JET SUBSTRUCTURE: BACK TO BASICS

Simone Marzani
Institute for Particle Physics Phenomenology
Durham University

Using Jet Substructure
University of Oregon, 23rd-26th April 2013

with Mrinal Dasgupta, Alessandro Fregoso and Gavin Salam (Alex Powling)
Our current understanding

**Boost 2010 proceedings:**

The [Monte Carlo] findings discussed above indicate that while [pruning, trimming and filtering] have qualitatively similar effects, there are important differences. For our choice of parameters, pruning acts most aggressively on the signal and background followed by trimming and filtering.

- To what extent are the taggers above similar?
- How does the statement of aggressive behaviour depend on the taggers’ parameters and on the jet’s kinematics?

- *Time to go back to basics,* i.e. to understand the perturbative behaviour of QCD jets with tagging algorithms
Comparison of taggers

The “right” MC study on QCD jets can be instructive
Comparison of taggers

Different taggers appear to behave quite similarly
Comparison of taggers

But only for a limited range of masses!
Questions that arise

- Can we understand the different shapes (flatness vs peaks)?
- What’s the origin of the transition points?
- How do they depend on the taggers’ parameters?
- What’s the perturbative structure of tagged mass distributions?
- The plain jet mass contains (soft & collinear) double logs:

\[
\Sigma(\rho) \equiv \frac{1}{\sigma} \int^\rho \frac{d\sigma}{d\rho'} d\rho' \sim \sum_n \alpha_s^n \ln^{2n} \frac{1}{\rho} + \ldots
\]

- Do the taggers ameliorate this behaviour?
- If so, what’s the applicability of FO calculations?
Trimming

1. Take all particles in a jet and re-cluster them with a smaller jet radius $R_{\text{sub}} < R$

2. Keep all subjets for which $p_t^{\text{subjet}} > z_{\text{cut}} p_t$

3. Recombine the subjets to form the trimmed jet
LO calculation

- LO eikonal calculation is already useful
- Consider the emission of a gluon in soft/collinear limit
  (small $z_c$ for convenience)

$$\frac{1}{\sigma} \frac{d\sigma}{dv} = \frac{\alpha_s C_F}{\pi} \int \frac{d\theta^2}{\theta^2} \int \frac{dx}{x} \Theta (R^2 - \theta^2) \left[ \Theta (R_{\text{sub}}^2 - \theta^2) + \Theta (\theta^2 - R_{\text{sub}}^2) \Theta (x - z_c) \right] \delta (v - x\theta^2)$$

$$v = \frac{m_j^2}{p_i^2}$$
LO calculation

• LO eikonal calculation is already useful
• Consider the emission of a gluon in soft/collinear limit (small $z_c$ for convenience)

$$\frac{1}{\sigma} \frac{d\sigma}{dv} = \frac{\alpha_s C_F}{\pi} \int \frac{d\theta^2}{\theta^2} \int \frac{dx}{x} \Theta \left( R^2 - \theta^2 \right) \left[ \Theta \left( R_{\text{sub}}^2 - \theta^2 \right) + \Theta \left( \theta^2 - R_{\text{sub}}^2 \right) \Theta(x - z_c) \right] \delta \left( v - x\theta^2 \right)$$

• Three regions:
LO calculation

• LO eikonal calculation is already useful
• Consider the emission of a gluon in soft/collinear limit (small $z_c$ for convenience)

\[
\frac{1}{\sigma} \frac{d\sigma}{dv} = \frac{\alpha_s C_F}{\pi} \int \frac{d\theta^2}{\theta^2} \int \frac{dx}{x} \Theta \left( R^2 - \theta^2 \right) \left[ \Theta \left( R_{sub}^2 - \theta^2 \right) + \Theta \left( \theta^2 - R_{sub}^2 \right) \Theta(x - z_c) \right] \delta \left( v - x\theta^2 \right)
\]

• Three regions: plain jet mass, single logs, jet mass with $R_{sub}$

\[ v = \frac{m_j^2}{p_t^2} \]

