) soft limit

Splitting Angle \( \Theta = \Theta_{qg} \) \( \Theta \rightarrow 0 \) collinear limit

Color Factors:

Quarks: \( \sum_{a} \frac{a_{ij} \cdot a_{jk}}{a_{si}} = C_{F} \delta_{ik} \)
\( C_{F} = \frac{N^{2} - 1}{2N} \rightarrow \frac{4}{3} \)

Gluons: \( \sum_{a} \frac{f_{abcd} f_{abc}}{a_{i}} = C_{A} \delta_{cd} \)
\( C_{A} = N \rightarrow 3 \)

(In just collinear limit, AP splitting functions)

(In just soft limit, antenna functions)

(Above formula for soft and collinear limit only)
Quark vs. Gluon Jets?

$C_A > C_F$, so gluon jets should be "fatter."

Need observable sensitive to this difference.

Goal: \( \frac{1}{d \sigma} \)

(Ideally, predict distributions in QCD,
Validate in Monte Carlo,
Test in LHC data.)

Pause for a moment: Why do we care?

Experiment: Different calibration for different kinds of jets.

QCD Theory: Build observables sensitive to color structure of QCD

BSM Theory: Many BSM signals are quark-rich, while backgrounds are gluon-rich.
Crucial Feature of an Observable
(if you want to calculate in perturbative QCD)

Infrared / Collinear Safety: (IRC safety)

Observable should be insensitive to
infinitely soft or infinitely collinear radiation.

Why?

Technically, needed so IR divergences cancel
between real and virtual diagrams. (KLN theorem)

Physically, very soft & collinear radiation controlled
by $A_{\text{coll}}$, and you want your observable
to be as insensitive as possible to non-perturbative effects.

IRC Safe Observables: Accurately described
in partonic language.

($p_T \rightarrow E$, $B \rightarrow \Theta$ for simplicity)

\[ C_1^{(B)} = \frac{\sum_{i \neq j} E_i E_j \Theta_{ij} \beta}{(\sum_i E_i)^2} \]

$\beta > 0$

Sum over particles in a single jet

Soft safe? Yes, $C_1$ doesn't change if $E_i \rightarrow 0$

Collinear safe? Yes, $C_1$ is additive, so $E \rightarrow E_1 + E_2$

has no effect (for $\beta > 0$)
Q: What is quark/gluon discrimination for $C_1$?

Hard to answer in general, but we'll go to "strongly-ordered" limit (sometimes called "leading log" or "double log" limit)

Hard quark/gluon emits hierarchical pattern of soft/collinear emissions

$$P_{\text{emit}}(z, \theta) \, dz \, d\theta = \frac{2\alpha_s}{\pi} \, C_F, A \, \frac{dz}{z} \frac{d\theta}{\theta}$$

Uniform emissions in $(\log \frac{1}{\theta}, \log \frac{1}{z})$ plane.

Strongly-ordered: Observable dominated by hardest emission(s)
What is $C_1$ in strongly ordered limit?

\[
\frac{g_2 E_2}{q_1} - q_2 \frac{E_2}{E_1} \quad \beta
\]

\[
C_1 = \frac{E_1 E_2 \Theta_{12}}{(E_1 + E_2)^2} \quad E_2 \ll E_1
\]

\[
z = \frac{E_2}{E_1 + E_2}
\]

\[
\beta \quad \Theta_{12} \quad \text{quark/gluon splitting angle}
\]

\[
\text{gluon energy fraction}
\]

\[
\log \left( \frac{1}{z} \right)
\]

as many emissions as you want.

\[
\text{dominant emission}
\]

\[
\text{no emissions here}
\]

\[
\log \frac{1}{E} = \log \frac{1}{z} + \beta \log \frac{1}{\Theta}
\]

