Jet Formation with Spectral Clustering

arXiv:2104.01972

H.Day-Hall[1,2]
Supervised by; S.Dasmahapatra[1], S.Moretti[1], C.H.Shepherd-Themistocleous[2]

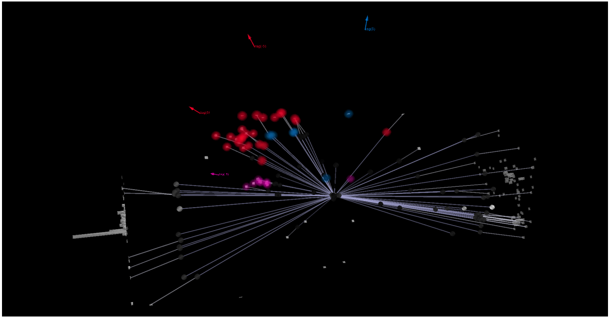
[1]University of Southampton, UK
[2]Rutherford Appleton Laboratory, UK

7th October 2021 at ICNFP 2021

Outline

Jet Formation

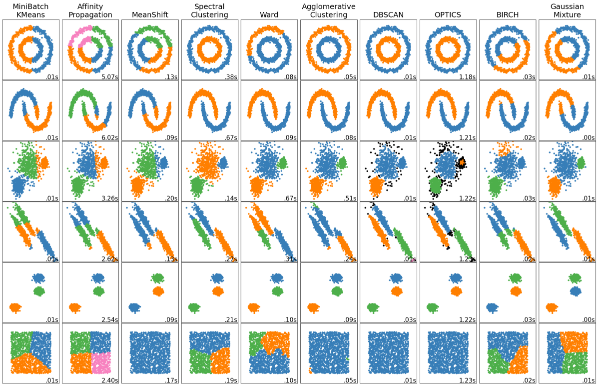
Jet formation involves grouping decay products to estimate the momentum and identity of the particle that decayed



This activity has been compared to reading tea leaves. . .
`www.quantumdiaries.org/2011/04/22/when-youre-a-jet-youre-a-jet-all-the-way/`

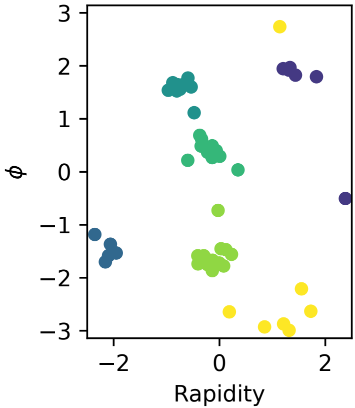
Spectral Clustering

Spectral clustering is a machine learning technique for picking out clusters. It doesn't use a neural net



`https://scikit-learn.org/stable/modules/clustering.html`

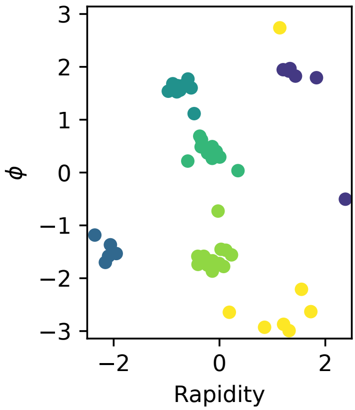
Spectral Clustering
Inputs



Here is a image representing hits on the unrolled barrel. Each point is a detected track, its colour roughly indicates which shower it was generated by. It contains;

- 4 *b*-showers from a light Higgs cascade.
- Background showers from initial state radiation.