Subtraction with hard collinear and finite $z_c$
Trimming: all orders

- Emissions within $R_{\text{sub}}$ are never tested for $z_{\text{cut}}$: double logs
- Intermediate region in which $z_{\text{cut}}$ is effective: single logs
- Essentially one gets exponentiation of LO (+ running coupling)
Trimming: all orders

- Emissions within $R_{\text{sub}}$ are never tested for $z_{\text{cut}}$: double logs
- Intermediate region in which $z_{\text{cut}}$ is effective: single logs
- Essentially one gets exponentiation of LO (+ running coupling)

All-order calculation done in the small-$z_{\text{cut}}$ limit
Trimming: all orders

- Emissions within $R_{\text{sub}}$ are never tested for $z_{\text{cut}}$: double logs
- Intermediate region in which $z_{\text{cut}}$ is effective: single logs
- Essentially one gets exponentiation of LO (+ running coupling)

All-order calculation done in the small-$z_{\text{cut}}$ limit
Trimming: all orders

- Emissions within $R_{\text{sub}}$ are never tested for $z_{\text{cut}}$: double logs
- Intermediate region in which $z_{\text{cut}}$ is effective: single logs
- Essentially one gets exponentiation of LO (+ running coupling)

- Our calculation captures $\alpha_s^n L^{2n}$ and $\alpha_s^n L^{2n-1}$ in the expansion
- To go beyond that one faces the usual troubles: non-global logs, clustering effects, etc.
- The transition points are correctly identified by the calculations
- The shapes are understood
Pruning

1. From an initial jet define pruning radius $R_{\text{prune}} \sim m / p_t$

2. Re-cluster the jet, vetoing recombination for which

$$z = \frac{\min(p_{ti}, p_{ti})}{|\vec{p}_{ti} + \vec{p}_{ti}|} < z_{\text{cut}}$$

$$d_{ij} > R_{\text{prune}}$$

i.e. soft and wide angle
LO calculation

- LO calculation similar to trimming
- Now the pruning radius is set dynamically $R_{\text{prune}}^2 \sim x\theta^2$

\[
\frac{1}{\sigma} \frac{d\sigma}{dv} = \frac{\alpha_s C_F}{\pi} \int \frac{d\theta^2}{\theta^2} \frac{dx}{x} \Theta (R^2 - \theta^2) \left[ \Theta (R_{\text{prune}}^2 - \theta^2) + \Theta (\theta^2 - R_{\text{prune}}^2) \Theta(x - z_{\text{cut}}) \right] \delta (v - x\theta^2)
\]

\[
v = \frac{m_j^2}{p_t^2}
\]
LO calculation

- LO calculation similar to trimming
- Now the pruning radius is set dynamically: $R_{\text{prune}}^2 \sim x\theta^2$

\[
\frac{d\sigma}{\sigma dv} = \frac{\alpha_s C_F}{\pi} \int \frac{d\theta^2}{\theta^2} \frac{dx}{x} \Theta (R^2 - \theta^2) \left[ \Theta (R_{\text{prune}}^2 - \theta^2) + \Theta (\theta^2 - R_{\text{prune}}^2) \Theta (x - z_{\text{cut}}) \right] \delta (v - x\theta^2)
\]

- Two regions:

![Graph showing the coefficient of $C_F\alpha_s/\pi$ for pruning $R=0.8$, $z_{\text{cut}}=0.1$. The graph displays a transition point and an analytic curve labeled as Event2.](attachment:graph.png)
LO calculation

- LO calculation similar to trimming
- Now the pruning radius is set dynamically \( R_{\text{prune}}^2 \approx x\theta^2 \)

\[
\frac{1}{\sigma} \frac{d\sigma}{dv} = \frac{\alpha_s C_F}{\pi} \int \frac{d\theta^2}{\theta^2} \frac{dx}{x} \Theta(R^2 - \theta^2) \left[ \Theta(R_{\text{prune}}^2 - \theta^2) + \Theta(\theta^2 - R_{\text{prune}}^2) \Theta(x - z_{\text{cut}}) \right] \delta(v - x\theta^2)
\]

- Two regions: plain jet mass and single-log region

\[
v = \frac{m_j^2}{p_t^2}
\]

transition point

leading behaviour in each region

Subtraction with hard collinear and finite \( z_c \)
What pruning is meant to do

Choose an $R_{\text{prune}}$ such that different hard prongs ($p_1, p_2$) end up in different hard subjets.
Discard any softer radiation.
What pruning is meant to do
Choose an $R_{\text{prune}}$ such that different hard prongs ($p_1$, $p_2$) end up in different hard subjets.
Discard any softer radiation.

