Fuzzy Jets

A new class of jet algorithms that use soft assignments during clustering

Benjamin Nachman
SLAC, Stanford University

Paper in Preparation

In collaboration with
Lester Mackey
Ariel Schwartzman
Conrad Stansbury
Jet Clustering as an Unsupervised Learning Task

Fundamental question: *which particles belong together?*

The state-of-the-art in organizing hadronic final states at the LHC is **hierarchical agglomerative** clustering:

**hierarchical**: hereditary structure to the final state classification  
**agglomerative**: every object starts as a cluster and is sequentially merged

These schemes require metrics on momenta  
\[ d_{ij} : \mathbb{R}^4 \times \mathbb{R}^4 \rightarrow \mathbb{R}^+, \quad d_{iB} : \mathbb{R}^4 \rightarrow \mathbb{R}^+ \]
and proceed as follows:

1. Assign each particle as a proto-jet.
2. Repeat until there are no proto-jets left: Let  
   \[ d_{kX} = \arg\min_{i,j} \{ d_{ij}, d_{iB} \}. \]  
   If \( X = l \), combine proto-jets \( k \) and \( l \) into a new proto-jet with  
   \[ \vec{X}_{\text{new}} = \vec{X}_l + \vec{X}_k \] (E-scheme). Else, declare proto-jet \( k \) a jet and remove it from the list.

\( \vec{X} \) = 4-vector, with components \( p_x, p_y, p_z \) and \( E \).
Membership Functions

**Definition** For each particle $i$, its membership function $f_i$ is a map $f_i : \{\text{jets}\} \rightarrow [0, 1]$ such that $\sum_{j=1}^{k} f_i = 1$.

For the state-of-the-art clustering schemes, every clustered object belongs to exactly one jet (with probability 1) and thus

$$f_i(j) = \begin{cases} 1 & \text{i was clustered into jet j} \\ 0 & \text{else} \end{cases}$$

New developments† have shown that introducing probabilistic memberships can lead to enhanced understanding/performance.

Our goal is to incorporate **fuzziness** during clustering in order to **learn** the features of jetty events and (sub)jetty structure.

† Examples:

Qjets [1201.1914] Re-cluster with randomness. For particle $i$ and jet $j$, $f_i(j)$ is the fraction of clusterings in which $i$ is† in $j$.

N.B. Not exactly Qjets, which is sensitive to joint membership functions.

PUPPI [1407.6013] $f_i(j)$ based on local $p_T$ density – high $f_i(j)$ indicates higher likelihood for hard scatter origin.

→ Today: a Preliminary Look at Fuzzy Jet Clustering
Probabilistic Clustering

Many standard methods of unsupervised learning incorporate probabilistic membership as part of the clustering procedure.

One of the most basic such algorithms is the mixture model for $k$ jets,

$$p(\tilde{X}) = \sum_{j=1}^{k} \pi_j \Phi(\tilde{X} | \Theta_j)$$

Which is equivalent to a generative model phrased as follows:

$\xi_i \in \{1, ..., k\} \equiv$ jet to which particle $i$ belongs

$\xi_i \overset{\text{IID}}{\sim} \text{Multinomial}(\vec{\pi}, 1)$, for $\sum_{i=1}^{\pi} \pi_i = 1$.

$\tilde{X}_i | \xi_i \overset{\text{IID}}{\sim} \Phi(\tilde{X}_i | \Theta_{\xi_i})$, for a probability density $\Phi$ depending on parameters $\Theta$.

The probabilistic part of the clustering is in the random variables $\xi_i$. 