Spectral Clustering
Inputs

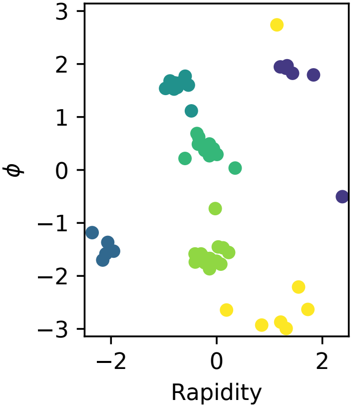


How likely are two particles to belong in the same jet?
Distance; $d_{i,j}$

$$= \sqrt{\delta\phi_{i,j}^2 + \delta y_{i,j}^2}$$

Spectral Clustering

Inputs



How likely are two particles to belong in the same jet?

Distance; $d_{i,j}$

$$= \sqrt{\delta\phi_{i,j}^2 + \delta y_{i,j}^2}$$

Similarity; $a_{i,j}$

$$= \exp\left(\frac{-d_{i,j}^2}{\sigma_v}\right)$$

Spectral Clustering

Cost function

Ideally, particles connected by high similarity would be in the same jet. Particles separated into different jets would only have low similarity between them.

Spectral Clustering
Cost function

Ideally, particles connected by high similarity would be in the same jet. Particles separated into different jets would only have low similarity between them.

The size of the groups should be about the same.

Spectral Clustering
Cost function

Ideally, particles connected by high similarity would be in the same jet. Particles separated into different jets would only have low similarity between them.

The size of the groups should be about the same.

Cost function;

$$\text{NCut} = \sum_k \frac{\sum_{i \in G_k, j \in \bar{G}_k} a_{i,j}}{|G_k|}$$

Where

The numerator is the similarity crossing the boundary of G_k.

$|G_k| = \sum_{i \in k} \sum_j a_{i,j}$ seeks to balance cluster size.

Spectral Clustering

Indicator vectors

Let's create some indicator vectors

$$f(k)_i = \begin{cases} \frac{1}{\sqrt{|G_k|}} & \text{if } i \in G_k \\ 0 & \text{otherwise} \end{cases}$$

If these could be found it would define the jets.

Spectral Clustering
Indicator vectors

Let's create some indicator vectors

$$f(k)_i = \begin{cases} \frac{1}{\sqrt{|G_k|}} & \text{if } i \in G_k \\ 0 & \text{otherwise} \end{cases}$$

If these could be found it would define the jets.

A Laplacian helps with that; let $D_{i,j} = \delta_{i,j} \left(\sum_k a_{i,k}\right)$

$$L_{i,j} = D_{i,j} - a_{i,j}$$

Spectral Clustering
Indicators

$$
\begin{aligned}
f(k)' L f(k) &= \sum_{i,j} f(k)_i L_{i,j} f(k)_j \\
&= \sum_{i,j} f(k)_i \left(\delta_{i,j} \sum_p a_{i,p} - a_{i,j} \right) f(k)_j \\
&= \sum_i \left(f(k)_i^2 \sum_p a_{i,p} - \sum_j f(k)_i f(k)_j a_{i,j} \right) \\
&= \sum_{i,j} a_{i,j} \left(f(k)_i^2 - f(k)_i f(k)_j \right) \\
&= \frac{1}{2} \sum_{i,j} a_{i,j} \left(f(k)_i - f(k)_j \right)^2
\end{aligned}
$$

Spectral Clustering
Indicator vectors

Use the definition of indicator vectors again;

$$f(k)_i = \begin{cases} \frac{1}{\sqrt{|G_k|}} & \text{if } i \in G_k \\ 0 & \text{otherwise} \end{cases}$$

This gives something familiar looking...

$$f(k)' L f(k) = \sum_{i \in G_k, j \notin G_k} \frac{a_{i,j}}{|G_k|}$$

Spectral Clustering

Indicator vectors

Use the definition of indicator vectors again;

$$f(k)_i = \begin{cases} \frac{1}{\sqrt{|G_k|}} & \text{if } i \in G_k \\ 0 & \text{otherwise} \end{cases}$$

This gives something familiar looking...