What pruning sometimes does
Chooses $R_{\text{prune}}$ based on a soft $p_3$ (dominates total jet mass), and leads to a single narrow subjet whose mass is also dominated by a soft emission ($p_2$, within $R_{\text{prune}}$ of $p_1$, so not pruned away).
Beyond LO

What pruning is meant to do
Choose an $R_{\text{prune}}$ such that different hard prongs ($p_1, p_2$) end up in different hard subjets.
Discard any softer radiation.

What pruning sometimes does
Chooses $R_{\text{prune}}$ based on a soft $p_3$ (dominates total jet mass), and leads to a single narrow subjet whose mass is also dominated by a soft emission ($p_2$, within $R_{\text{prune}}$ of $p_1$, so not pruned away).
Structure beyond LO

• Because of its anomalous component the logarithmic structure at NLO worsens: \( \sim \alpha_s^2 L^4 \) (as plain jet mass)
• Explicit calculation shows that the anomalous component is active for \( \rho < z_{\text{cut}}^2 \)
• A simple fix: require at least one successful merging with \( \Delta R > R_{\text{prune}} \) and \( z > z_{\text{cut}} \) (sane pruning)

• It is convenient to resum the two components separately
• Sane pruning: essentially Sudakov suppression of LO \( \sim \alpha_s^n L^{2n-1} \)
• Anomalous pruning: more complicated convolution structure, which leads to \( \sim \alpha_s^n L^{2n} \)
All-order results

- Full Pruning: single-log region for $z_{\text{cut}}^2 < \rho < z_{\text{cut}}$
- We control $\alpha_s^n L^{2n}$ and $\alpha_s^n L^{2n-1}$ in the expansion
- NG logs present but parametrically reduced
All-order results

- Full Pruning: single-log region for $z_{\text{cut}}^2 < \rho < z_{\text{cut}}$
- We control $\alpha_s^n L^{2n}$ and $\alpha_s^n L^{2n-1}$ in the expansion
- NG logs present but parametrically reduced

**Analytic Calculation: quark jets**

$m$ [GeV], for $p_t = 3$ TeV, $R = 1$

**Pythia 6 MC: quark jets**

$m$ [GeV], for $p_t = 3$ TeV, $R = 1$

All-order calculation done in the small-$z_{\text{cut}}$ limit
All-order results

- Full Pruning: single-log region for $z_{cut}^2 < \rho < z_{cut}$
- We control $\alpha_s^n L^{2n}$ and $\alpha_s^n L^{2n-1}$ in the expansion
- NG logs present but parametrically reduced

All-order calculation done in the small-$z_{cut}$ limit
1. Undo the last stage of the C/A clustering. Label the two subjets $j_1$ and $j_2$ ($m_1 > m_2$)

2. If $m_1 < \mu m$ (mass drop) and the splitting was not too asymmetric ($y_{ij} > y_{cut}$), tag the jet.

3. Otherwise redefine $j = j_1$ and iterate.
Mass Drop Tagger at LO

1. Undo the last stage of the C/A clustering. Label the two subjets $j_1$ and $j_2$ ($m_1 > m_2$)

2. If $m_1 < \mu m$ (mass drop) and the splitting was not too asymmetric ($y_{ij} > y_{cut}$), tag the jet.

3. Otherwise redefine $j = j_1$ and iterate.

In the small-$y_{cut}$ limit the result is identical to LO pruning: single-log distribution

![Graph showing the coefficient of $C_F \alpha_s / \pi$ for mass-drop $R=0.8, y_{cut}=0.1$. The graph compares the analytic results with hard collinear and finite $y_c$. The subtraction occurs at $y_{cut}$.](image_url)
Problems beyond LO