$\Theta_j$ are the jets

$\pi_j$ are (learned) priors
Consider a special case in which $\Phi$ is a Gaussian in $y - \phi$ space:

$$
\Phi(\vec{\rho}|\vec{\mu}_j, \Sigma_j) = 
\exp\left(-\frac{1}{2}(\vec{\rho} - \vec{\mu}_j)^\top \Sigma_j^{-1}(\vec{\rho} - \vec{\mu}_j)\right)
\sqrt{|2\pi\Sigma_j|}
$$

$\vec{\rho}$ is the $y - \phi$ (sub)coordinates of $\vec{X}$. $\Sigma$ is a $2 \times 2$ invertible matrix.

For a fixed $k$, the goal is to learn

1. $\vec{\mu}$ (jet positions)
2. $\Sigma$ (jet size and shape)
EM Algorithm

There is a well-known procedure for iteratively determining the distribution of the $\xi_i$ known as the Expectation-Maximization (EM) algorithm:

1. Initialize $\vec{\pi}$, $\vec{\mu}_j$ and $\Sigma_j$ (more on this later)
2. Alternate until convergence

**E Step** Compute membership functions $f_i(j) = \Pr(\xi_i = j | \vec{\mu}_j, \Sigma_j, \vec{\pi})$

**M Step** Given the $f_i(j)$, compute the new cluster centers and shapes:

$$
\pi_j = \sum_{i=1}^{n} \frac{f_i(j)}{n} \quad \vec{\mu}_j = \frac{\sum_i f_i(j) \vec{\rho}_i}{\sum_i f_i(j)} \quad \Sigma_j = \frac{\sum_i f_i(j)(\vec{\rho}_i - \vec{\mu}_j)(\vec{\rho}_i - \vec{\mu}_j)^T}{\sum_i f_i(j)}
$$

This is not sequential recombination!

N.B. This is Lloyd’s algorithm for solving $k$-means in the limit that $\Sigma = \sigma^2 I \to 0$ and $\pi_i = \frac{1}{k}$. 
Illustration: $k = 2$

- **Initialization**
- **1$^{\text{st}}$ E step**
- **1$^{\text{st}}$ M step**
- **2$^{\text{nd}}$ E step**
- **2$^{\text{nd}}$ M step**

In the E-step, darker colors correspond to higher value of $f_i$(blue jet).

- **3$^{\text{rd}}$ E step**
- **3$^{\text{rd}}$ M step**
- **4$^{\text{th}}$ E step**
- **4$^{\text{th}}$ M step**
- **5$^{\text{th}}$ M step: Converged**

The algorithm will converge, but not guaranteed to reach global optimum
Modified Mixture Model (mMM)

Out-of-the-box, Gaussian Mixture Model jets are not IRC safe.

However, there is a simple modification that makes a broad class of modified mixture model jets IR(C) safe, given by the likelihood

$$
\mathcal{L}(\vec{\pi}, \vec{\Theta}) = \sum_{i=1}^{N} p_{Ti}^\alpha \left( \sum_{j=1}^{k} \pi_j \Phi(\vec{\rho}_i | \vec{\Theta}_j) \right)
$$

- For $\alpha \to 0$, we recover the standard mixture model.
- For $\alpha = 1$, the jets are IRC safe for any $\Phi$ that depends only on $\vec{\rho}$.
- For $\alpha > 0$, the jets are generally IR safe.

Physically: $\alpha$ gives a larger weight to the hard structure of a jet.
Modification to the EM Algorithm

The EM algorithm still applies with one minor change: $f_i(j) \mapsto p_{T_i}^{\alpha} f_i(j)$.

For any kernel $\Phi$, one needs to re-derive the $M$-step, but this is straightforward for simple shapes:

- **Uniform**
- **Truncated Gaussian**
- **Full Gaussian**
- $\Sigma = \sigma^2 I$

**Preliminary** results shown in the following slides constructed with **Pythia 8.170** and **FastJet 3.0.3**.
Choosing the Initialization

For sequential recombination, $R$ is specified ahead of time and $k$ is learned.

For mMM jets, $k$ is specified ahead of time and (more than) $R$ is learned.

