QJets

Tuhin S. Roy
Los Alamos National Laboratory
Outline

- **Introduction**
  - The central idea behind Qjets.
  - How Qjets can be used with Pruning
  - The Qjets clustering algorithm

- **QJets + Pruning**
  - Volatility and its uses
    - results of ATLAS studies.
  - Statistics in the context of Qjets
    - Applications in stabilizing statistical fluctuations in jet measurements.
QJets: The Idea

An event → calorimeter cells

Particle flow → traditional jet clustering algorithm such as C/A, $k_T$

Jets → jet-observables
QJets: The Idea

An event ➔ calorimeter cells ➔ particle flow ➔ traditional jet clustering algorithm such as C/A, $k_T$ ➔ jets

Graph: Fraction of jets per bin vs. $m_J$ [GeV]
QJets: The Idea

An event → calorimeter cells

particle flow → traditional jet clustering algorithm such as C/A, $k_T$

jets → variations of observables as new observables
Qjets is an idea that explores the dimension of clustering history.
Qjets is an idea that explores the dimension of clustering history. It is a challenging task -- let us start with something simpler.
QJets: The Idea

Qjets is an idea that explores the dimension of clustering history

Start with the constituents of a jet → jet algorithm

k_T

C/A

answers to a question that depends on clustering history would be different

it is a challenging task -- let us start with something simpler
Qjets give different answers to a question that depends on clustering history

ex: pruning
Pruning

Start with the constituents of a given jet and rebuild the jet along C/A or $k_T$
Pruning

At every step of clustering check whether the branch to be added is soft and wide angled.

- if yes discard the softer four-vector.

soft if: \( \frac{\min(p_{T_i}, p_{T_j})}{|p_{T_i} + p_{T_j}|} < z_{\text{cut}} \)

wide-angled if: \( \Delta R_{ij} > D_{\text{cut}} \)
Pruning

Pruned Jet
Pruning

- Four-vectors that are pruned are actually branches of the tree.

- Pruned jets depend crucially on the tree-structure or the clustering algorithm used to construct the jet.
As in a sequential recombination algorithm, assign every pair of four-vectors a distance measure \( d_{ij} \).

However, unlike a normal sequential algorithm (where the pair with the smallest measure is clustered), here a given pair is randomly selected for merging with probability

\[
\Omega_{ij} = \frac{1}{N} \exp\left(-\alpha \frac{d_{ij}}{d_{\text{min}}}ight)
\]

where \( d_{\text{min}} = \text{Min}\{d_{ij}\} \) is the minimum distance between any pair of vectors.
\[ \Omega_{ij} = \frac{1}{N} \exp \left( -\alpha \frac{d_{ij}}{d_{\text{min}}} \right) \]

d_{ij} : we take C/A or kT measure

\( \alpha \to \infty \) Classical regime: only path corresponding to \( d_{\text{min}} \) is selected

\( \alpha > 0 \) physical regime: physical paths are preferred

\( \alpha \to 0 \) democratic regime: all paths have same weight

\( \alpha < 0 \) unphysical regime: physical paths are de-weighted
**QJets + Pruning**

A collection of 4 vectors

just Pruning

one pruned jet

one pruned jetmass

Qjets + Pruning

N pruned Qjets

N pruned jetmasses
QJets + Pruning

Ex. W decaying to two quarks

The original jet is made using anti-kT algorithm with $R = 0.7$ and $p_T > 500\text{GeV}$
QJets + Pruning

Ex. W decaying to two quarks

The original jet is made using anti-kT algorithm with $R = 0.7$ and $pT > 500\text{GeV}$

How can this distribution be used?
Outline

- **Introduction**
  - The central idea behind Qjets.
  - How Qjets can be used with Pruning
  - The Qjets clustering algorithm

- **QJets + Pruning**
  - Volatility and its uses
    - results of ATLAS studies.
  - Statistics in the context of Qjets
    - Applications in stabilizing statistical fluctuations in jet measurements.
QPruning vs. Pruning

Let us take a sample jet

How can this distribution be used?

- Simply use the shape of the distribution to discriminate signal from background.

Application in signal discovery

- Use the distribution to reduce statistical fluctuations in measurements.