Key Exercise: Probability to get a value of $C_1$ less than $C_1^{\text{max}}$:

\[
\mathcal{P}_{C_1}(C_1^{\text{max}}) = \mathcal{P}_{C_1} \left( \text{area under } C_1^{\text{max}} \text{ curve} \right)
\]

\[
\mathcal{P}_{C_1}(c_1) = e^{-\frac{2\pi s}{\pi} \mathcal{P}_{C_1}} \left( \text{area under } C_1^{\text{max}} \text{ curve} \right)
\]

\[
= \left( \frac{1}{\mathcal{P}_{C_1}} \right)^{\mathcal{P}_{C_1}} + \frac{1}{2} \frac{\mathcal{P}_{C_1}^2}{\mathcal{P}_{C_1}} + \cdots
\]

\[
\left( \text{no emissions } O(\mathcal{P}_{C_1}) \quad O(\mathcal{P}_{C_1}^2) \right)
\]
\[ \Sigma_{F} \left( c_{\text{max}} \right) = e^{-\frac{a_{s}}{\pi} \frac{C_{F}}{b} \log^{2} \frac{R}{c_{\text{max}}}} \]

\[ \Sigma_{g} \left( c_{\text{max}} \right) = e^{-\frac{a_{s}}{\pi} \frac{C_{A}}{b} \log^{2} \frac{R}{c_{\text{max}}}} \]

-called a Sudakov form factor

Time to interpret this result!

1) What is the cross section?

\[ \frac{1}{\sigma} \frac{d\sigma}{dc_{1}} = \frac{d}{dc_{1}} \Sigma_{g} \left( c_{1} \right) \]

\[ = \frac{2a_{s}}{\pi} \frac{C_{F}}{b} \frac{1}{c_{1}} \log \frac{R}{c_{1}} e^{-\frac{a_{s}}{\pi} \frac{C_{A}}{b} \log^{2} \frac{R}{c_{1}}} \]

At fixed order \( O(a_{s}) \)

\[ \frac{1}{\sigma} \frac{d\sigma}{dc_{1}} \text{ factor of 9.4} \]

Delta function at origin from virtual diagrams to make sure distribution is normalized.

In strongly ordered limit

\[ \frac{1}{\sigma} \frac{d\sigma}{dc_{1}} \rightarrow \text{ we see the separation.} \]
2) What is the discrimination power?

Place a cut $C_1 < C_{1\text{ cut}} \Rightarrow \text{"quark"}$
$C_1 > C_{1\text{ cut}} \Rightarrow \text{"gluon"}$

Want to know efficiency $\Sigma_q (C_1)$ vs. mistag $\Sigma_g (C_1)$

Key: $\Sigma_g = (\Sigma_q)^{CA/C_F \to 1/4}$

In strongly-ordered limit, independent of $\beta$!

(i.e. at this order, only difference between quarks and gluons is $C_F$ vs. $C_A$.

Have to work much harder to optimize $q$ vs. $g$ discrimination.)
Higher Order Effects
- Multiple Emissions
- Subleading Terms in Splitting Functions
- Fixed Order Corrections
- Running $\alpha_s$

General Strategy for Jet Substructure
- Figure out your goal
  ("Discriminate Quarks vs. Gluons")
- Identify Underlying Physics
  ("$C_F$ vs. $C_A$")
- Construct (clever?) observable to probe that physics
  ($C_i = \frac{\sum_{ij} E_i E_j \Theta_{ij}}{(\sum E_i)^2}$)
- Use analytic / Monte Carlo methods to determine distributions / discrimination power
  ("strongly-ordered limit")
- Convince experimentalists to apply to data.

Next time: Multi-Prong Jets & Jet Grooming
(More complicated observables) (Dealing with contamination)
Part 3: Boosted Objects

Last time: Quarks vs. Gluons Intrinsic Jet-Like
Used $C_F$ vs. $C_A$ to discriminate
Analytically tractable in strongly-ordered limit

This time:
\[ \{ \begin{align*}
W/Z/H & : \text{Intrinsically 2 Jet-Like} \\
top \text{ quark} & : \text{Intrinsically 3 Jet-Like}
\end{align*} \]
Active research ongoing to get
analytic control. This lecture will
be more heuristic.

Key point: In extreme kinematic circumstances,
$W/Z/H/top$ can look like
single jet.

\[ m_H^2 \leq p_{Tb} p_{Tb} \Delta R_{b_b} \]
\[ p_{T_H} \approx p_{Tb} + p_{Tb} \]

For $p_{T_H} \gg \frac{2m_H}{R_{Jet}}$
boosted Higgs looks like
a single Jet.
"But Higgs jet has a mass of 126 GeV whereas quarks & gluons are massless."