$$f(k)' L f(k) = \sum_{i \in G_k, j \notin G_k} \frac{a_{i,j}}{|G_k|}$$

It's part of the cost function;

$$\text{NCut} = \sum_k \frac{\sum_{i \in G_k, j \in \bar{G}_k} a_{i,j}}{|G_k|}$$

Spectral Clustering
Indicator vectors

Use the definition of indicator vectors again;

$$f(k)_i = \begin{cases} \frac{1}{\sqrt{|G_k|}} & \text{if } i \in G_k \\ 0 & \text{otherwise} \end{cases}$$

This gives something familiar looking...

$$f(k)' L f(k) = \sum_{i \in G_k, j \notin G_k} \frac{a_{i,j}}{|G_k|}$$

It's part of the cost function;

$$\text{NCut} = \sum_k \frac{\sum_{i \in G_k, j \in \bar{G}_k} a_{i,j}}{|G_k|}$$

So if F is a matrix with rows of $f(n)$;

$$\text{NCut} = \text{tr}(F' L F) = F' T F$$

Spectral Clustering

Indicator vectors

Use the definition of indicator vectors again;

$$f(k)_i = \begin{cases} \frac{1}{\sqrt{|G_k|}} & \text{if } i \in G_k \\ 0 & \text{otherwise} \end{cases}$$

This gives something familiar looking...

$$f(k)' L f(k) = \sum_{i \in G_k, j \notin G_k} \frac{a_{i,j}}{|G_k|}$$

It's part of the cost function;

$$\text{NCut} = \sum_k \frac{\sum_{i \in G_k, j \in \bar{G}_k} a_{i,j}}{|G_k|}$$

So if F is a matrix with rows of $f(n)$;

$$\text{NCut} = \mathrm{tr}(F' L F) = F' T F$$

Finally, let $T = D^{\frac{1}{2}} F$. Then $T'T = I$, so

$$\text{NCut} = \frac{T' D^{-\frac{1}{2}} L D^{-\frac{1}{2}} T}{T'T}$$

Spectral Clustering
Embedding space

To minimise this cost function, take eigenvectors of
$L = D^{-\frac{1}{2}}(D - A)D^{-\frac{1}{2}} \rightarrow L\mathbf{v_n} = \lambda_n \mathbf{v_n}$

Spectral Clustering
Embedding space

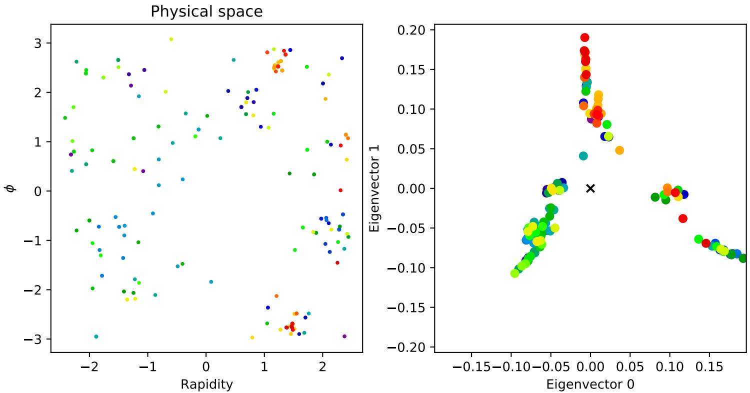
To minimise this cost function, take eigenvectors of
$L = D^{-\frac{1}{2}}(D - A)D^{-\frac{1}{2}} \rightarrow L\mathbf{v_n} = \lambda_n \mathbf{v_n}$

$$\mathbf{v_1}, \mathbf{v_2} = \begin{bmatrix} \mathbf{v_1}_1 \\ \mathbf{v_1}_2 \\ \vdots \\ \mathbf{v_1}_N \end{bmatrix}, \begin{bmatrix} \mathbf{v_2}_1 \\ \mathbf{v_2}_2 \\ \vdots \\ \mathbf{v_2}_N \end{bmatrix}$$