Application in determination of cross-section, mass etc.
Volatility of a jet

\[ \nu = \frac{\omega_p}{m_p} \]

\( \omega_p \) = width of jet mass distribution
\( m_p \) = averaged pruned jet mass
Performance and Validation of Q-jets at the ATLAS Detector in pp Collisions at $\sqrt{s} = 8$ TeV in 2012

The ATLAS Collaboration

Abstract

The Q-jets technique introduces the idea of interpreting jets through multiple sets of possible showering histories. This approach allows jet observables, such as the jet mass, to be evaluated not simply as single values, but rather as distributions. The resulting distributions can be interpreted statistically to form new observables, allowing the separation of boosted, hadronically-decaying particles from light quark and gluon backgrounds. We present a study of Q-jets in boosted, hadronically-decaying $W$ boson and dijet samples, demonstrating the discriminating power of this technique. Different Q-jet parameters and observables are studied, and an optimal configuration based on physics performance and computational efficiency is proposed, leading to a factor of 15 in dijet rejection at a 50% efficiency for jets from boosted, hadronically decaying $W$ bosons. The impact of pile-up on the performance of this method is tested up to an average of 40 additional interactions per event and found to be weak. A performance comparison between the Q-jets algorithm and $N$-subjettiness, a previously measured substructure observable which determines the compatibility of a jet with the $N$-subjet hypothesis, is presented.

Copyright 2013 CERN for the benefit of the ATLAS Collaboration. Reproduction of this article or parts of it is allowed as specified in the CC-BY-3.0 license.
Volatility of a jet: experimental results

Summary:

- The volatility distribution for QCD jets and W-jets have been reproduced.
- The optimized separation is obtained for $\alpha = 0.1$ and $N_{Qjets} > 25$.
- Volatility distribution does not have strong dependence on Pile-up.
- Good data/MC agreement.
- 15 QCD jet rejection at 50% W-jet tagging efficiency.
- Comparable performance with N-subjettiness.
- Volatility and N-subjettiness are weakly (ATLAS)/strongly (CMS) correlated!
**Volatility of a jet:** experimental results

**Summary:**

- The volatility distribution for QCD jets and W-jets have been reproduced.

- The optimized separation is obtained for $\alpha=0.1$ and $N_{Q\text{jets}} > 25$

- Volatility distribution does not have strong dependence on Pile-up

- $W$ vs. Dijets

- $C/A$ Pruning, $N_{Q\text{jets}} = 75$

- $W$ jets and dijets

- $\alpha^2$ and $D_{\text{subjettiness}}$

- $\text{W jets}$ and dijets

- $N_{Q\text{jets}}$ are comparable in performance with $N_{\text{subjettiness}}$

- $\text{Wjets}$ and dijets and (b) the significance of the volatility

- The significance is defined as the ratio of the difference between the mean dijet and mean $W$-jet selection volatilities to the sum in quadrature of the respective RMS values.

- Volatility of a jet: $W$-jets and dijets and (b) the significance of the volatility

- The optimized separation in mean and optimal significance is observed at $\alpha = 0.1$ and $N_{Q\text{jets}} = 75$

- The volatility distribution for QCD jets and W-jets have been reproduced.

- The optimized separation is obtained for $\alpha=0.1$ and $N_{Q\text{jets}} > 25$

- Volatility distribution does not have strong dependence on Pile-up

- $W$ vs. Dijets

- $C/A$ Pruning, $N_{Q\text{jets}} = 75$

- $W$ jets and dijets

- $\alpha^2$ and $D_{\text{subjettiness}}$

- $\text{Wjets}$ and dijets

- $N_{Q\text{jets}}$ are comparable in performance with $N_{\text{subjettiness}}$
**Volutility of a jet:** experimental results

**Summary:**

- The volatility distribution for QCD jets and W-jets have been reproduced.
- The optimized separation is obtained for $\alpha=0.1$ and $N_{Qjets} > 25$
- Volatility distribution does not have strong dependence on Pile-up
- Good data/MC agreement
- 15 QCD jet rejection at 50% W-jet tagging efficiency
- Comparable performance with N-subjettiness
- Volatility and N-subjettiness are weakly (ATLAS)/strongly (CMS) correlated!
Summary:

