QCD CALCULATIONS FOR JET SUBSTRUCTURE

Simone Marzani
Institute for Particle Physics Phenomenology
Durham University

RADCOR 2013
11th International Symposium on Radiative Corrections
(Application of Quantum Field Theory to Phenomenology)
22nd-27th September 2013

with
M. Dasgupta, A. Fregoso, G. P. Salam and A. Powling
arXiv:1307.0007 and 1307.0013
The boosted regime

- The LHC is exploring phenomena at energies above the EW scale
- Z/W/H/top can no longer be considered heavy particles
- These particles are abundantly produced with a large boost
- Their hadron decays are collimated and can be reconstructed within a single jet. Need to distinguish:

\[ p_t \gg m \]
Grooming and tagging

• The last few years have seen a rapid development in substructure techniques: O(10-20) powerful methods to tag jet substructure

• Many of the methods have been tried out in searches and work; they will be crucial for searches in the years to come

• Many methods can lead to some confusion
• Do we understand how / why they work?
• Only analytic understanding can give this field robustness
Where to start?

Cannot possibly study all tools
These 3 are widely used
We concentrate on background (QCD jets)

Trimming

- recluster on scale \( R_{\text{sub}} \)
- discard subjets with \( < z_{\text{cut}} p_t \)

Krohn, Thaler and Wang (2010)

Pruning

- recluster \( R_{\text{prune}} \sim m_j/p_t \)
- discard large-angle soft clusterings

Ellis, Vermillion and Walsh (2009)

Mass-drop tagger (MDT, aka BDRS)

- decluster & discard soft junk
- repeat until find hard struct

Butterworth, Davison, Rubin and Salam (2008)
Our understanding so far

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?

• The “right” MC study can be instructive
Comparison of taggers

Plain jet mass: characteristic Sudakov peak
Comparison of taggers

Different taggers appear to behave quite similarly
Comparison of taggers

But only for a limited kinematic region!
Comparison of taggers

Let’s translate from QCD variables to ``search'' variables:

\[ \rho \rightarrow m, \text{ for } p_t = 3\text{ TeV}, R = 1.0 \]
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?
• Cumulative distribution for plain jet mass contains (soft & collinear) double logs

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

• Do the taggers ameliorate this behaviour?
• If so, what's the applicability of FO calculations?
Trimming at LO

\[ \frac{\rho}{\sigma} \frac{d\sigma^{(\text{trim,LO})}}{d\rho} = \frac{\alpha_s C_F}{\pi} \left( \ln \frac{r^2}{\rho} - \frac{3}{4} \right) \]

Intermediate region in which \( z_{\text{cut}} \) is effective: single logs

\[ \frac{\alpha_s C_F}{\pi} \left( \ln \frac{1}{z_{\text{cut}}} - \frac{3}{4} \right) \]

Emissions within \( R_{\text{sub}} \) are never tested for \( z_{\text{cut}} \): double logs

\[ r = \frac{R_{\text{sub}}}{R} \]

\[ \rho = \frac{m^2}{(p_t^2 R^2)} \]

\( m \) [GeV], for \( p_t = 3 \) TeV, \( R=1 \)

trimmed quark jets: LO
Trimming: all orders

One gets exponentiation of LO (+ running coupling)

\[
\frac{d\sigma^{\text{trim, resum}}}{d\rho} = \frac{d\sigma^{\text{trim, LO}}}{d\rho} \exp \left[-\int_\rho d\rho' \frac{1}{\sigma} \frac{d\sigma^{\text{trim, LO}}}{d\rho'}\right]
\]

Pythia 6 MC: quark jets
m [GeV], for \(p_t = 3\) TeV, \(R = 1\)

Trimming
\(R_{\text{sub}} = 0.2, z_{\text{cut}} = 0.05\)
\(R_{\text{sub}} = 0.2, z_{\text{cut}} = 0.1\)

Analytic Calculation: quark jets
m [GeV], for \(p_t = 3\) TeV, \(R = 1\)

All-order calculation done in the small-\(z_{\text{cut}}\) limit
Pruning & MDT at LO

- The pruning radius is set dynamically: $R_{\text{prune}} < d_{ij}$
- The 2 prongs are always tested for $z_{\text{cut}}$: single logs

MDT result is identical to LO pruning (in the small-$\gamma_{\text{cut}}$ limit)
Structures beyond LO

All-order MDT and pruning distributions are NOT given by exponentiation of LO

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).