Wrong! Quark and gluon jets are not massless.

**Key Exercise:** In soft-collinear limit show that

\[
\langle m_{\text{jet}}^2 \rangle \approx -s \frac{C_F A}{\pi} \frac{P_{T\text{jet}}}{R_{\text{jet}}} R^2
\]

versus

\[
m_H^2 \approx \frac{P_{T\text{jet}}^2}{2} R_{\text{jet}}^2
\]

Need observables to distinguish these two cases.

**Obvious choice:** Jet mass

But want other observables to test for 2-prong nature of Higgs jet.
Leading Order Structure

- H \rightarrow b

Democratic energy sharing
Flat distribution in $z = \frac{E_b}{E_H}$

vs.

- q \rightarrow q + g

Soft singularity
Expect $1/z$-like behavior.

Subleading Structure

- H \rightarrow b + g
  - color singlet
  - radiation confined to $b\bar{b}$ dipole

- q \rightarrow q + g
  - quark/gluon color-connected to rest of event

Generic Boosted Object Strategies

- Toss out wide-angle soft radiation (see jet grooming)
  - (large contribution to mass, but unlikely to come from hard subject)

- Identify democratic energy sharing (i.e. measure $z$)

- Identify $N$-prong nature of jet ($N=2 \text{ or } 4$)

- Probe color structure of jet (singlet vs. nonsinglet)
Algorithms / Observables for each of these strategies

(No way I can talk about all of them, so I'll mention just two.)

BDRS (2008): Started whole jet substructure industry for $pp \rightarrow Z\ell\ell$

*Actually, earlier (forgotten) work by Mike Seymour (1994) [1991]

1) Recursively cluster a jet using "CA algorithm"
(join nearest neighbors until you reach jet radius $R$)

\[
\text{Ejet} = \sum E_i, \quad \text{tree-like structure}
\]

2) Un-wind the jet, testing for symmetry and mass drop
($E_1 > E_2$) Require $\frac{E_2}{E_1} \geq \text{cut}$, $\frac{\text{max}(m_1, m_2)}{M_{\text{cut}}} < \text{cut}$

3) If test fails, throw out softer subject, continue recursion on harder subject.
If test succeeds, you've found a candidate Higgs jet.
Why does RDRS work?

$\mu_{cut} \Rightarrow$ good test for democratic energy sharing

$\mu_{cut} \Rightarrow$ QCD jets get mass from many soft wide-angle emissions, whereas Higgs has a hard $H \rightarrow b\bar{b}$ splitting

Can be generalized to boosted top quarks (3-prong jets)

$\Rightarrow$

N-subjettiness (2010): Jet shape sensitive to N-prong structure

First applied to boosted tops

$$\tau_N(\beta) = \frac{1}{\text{norm}} \sum \text{Pt.} \min \chi^2 \triangle R_{i,j}, \triangle R_{i,j}, \ldots \triangle R_{i,N}$$

N axes chosen by some method, e.g., by minimizing $\tau_N(\beta)$

$$\tau_N(\alpha)$$
Consider a boosted top \((t \rightarrow bW \rightarrow jbjj)\)

vs. QCD jet with \(m_{jett} \approx 170\) GeV

**Top jet**

- Axes "lock-on" to correct substructure
- \(\tau_3\) is small

**QCD jet**

- In determinate substructure
- \(\tau_3\) is ambiguous

\(\tau_2\)

- Not very 2-subjet-like
- \(\tau_2\) is large

\(\tau_3\)

- Still indeterminate substructure
- \(\tau_2 \approx \tau_3\) is ambiguous
Generally, for an N-prong jet $\tau_N < \tau_{N-1}$

$\tau_N$ is a good discriminant for $N = 3$ tops

$\tau_{N-1}$ is for $N = 2, W/Z/H$.

Why does $N$-subjettiness work?

→ Boosted object already has mass, but from prongs, not from gluon radiation $\tau_N < \tau_{N-1}$

→ QCD jets get mass from energetic core with wide angle emissions $\tau_N \propto \tau_{N-1}$

Both BDRS & $N$-subjettiness used in Higgs studies

BDRS: $h \to b\bar{b}$

$\tau_2/\tau_1$: $h \to WW \to (jj)(jj)$

Heavier Higgs boosted, help identify hadronic $W$. 
(Aside: Can you combine best features of BORG and $\tilde{r}_2/\tilde{r}_1$?

