Spectral Clustering for Jet Physics

What is a jet?
Many particles produced in high energy collisions decay before they can be detected. Clustering the stable particles that are produced is key to reconstructing the particle that decayed. This cluster is called a jet [2].

What is spectral clustering?
Spectral clustering is a Machine Learning (ML) technique for detecting clusters in points with noisy configurations. It requires the construction of a graph Laplacian and calculation of the eigenvectors of the Laplacian to optimise the groupings of the points [3].

How to form the embedding space
Here is a minimal description of how to form an embedding space. The full process is described in [1].

1. Calculate distances between all pairs of points.
   \[ d_{i,j} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2} \]
2. Calculate affinities from distances.
   \[ a_{i,j} = \exp\left(-\frac{d_{i,j}^2}{\sigma^2}\right) \]
3. Construct Laplacian matrix.
   \[ D_{i,j} = \delta_{i,j} \sum_k a_{i,k} \quad A_{i,j} = a_{i,j} \]
   \[ L = D^{-1/2}(D - A)D^{-1/2} \]
4. Find eigenvectors of Laplacian with small eigenvalues.
   \[ Lx_n = \lambda_n x_n \quad \lambda_n < \lambda_{limit} \]
5. Define embedding coordinates from eigenvectors.
   \[ z_i = (x_{n=0,i}, x_{n=1,i}, \ldots, x_{n,1}) \]

Clustering process
To begin with, all particles are considered pseudojets (candidate jets).

1. Using current pseudojets as points, calculate the embedding space.
2. If there are no dimensions in the embedding space, or the mean distance between points is too high stop clustering.
3. Identify the closest 2 points in embedding space by angle about the origin.
4. Merge the 2 pseudojets corresponding to the points by combing their 4 momentum (E-scheme recombination).
5. Return to step 1.

When clustering ends, all pseudojets become jets. Jets with less than 2 tracks, or low transverse momentum are removed.

[1] S. Dasmahapatra et al. Spectral Clustering for Jet Physics. 2021. arXiv: 2104.01972 [hep-ph].
[2] Gavin P. Salam. “Towards jetography”. In: The European Physical Journal C 67.3-4 (May 2010), pp. 637–686. ISSN : 1434-6052. DOI : 10.1140/epjc/s10052-010-1314-6.
[3] Ulrike von Luxburg. A Tutorial on Spectral Clustering. 2007. arXiv: 0711.0189 [cs.DS].
Infrared and collinear safety
Jet finding algorithms must be insensitive to soft and collinear emissions. That is, the locations and sizes of the jets must not be influenced to particles in this limit.

Soft emission; any angle, very low energy. Collinear splitting; both products have almost identical direction.

Neither of these may effect the momentum of the jets produced, else the values of the shape variables would be influenced.

Multi-configuration scan
In the top right we compare shape variables for one configuration, but what if we change the parameters? Are the clustering algorithms still IR-safe if the jet radius (or other parameters) change?

The NLO dataset has soft emissions and collinear splittings, but the LO dataset does not. This is known to be IR-safe. Can we determine if this is IR-safe?

Here we check a single parameter configuration (jet radius etc.) for two algorithms. As the shape variables (Mass, Thrust, Sphericity and Oblateness) have the same distributions on LO and NLO data we know both algorithms are insensitive to infrared and collinear radiation (IR-safe).

The Jensen Shannon score is a summary statistic that will let us compare many parameter configurations in one plot.

Spectral Clustering for Jet Physics
Spectral Clustering performance

Jet formation algorithms are evaluated based on two metrics: jet multiplicity and mass resolution.

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

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

Semi-leptonic top decay $e^-$ produces the best mass peaks. Spectral is an improvement on anti-kt radius 0.8.

Anti-kt radius 0.4 produces the best mass peaks. Spectral is partially mimicking this.

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.

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

Spectral is capable of reconstructing a variety of events without changing its parameters.