What MDT does wrong
If the energy condition fails, MDT iterates on the more massive subjet. It can follow a soft branch ($p_2 + p_3 < y_{\text{cut}} p_{\text{jet}}$), when the “right” answer was that the (massless) hard branch had no substructure.
The modified Mass Drop Tagger

- The soft-branch issue 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: always follow the subjet with highest transverse mass

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

- In the small $y_{\text{cut}}$ limit, it is just the exponentiation of LO
- The mMDT has single logs to all orders (i.e. $\alpha_s^n \ln^n$)

Remarkable agreement!
Interesting feature: flat mass distribution (more in backup slides)
All-order structure of mMDT

- In the small $y_{\text{cut}}$ limit, it is just the exponentiation of LO
- The mMDT has single logs to all orders (i.e. $\alpha_s^n L^n$)

- Single logs: extended validity of FO calculations

LO v. NLO v. resummation (quark jets)

$m$ [GeV], for $p_t = 3\text{ TeV}, R = 1$

- $10$
- $100$
- $1000$

$\rho / d\sigma / dp$

$p_{t,\text{jet}} > 3\text{ TeV}$

mMDT ($y_{\text{cut}} = 0.13$)

Leading Order
Next-to-Leading Order
Resummed

only a qualitative comparison: different process/kinematics!
All-order structure of mMDT

- In the small $y_{\text{cut}}$ limit, it is just the exponentiation of LO
- The mMDT has single logs to all orders (i.e. $\alpha_s^n L^n$)

- Single logs: extended validity of FO calculations
- Single logs of collinear origin
- Remarkable consequence: mMDT is free of non-global logs!
Pruning: I & Y components

- Pruning @ NLO $\sim \alpha_s^2 L^4$ (like plain jet mass)
- Single-pronged component (I-pruning) is active for $\rho < z_{cut}^2$
- A simple modification: require at least one successful merging with $\Delta R > R_{prune}$ and $z > z_{cut}$ (Y-pruning)

- It is convenient to resum the two components separately
- Y-pruning: essentially Sudakov suppression of LO $\sim \alpha_s^n L^{2n-1}$
- I-pruning: convolution between the pruned and the original mass $\sim \alpha_s^n L^{2n}$
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 deferred to NNLO

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

- Most taggers have reduced sensitivity to NP physics
- mMDT particularly so (it's the most calculable)
- Y-pruning sensitive to UE because of the role played by the fat jet mass
Performances for finding signals (Ws)

Y-pruning gives a visible improvement
In summary ...

- Analytic studies of the taggers reveal their properties
- Particularly useful if MCs don’t agree
In summary ...

- Analytic studies of the taggers reveal their properties.
- Particularly useful if MCs don’t agree.
In summary ...

• Analytic studies of the taggers reveal their properties
• Particularly useful if MCs don’t agree

• They also indicate how to develop better taggers

• **Y-pruning:**
  • improved log behaviour wrt pruning ($\alpha_s^n L^{2n-1} vs \alpha_s^n L^{2n}$)
  • better rejection of QCD background

• **mMDT:**
  • exceptionally simple structure (single logs, no non-global)
  • reduced sensitivity to non-perturbative physics
BACKUP SLIDES
of Y-pruning, namely that at high Sudakov suppression. Despite this apparent advantage, one drawback.

exploits the same double-logarithmic background suppress mass and the resulting pruning radius. It remains of interest by underlying event and pileup, because of the way in which the tagging

It would also, of course, be interesting to extend our analysis of observables specifically designed to show sensitivity to colour factor in the underlying event and pileup, because of the way in which the tagging

p

the background suppression dominates, leading to improved performance of pruning relative to mMDT is mitigated. Most interesting, perhaps, is the fact that the non-perturbative (NP) region starts, with which the non-perturbative (NP) region starts, with

\( \mu_{\text{NP}} p_t R \)

\( \mu_{\text{NP}} p_t R_{\text{sub}} \)

\( \mu_{\text{NP}} p_t R \)

\( \mu_{\text{NP}} p_t R \)

\( \mu_{\text{NP}} p_t R \)

\( \mu_{\text{NP}}^2 / \rho_c \)

\( \langle \text{Sudakov tail} \rangle \)

\( \text{yes} \)

\( \text{no} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)

\( \text{yes} \)
Lund diagrams for mMDT
Lund diagrams for pruning

\[ E_{\text{jet}} = z_{\text{cut}} \]

\[ \rho > z_{\text{cut}} \]

\[ z_{\text{cut}} < \rho < z_{\text{cut}} \]

\[ \rho < z_{\text{cut}} \]

Pruning
Hadronisation effects for mMDT

Hadronisation produces:
1. a shift in the squared jet mass
2. a shift in the jet’s (or prong’s) momentum

Same power behaviour but with competing signs:

\[
\frac{d\sigma^{\text{NP}}}{dm} \approx \frac{d\sigma^{\text{PT}}}{dm} \left[ 1 + a \frac{\Lambda_{\text{NP}}}{m} \right]
\]
Examples of NLO checks

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

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

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

Coefficient of $(C_F \alpha_s/n)^2$ for trimming $R=0.8$, $R_{sub}=0.2$, $z_{cut}=0.15$
Other properties of mMDT

- **Flatness** of the background is a desirable property (data-driven analysis, side bands)
- $\gamma_{\text{cut}}$ can be adjusted to obtain it (analytic relation)
- **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 ($N_{\text{filt}}^{\text{LL}}$) terms
• ATLAS measured the jet mass with MDT
• Different version of the tagger with $R_{\text{min}}=0.3$ between the prongs
• ATLAS measured the jet mass with MDT
• Different version of the tagger with \( R_{\text{min}} = 0.3 \) between the prongs

• This cut significantly changes the tagger’s behaviour: mass minimum
• The single-log region is reduced (and can even disappear)
• We hope that future studies will be able to avoid this