Spectral Clustering
Embedding space

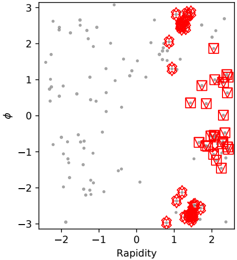
The eigenvectors of L create a multidimensional embedding space.

$$Lv_n = \lambda_n v_n$$



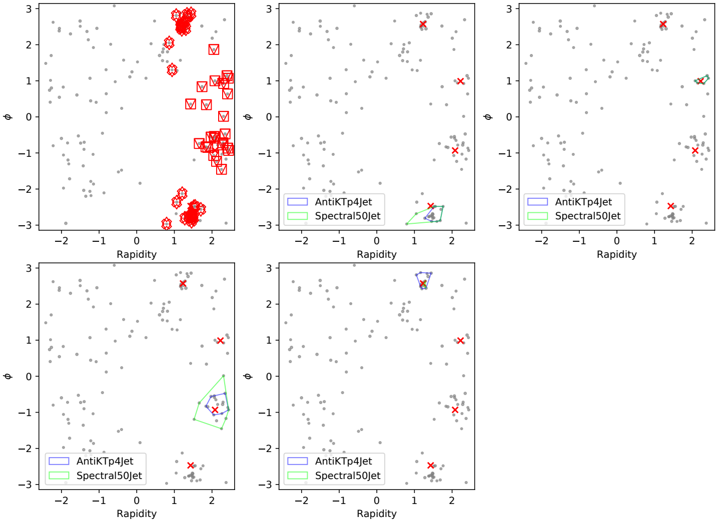
Spectral Clustering

Jet shape



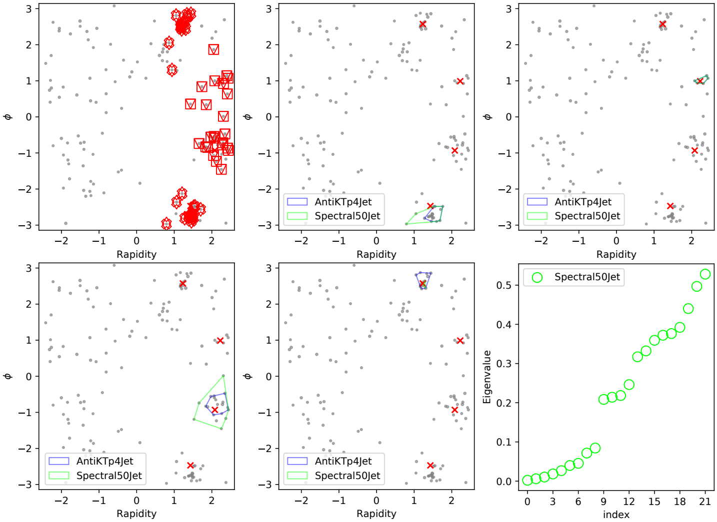
Spectral Clustering

Jet shape



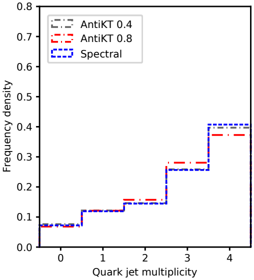
Spectral Clustering

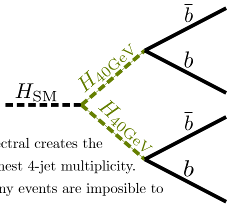
Jet shape



Results

Light Higgs cascade

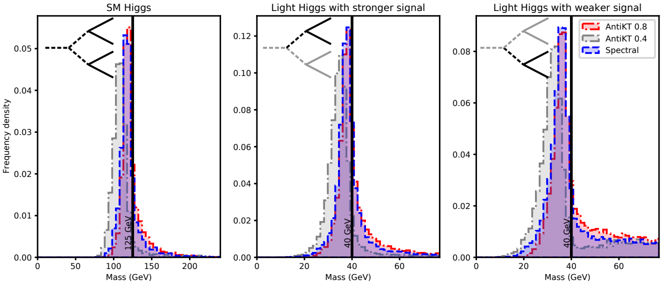




Spectral and anti-kt radius 0.8 both give good narrow mass peaks with the correct location.