- The volatility distribution for QCD jets and W-jets have been reproduced.
- The optimized separation is obtained for $\alpha = 0.1$ and $N_{Qjets} > 25$.
- Volatility distribution does not have strong dependence on Pile-up.
- Good data/MC agreement.
- 15 QCD jet rejection at 50% W-jet tagging efficiency.
- Comparable performance with N-subjettiness.
- Volatility and N-subjettiness are weakly (ATLAS) / strongly (CMS) correlated!
**Volatility of a jet: experimental results**

**Summary:**

- The volatility distribution for QCD jets and W-jets have been reproduced.
- The optimized separation is obtained for $\alpha = 0.1$ and $N_{Qjets} > 25$.
- Volatility distribution does not have strong dependence on pile-up.
- Good data/MC agreement.
- 15 QCD jet rejection at 50% W-jet tag efficiency.
- Comparable performance with N-subjettiness.
- Volatility and N-subjettiness are weakly correlated -- suggests useful potential combination.

![Diagram showing volatility and W-jet efficiency](image-url)
Summary:

- The volatility distribution for QCD jets and W-jets have been reproduced.
- The optimized separation is obtained for $\alpha = 0.1$ and $N_{\text{Qjets}} > 25$.
- Volatility distribution does not have strong dependence on Pile-up.
- Good data/MC agreement.
- 15 QCD jet rejection at 50% W-jet tag efficiency.
- Comparable performance with N-subjettiness.
- Volatility and N-subjettiness are weakly (ATLAS)/strongly (CMS) correlated!
Summary:

- The volatility distribution for QCD jets and W-jets have been reproduced.
- The optimized separation is obtained for $\alpha = 0.1$ and $N_{Qjets} > 25$.
- Volatility distribution does not have strong dependence on Pile-up.
- Good data/MC agreement.
- 15 QCD jet rejection at 50% W-jet tag efficiency.
- Comparable performance with N-subjettiness.
- Volatility and N-subjettiness are weakly (ATLAS)/ strongly (CMS) correlated!
QPruning vs. Pruning

Let us take a sample jet

How can this distribution be used?

Simply use the shape of the distribution to discriminate signal from background

Use the distribution to reduce statistical fluctuations in measurements

Application in determination of cross-section, mass etc.

Application in signal discovery
QPruning vs. Pruning

Consider candidates for a W jet

- **QPruned jetmass distribution**
- **Mass window for W**
  - 70 GeV
  - 90 GeV

**Classical pruned jetmass**

- Pruned mass is either in or out of the bin
- Tagging efficiency is either 0 or 1

- Tagging efficiency is a number between 0 to 1
QPruning vs. Pruning

Consider candidates for a W jet

Mass window for W

Pruning -> QPruning

A transition from a discrete (binomial distribution) to a continuous distribution
Statistical aspects of QJets

Consider a master set of jets

In an experiment

- pick N jets at random out of the master set
- Qjet+prune every jet and determine
  1. tagging efficiency (τ_j)
  2. An observable averaged over all qjets (q_j)
     ex. average mass of the distribution (μ_j)

Experimental Observables:

\[ N_{\text{expt}} = \sum_j \tau_j \]
\[ Q_{\text{expt}} = \frac{\sum_j \tau_j q_j}{\sum_j \tau_j} \]

we want to determine δN_{\text{expt}} and δQ_{\text{expt}}
**Statistical aspects of QJets**

we want to determine $\delta N_{\text{expt}}$ and $\delta Q_{\text{expt}}$

In an experiment

- pick $N$ jets at random out of the master set
- $Q$jet+prune every jet and determine
  1. tagging efficiency $(\tau_j)$
  2. An observable averaged over all qjets $(q_j)$
     ex. average mass of the distribution $(\mu_j)$