There are many possibilities for initializing; for example,

1. Run a sequential recombination scheme for $k$ and the $\vec{\mu}_{i=1..k}$.
2. Choose $k$ using some standard procedure (e.g. gap statistic) and then initialize uniformly.
3. Let every particle be a cluster and let clusters merge.

$\pi \propto 1, \Sigma = 0.5^2 I, \alpha = 1$

Cluster Merging Scheme

- More plots in backup
Learning $\sigma$ for mGMM Fuzzy Jets

We have studied several kernels $\Phi$ and many configurations.

For illustration, the next slides show mGMM jets with $\Sigma \propto \sigma^2 I$.

Initialization: anti-$k_t$ jets with $R = 1.0$ and $p_T > 5$ GeV.

Background QCD dijets, $\hat{p}_T > 200$ GeV
- 3-prong $Z' \rightarrow t\bar{t}$ (fully hadronic)
- 2-prong $W' \rightarrow W(\rightarrow qq')Z(\rightarrow e\bar{e})$

N.B. Signal events are weighted to match the background $p_T$ spectrum.

Define the kinematics of fuzzy jet $j$ by

$$\vec{X}_{\text{jet}}^j = \sum_{i=1}^{N} \vec{X}_{\text{particle}}^i \times \delta(\max_q f_i(q), j)$$

More information on jet kinematics ▶ in the backup

Ellipses are the $1\sigma$ contours
Pink stars are the top quarks

Fuzzy jets can (and do) overlap!
Comparing Jet Kinematics: Fuzzy Jets vs anti-\( k_t \)

Fuzzy mGMM jets with \( \Sigma \propto \sigma^2 I \) are very similar to anti-\( k_t \) R=1.0 jets.

\( p_T \) is determined mostly by the hard core: fuzzy jets learn the same core

Mass depends on the shape of the jet: fuzzy jets learn different shapes
The size $\sigma$ of fuzzy jets

Unlike usual sequential recombination$^\dagger$, jet size is learned by fuzzy jets. One measure is the Gaussian $\sigma$, though not directly comparable with $R$.

For fixed $p_T$, fuzzy jets is learning the sizes for these processes!

The subleading jets tend to be larger on average than the leading jets.

$^\dagger$ With modification, size can vary in sequential recombination - see *Jets with Variable R*
Correlations between $\sigma$ and anti-$k_t$ jet mass

As expected, for a fixed $p_T$, the higher mass jets have a larger size.

However, the correlation is not 100% - the fuzzy jet $\sigma$ is not just the characteristic anti-$k_t$ jet size.
Correlations between $\sigma$ and fuzzy jet mass

Even though $\sigma$ has a component uncorrelated with the characteristic anti-$k_t$ jet size, it is highly correlated with the characteristic fuzzy jet size.

![Graph showing correlations between $\sigma$ and fuzzy jet mass](image-url)
Tagging Hadronic $W$ boson and top quark decays

In some regimes, the uncorrelated component of $\sigma$ (and the mass) may be useful in discriminating heavy particle decays.

Still need to understand correlations with other (substructure) variables.
Today’s focus was on learning. In order to fully take advantage of the fuzzy part of mMM jets, one needs to incorporate correlations.

The generative model for mMM jets assume independence – fuzzy jets ‘volatility’ (à la Qjets) could be more powerful with correlations.

There are unsupervised learning algorithms that do this, but scale with complexity.

Incorporating correlations may necessary to learn the tree structure of QCD radiation.

With \( f_i(j) \), can compute a distribution for any jet observable. Details ▶ in the backup.
Conclusions

Fuzzy jets are an alternative clustering scheme that naturally incorporate probabilistic membership to learn features of hadronic final states.

- Given any probability density $\Phi$ that only depends on $y$ and $\phi$, there is a corresponding modified mixture model jet.
- Unlike sequential recombination schemes, these algorithms learn the shape of jets.
- In some regimes, this learning may help improve tagging performance.