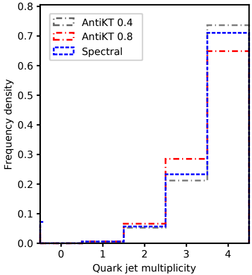
Spectral creates the highest 4-jet multiplicity. Many events are imposible to reconstruct due to low energy.

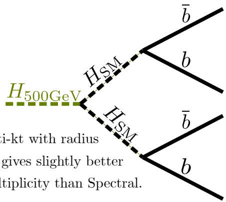
Only spectral is producing good mass peaks and multiplicity



Results

Heavy Higgs cascade

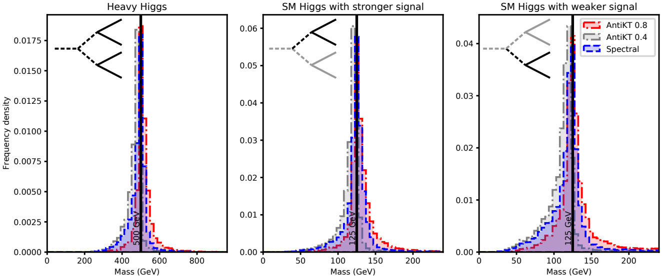




Spectral and anti-kt radius 0.8 both give good narrow mass peaks with the correct location.

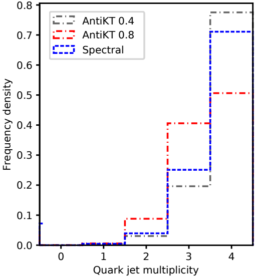
Anti-kt with radius 0.4 gives slightly better multiplicity than Spectral.

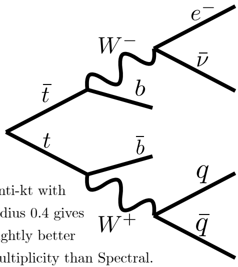
Spectral is producing good mass peaks and multiplicity



Results

Semileptonic top decay

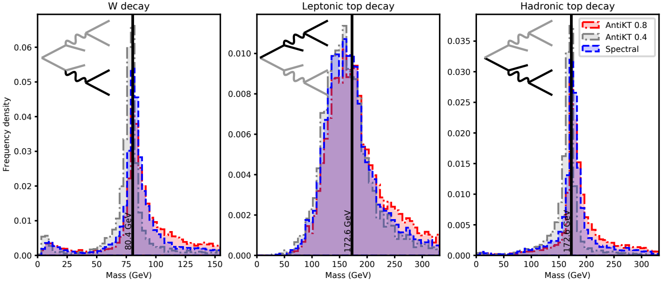




Anti-kt radius 0.4 produces the best mass peaks. Spectral is an improvement on anti-kt radius 0.8.

Anti-kt with radius 0.4 gives slightly better multiplicity than Spectral.

Anti-kt 0.4 is producing good mass peaks and multiplicity, spectral is partially mimicking this.



Conclusion

Spectral clustering offers a deterministic crisp jet formation algorithm, without any use of learnt parameters or black box elements. It is remarkable for its ability to adapt to various data sets.

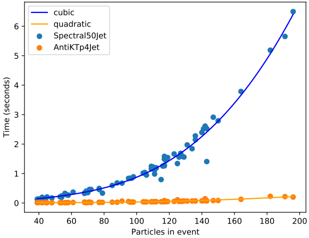
arXiv:2104.01972

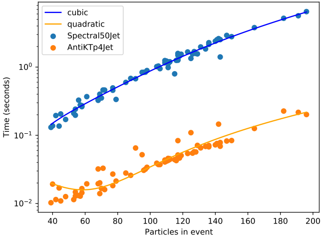
End

Thank you!