**BORG**: Does not explicitly test for prongs

**$\tilde{r}_2/\tilde{r}_1$**: Does not explicitly test for symmetry

**Generalized Energy Correlation Function** (2013)

$$C_{2}^{(2)} = \frac{\sum E_i E_j E_k (\theta_{ij} \theta_{jk} \theta_{ik})^{1/5} \sum E_i}{\left( \sum E_i E_j \theta_{ij} \right)^2}$$

Combines features of BORG & $\tilde{r}_2/\tilde{r}_1$,

for 2-prong searches ($\tilde{w}/\tilde{z}/\tilde{H}$)

On-going work to analytically understand these various 2-prong methods.)
Part 4: Jet Grooming

What is a jet?

Sources of Contamination

Irreducible
- ISR (soft radiation from initial state)
- Underlying Event (secondary parton collisions)

Reducible, but huge issue for high luminosity LHC.
- Pileup (secondary hadron collisions)

Why does it matter?

\[
\frac{1}{\Delta \rho} \frac{\partial}{\partial \rho} m
\]

Overall shift (relatively easy to correct)
- Degraded mass resolution (goal of grooming methods)
- Boosted object at parton level

stuff we measure 
\rightarrow stuff we want + contamination
Many methods have been proposed.
(e.g. trimming, filtering, pruning, mass drop...)
... but I'll mention just two here.

**Area Subtraction (2008)**: Specifically for pileup.

![Diagram of area subtraction](image)

**0-th order assumption**: Pileup is uniform in \( \eta - \phi \) plane
(can refine)

- Estimate pileup density \( \rho \)
  (calculate median contamination in mini-jets)
- Add uniform spray of ghost particles
  with energy density \( \rho \) across entire event

Effectively deals with event-by-event fluctuations of pileup (but not fluctuations within an event.)
Jet Grooming Methods aim to generally remove soft, wide angle emissions.

What is soft?

Test whether $E_{trg}$ is small?

No, not IRC safe

Need to define "soft" in IRC safe way.

(Recent paper on "Modified Mass Drop", which I suspect will be somehow the future of jet grooming. I need to understand it better, so focus on a grooming method more familiar to me.)

Jet Trimming (2009)

Ordinary jet $P_{jet}^M = \sum_{i:ejet} P_i^M$

Trimmed jet $E_{jet} = \sum_{i:ejet} P_i^M w_i$ Weight

$w_i = \begin{cases} 1 & \text{if "hard"} \\ 0 & \text{if "soft"} \end{cases}$

Originally weights were defined through jet algorithm, but a bit easier to understand through analytic formula. (upcoming work)
\[ w_i = \Theta \left( \frac{\sum_{\text{jet}} \theta(\ell_j < R_{jet})}{\sum_{\text{jet}} \ell_j} > \text{fact} \right) \]

Measure energy in a region of size $R_{jet}$

Keep if energy fraction is above a threshold.

"Active" removal of jet contamination, both pileup and ISR/UE. Recent ATLAS study showing that there is approximately flat response as $N_{PU}$ increases.

Trimming is a "generic" method, agnostic as to whether the jet is quark/gluon/boosted object.

Other grooming methods optimized for boosted objects.

---

I've given you a brief tour of jet substructure.

Much more information in BOOST reports.

Shelton's TASI lectures a good place to start for more information/references.
The Future of Jet Substructure.

An active field (cf. BOOST workshops) with room for new ideas.

- How will we push the boundaries?

Experiment:

Near Term: Apply q, w.s.g., boosted objects, jet grooming to wide variety of searches;
Data driven validation methods.

Long Term: How to define trigger jets in high pileup environment.

QCD Theory:

Near Term: Analytic methods for "2-prong" triggers;
Gain understanding of differences between methods, validity of Monte Carlo

Long Term: Tackle question of impact of non-perturbative physics

BSM Theory:

Near Term: Apply to as many BSM searches as possible;
Jet substructure vs. jet superstructure (jet vs. event)

Long Term: Find jet blind spots, especially with high luminosity;
Other opportunities in boosted regime?