What MDT does wrong:
If the $y_{ij}$ condition fails, MDT iterates on
the more massive subjet. It can follow a
soft branch ($p_2 + p_3 < y_{cut} p_{tjet}$), when the
“right” answer was that the (massless)
hard branch had no substructure

- This can be considered a flaw of the tagger
- It worsens the logarithmic structure $\sim \alpha_s^2 L^3$
- It makes all-order treatment difficult
- It calls for a modification
Modified Mass Drop Tagger

1. Undo the last stage of the C/A clustering. Label the two subjets $j_1$ and $j_2$ ($m_1 > m_2$)
2. If $m_1 < \mu m$ (mass drop) and the splitting was not too asymmetric ($y_{ij} > y_{cut}$), tag the jet.
3. Otherwise redefine $j$ to be the subjet with highest transverse mass and iterate.

- In practice the soft-branch contribution is very small
- However, this modification makes the all-order structure particularly interesting
All-order structure of mMDT

• The mMDT has single logs to all orders (i.e. $\sim \alpha_s^n L^n$)
• In the small $\gamma_{\text{cut}}$ limit it is just the exponentiation of LO
• Beyond that flavour mixing can happen (under control)
All-order structure of mMDT

- The mMDT has single logs to all orders (i.e. $\sim \alpha_s^n L^n$)
- In the small $y_{cut}$ limit it is just the exponentiation of LO
- Beyond that flavour mixing can happen (under control)
All-order structure of mMDT

- The mMDT has single logs to all orders (i.e. $\sim \alpha_s^n L^n$)
- In the small $y_{\text{cut}}$ limit it is just the exponentiation of LO
- Beyond that flavour mixing can happen (under control)
Properties of mMDT

- **Flatness** of the background is a desirable property (data-driven analysis)
- $\gamma_{cut}$ can be adjusted to obtain it (analytic relation)
- FO calculation might be applicable
- **Role of $\mu$,** not mentioned so far
- It contributes to subleading logs and has small impact if not too small ($\mu > 0.4$)
- **Filtering** only affects subleading terms
- It has only single logs, which are of collinear origin
Properties of mMDT

- **Flatness** of the background is a desirable property (data-driven analysis)
- $\gamma_{\text{cut}}$ can be adjusted to obtain it (analytic relation)
- FO calculation might be applicable
- **Role of $\mu$**, not mentioned so far
- It contributes to subleading logs and has small impact if not too small ($\mu > 0.4$)
- **Filtering** only affects subleading terms
- It has only **single logs**, which are of collinear origin
- Important consequence:

**mMDT is FREE of non-global logs!**
In summary ...

- Analytic studies of the taggers reveal their properties
- Particularly useful if MCs don’t agree
- They also lead to the design of better taggers
- Sane pruning can be an example (but need further tests)
- mMDT is remarkable: single-jet observable free of non global logs

- We’ve also investigated aspects of NP effects (not presented here)

- Future work will involve looking at signal-jets as well
BACK UP SLIDES
Examples of NLO checks

Coefficient of $(C_F \alpha_s/\pi)^2$ for modified mass-drop $R=0.8$, $y_{cut}=0.1$

Coefficient of $(C_F \alpha_s/\pi)^2$ for pruning $R=0.8$, $z_{cut}=0.4$

Coefficient of $C_F C_A (\alpha_s/\pi)^2$ for modified mass-drop $R=0.8$, $y_{cut}=0.2$

Coefficient of $(C_F \alpha_s/\pi)^2$ for trimming $R=0.8$, $R_{sub}=0.2$, $z_{cut}=0.15$
Effect of filtering: quark jets
$m$ [GeV], for $p_t = 3$ TeV, $R = 1$
$m_{\text{MDT}} (y_{\text{cut}}=0.13)$

Effect of $\mu$ parameter: quark jets
$m$ [GeV], for $p_t = 3$ TeV, $R = 1$
$\mu = 1.00$
$\mu = 0.67$
$\mu = 0.40$
$\mu = 0.30$
$\mu = 0.20$
Sometimes MCs don’t agree
Analytics can tell you which one is right