Further exploration and comparisons are required to fully understand and exploit the potential of Fuzzy Jets!
BACKUP
• Cluster Merging Scheme
  ▶ Reference in the talk  ▶ Additional Plots

• Learning $\sigma$ Additional Plots at high $p_T$
  ▶ Reference in the talk  ▶ Additional Plots

• Learning $\sigma$ Comparisons with trimmed anti-$k_t$ mass
  ▶ Reference in the talk  ▶ Additional Plots

• Learning $\sigma$ Plots at low $p_T$
  ▶ Reference in the talk  ▶ Additional Plots

• Learning substructure Additional Plots
  ▶ Backup Slides

• Using Distributions
  ▶ Reference in the talk  ▶ Additional Plots
Need to fix a parameter $r_{\text{merge}}$. We find that if small enough, the actual value does not matter.

(chosen here to be $r = 0.01$)

With the appropriate choices of $\sigma$ and $\alpha$, one recovers a set of jets very similar to C/A (only depend on $\vec{\rho}$).

The next two slides show some examples of event displays with the clustering history overlaid, with learning $\Sigma = \sigma^2 I$ and a full covariance matrix $\Sigma$. 

\[ \sqrt{s} = 8 \text{ TeV PYTHIA } Z' \rightarrow t\bar{t}, \ m_{Z'} = 1.5 \text{ TeV} \]
$\sqrt{s} = 8$ TeV PYTHIA $Z' \rightarrow t\bar{t}$, $m_{Z'} = 1.5$ TeV

$\pi \propto 1$, $\Sigma = \sigma^2 I$, $\alpha = 1$

Cluster Merging Scheme
$\sqrt{s} = 8$ TeV PYTHIA $Z' \rightarrow t\bar{t}, m_{Z'} = 1.5$ TeV

$\pi \propto 1, \alpha = 1$

Cluster Merging Scheme
Membership functions can generate distributions for each jet observable. 

\[
\Pr(\mathcal{O}(J) \leq c) = \sum_{S} \Pr(\mathcal{O}(J) \leq c | S \in J) \Pr(S \in J)
\]

\(J\) is a jet
\(\mathcal{O}\) is a jet observable (like mass)
\(S\) is a set of constituents
\(\Pr(S \in J)\) is determined by the \(f_i(J)\)

From this distribution, one can use various moments as new jet observables.

E.g. coefficient of variation for the mass (Qjets volatility)
Correlations between $\sigma$ and anti-$k_t$ jet mass

**Pythia 8 - QCD, Correlation = 0.58**

**Pythia 8 - W→qq', Correlation = 0.64**
Correlations between $\sigma$ and fuzzy jet mass

**Pythia 8, QCD, Correlation = 0.92**

**Pythia 8, W$\rightarrow$ qq', Correlation = 0.77**
Comparisons with Trimmed Jet Mass

Pythia 8
$\sqrt{s} = 8$ TeV

$W \rightarrow qq'$

$Z' \rightarrow t\bar{t}$

QCD

Leading anti-$k_T$ Jet Mass [GeV]

Normalized to Unity

0
0.1
0.2
0.3
0.4
0.5
0.6

Leading Jet Mass [GeV]

Normalized to Unity

0
0.1
0.2
0.3
0.4
0.5
0.6

Benjamin Nachman (SLAC)
Comparisons with Trimmed Jet Mass II

Pythia 8

1 - QCD Efficiency

W→qq' Efficiency

350 GeV < \pT < 450 GeV

- Fuzzy σ
- anti-kT Mass
- Fuzzy Mass
- Random Tagger

R=1.0

Leading anti-kT

Benjamin Nachman (SLAC)
Comparisons with Trimmed Jet Mass III

Pythia 8  QCD, Correlation = 0.56

Leading anti-$k_t$ Jet Mass [GeV]