Spectral Clustering
Timing

The algorithm is currently $\mathcal{O}(N^3)$.





Spectral Clustering
Parameters

- σ_v and α are used to define the similarity; $a_{i,j} = e^{-d_{i,j}^{\alpha}/\sigma_v}$.

Spectral Clustering
Parameters

- σ_v and α are used to define the similarity; $a_{i,j} = e^{-d_{i,j}^{\alpha}/\sigma_v}$.
- k_{NN} removes some similarities to reduce noise.

Spectral Clustering
Parameters

- σ_v and α are used to define the similarity; $a_{i,j} = e^{-d_{i,j}^{\alpha}/\sigma_v}$.
- k_{NN} removes some similarities to reduce noise.
- λ_{limit} and β make use of the information in the eigenvalues to prioritise the eigenvectors.

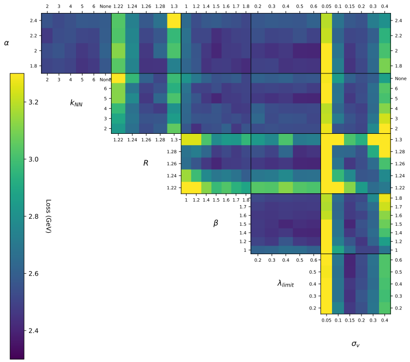
Spectral Clustering

Parameters

- σ_v and α are used to define the similarity; $a_{i,j} = e^{-d_{i,j}^{\alpha}/\sigma_v}$.
- k_{NN} removes some similarities to reduce noise.
- λ_{limit} and β make use of the information in the eigenvalues to prioritise the eigenvectors.
- When the mean distance in the embedding space rises over R, the clustering stops.

Spectral Clustering
Parameters

The parameters are not fine tuned. This can be seen if we plot the performance as the parameters change;

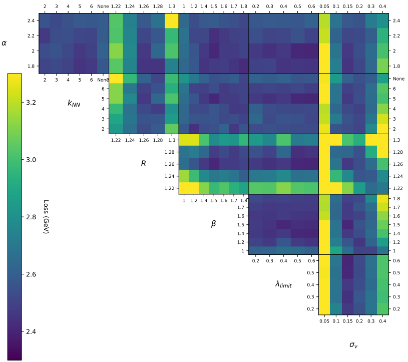


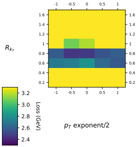
Dark blue areas, indicating low loss, show parameter combinations with good performance.

Spectral Clustering

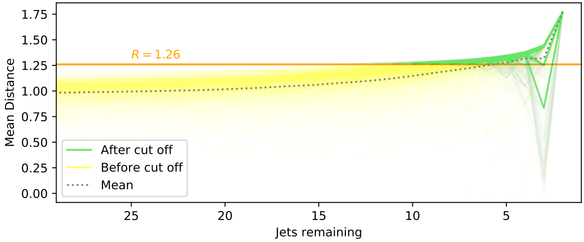
Parameters

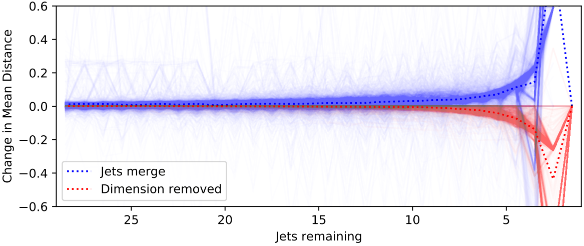
For contrast, on the right is the parameter space of the generalised k_T algorithm





Spectral Clustering
Stopping condition





Spectral Clustering

Shape variables

Shape variables are dependant on the clustering algorithm