**Standard Procedure:**

- repeat the experiment many times and calculate $\delta N_{\text{expt}}$ and $\delta Q_{\text{expt}}$

one can derive $\delta N_{\text{expt}}$ and $\delta Q_{\text{expt}}$ analytically
Statistical aspects of QJets

one can derive $\delta N_{\text{expt}}$ and $\delta Q_{\text{expt}}$ analytically

$$\delta N_{\text{expt}} = \sqrt{N_{\text{expt}}} \times \sqrt{\langle \tau \rangle + \frac{\text{var}(\tau)}{\langle \tau \rangle}}$$

$$\delta Q_{\text{expt}} = \frac{\langle Q_{\text{expt}} \rangle}{\sqrt{N}} \times \sqrt{\frac{\text{var}(\tau)}{\langle \tau \rangle^2} + \frac{\text{var}(q\tau)}{\langle q\tau \rangle^2} - 2 \frac{\text{cov}(\tau, q\tau)}{\langle \tau \rangle \langle q\tau \rangle}} + O\left(\frac{1}{N}\right)$$
Statistical aspects of QJets

one can derive $\delta N_{\text{expt}}$ and $\delta Q_{\text{expt}}$ analytically

$$\delta N_{\text{expt}} = \sqrt{N_{\text{expt}}} \times \sqrt{\langle \tau \rangle + \frac{\text{var}(\tau)}{\langle \tau \rangle}}$$

$$\sqrt{\langle \tau \rangle + \frac{\text{var}(\tau)}{\langle \tau \rangle}} \leq 1 \text{ for all distributions}$$

$$= 1 \text{ for binomial distributions (i.e. for pruning)}$$
Statistical aspects of QJets

one can derive $\delta N_{\text{expt}}$ and $\delta Q_{\text{expt}}$ analytically

$$\delta N_{\text{expt}} = \sqrt{N_{\text{expt}}} \times \sqrt{\langle \tau \rangle + \frac{\text{var}(\tau)}{\langle \tau \rangle}}$$

![Graph showing the relationship between $\langle \tau \rangle$ and $\frac{\text{var}(\tau)}{\langle \tau \rangle}$ for different types of jets.]

- Just pruning
- W jets: Qjets + pruning
- QCD jets: Qjets + pruning
**Statistical aspects of QJets**

One can derive $\delta N_{\text{expt}}$ and $\delta Q_{\text{expt}}$ analytically.

$$
\delta Q_{\text{expt}} = \frac{\langle Q_{\text{expt}} \rangle}{\sqrt{N}} \times \sqrt{\frac{\text{var}(\tau)}{\langle \tau \rangle^2} + \frac{\text{var}(q\tau)}{\langle q\tau \rangle^2} - 2 \frac{\text{cov}(\tau, q\tau)}{\langle \tau \rangle \langle q\tau \rangle}} + O\left(\frac{1}{N}\right)
$$

However, depending on the algorithm $\delta Q_{\text{expt}}$ can go either way.

\[ \sqrt{\frac{\text{var}\left(q\tau\right)_0}{\langle q\tau \rangle^2_0} - \frac{\text{var}\left(\tau\right)_0}{\langle \tau \rangle^2_0}} \]
For numerical example, we use $q = \mu = \text{average jetmass in the bin}$

one can derive $\delta N_{\text{expt}}$ and $\delta Q_{\text{expt}}$ analytically

$$
\delta Q_{\text{expt}} = \frac{\langle Q_{\text{expt}} \rangle}{\sqrt{N}} \times \sqrt{\frac{\text{var}(\tau)}{\langle \tau \rangle^2} + \frac{\text{var}(q\tau)}{\langle q\tau \rangle^2} - 2 \frac{\text{cov}(\tau, q\tau)}{\langle \tau \rangle \langle q\tau \rangle}} + O\left(\frac{1}{N}\right)
$$
Conclusion

Grooming tools (pruning, trimming, filtering) even though designed for boosted search, are useful and essential for non-boosted cases.

We introduced QJets: a non-deterministic jet clustering algorithm.

- QJets Clustering lets us look inside a jet in a new way.
- QJets + pruning renders stability to jet observables and provides new discriminants for the discovery of signal jets.

- Q-jets are a new way to interpret jets: focus on multiple possible clustering histories, motivated by non-invertibility of parton shower
  - The first time such an idea is being considered!

- Just the tip of the iceberg: volatility is the first application of Q-jets at ATLAS– looking forward to seeing more!

Maximilian Swiatlowski, for the ATLAS Collaboration