Learned $\sigma$

350 GeV $< p_t < 450$ GeV

Leading anti-$k_t$, $R=1.0$
Comparisons with Trimmed Jet Mass IV

Pythia 8 QCD, Correlation = 0.56

Learned σ

Leading anti-\(k_t\) Jet Mass/Jet \(p_T\)

350 GeV < \(p_T\) < 450 GeV

Benjamin Nachman (SLAC)
Lower $p_T$ bin: Jet Mass

![Graphs showing jet mass distributions for different processes.](Pythia 8, $\sqrt{s} = 8$ TeV, $250$ GeV $< p_T < 350$ GeV, leading anti-$k_T$, QCD)

**Pythia 8**

$\sqrt{s} = 8$ TeV

- $W \rightarrow qq'$
- $Z' \rightarrow t\bar{t}$
- QCD

Normalized to Unity

Leading Jet Mass [GeV]

Normalized to Unity

Leading anti-$k_T$ Jet Mass [GeV]
Lower $p_T$ bin: Fuzzy Mass and $\sigma$

Pythia 8 QCD, Correlation = 0.92

$\sigma$ Learned

Pythia 8 $W \rightarrow q\bar{q}'$, Correlation = 0.79

$R=1.0$

Leading anti-$k_T$

$250 \text{ GeV} < p_T < 350 \text{ GeV}$
Lower $p_T$ bin: Jet Mass and $\sigma$

- **Pythia 8** QCD, Correlation = 0.53

- **Pythia 8** $W \rightarrow qq'$, Correlation = 0.44

$< 350$ GeV

$R=1.0$

Leading anti-$k_T$ Jet Mass [GeV]
Lower $p_T$ bin: Fuzzy Mass/$p_T$ and $\sigma$

- **Pythia 8 QCD, Correlation = 0.94**

- **Pythia 8 $W \rightarrow q\bar{q}'$, Correlation = 0.84**
Lower $p_T$ bin: Jet Mass/$p_T$ and $\sigma$

**Pythia 8**

- **QCD**, Correlation = 0.54
- **W → qq'**, Correlation = 0.47

$< 350 \text{ GeV}$

$R=1.0$

Leading anti-$k_t$ Jet Mass/Jet $p_T$

**Learned $\sigma$**
Pythia 8
\( \sqrt{s} = 8 \text{ TeV} \)

- \(Z' \rightarrow t \bar{t}\)
- \(W \rightarrow qq'\)
- \(\text{QCD}\)

\(250 \text{ GeV} < p_T < 350 \text{ GeV}\)

\(R = 1.0\)

Leading anti-\(k_T\)

\(\sigma_{\text{Average Learned}}\)
Learning Substructure

One can apply the same techniques to learn the shapes of jet substructure. For example, $k = 2$ fuzzy jets with anti-$k_t$ $R = 1.0$ constituents as input:

Coordinates rotated and scaled - leading subjet is at $(0, 0)$ and has $\sigma = 1$. 

Pythia $\sqrt{s} = 8$ TeV QCD

One small and one broad subjet

Relatively symmetric subjets
Two observables based on the representative event displays: $\Delta R$ between learned fuzzy jets and the relative size of the second fuzzy subjet ($\sigma_2/\sigma_1$).

$W$ Further apart (correlated with $\Delta R$ of seeded $R=0.3$ anti-$k_t$ jets)

$W$ subjets are more symmetric in size than for QCD.

$\sqrt{s} = 8$ TeV

Pythia 8

Leading anti-$k_T$, $R=1.0$

200 GeV < $p_T$ < 350 GeV
Learning substructure: Correlations

Pythia 8  QCD, Correlation = 0.33

Pythia 8  W', Correlation = 0.48

Benjamin Nachman (SLAC)  Fuzzy Jets  August 18, 2014
Learning substructure: Correlations II

Pythia 8 QCD, Correlation = 0.07

Pythia 8 W’, Correlation = 0.23