Efficient Evaluation of Spectral EMD

Andrew Larkoski

June 26, 2023

1 Closed-Form Evaluation of Spectral EMD

We will work to evaluate the $p = 2$ spectral EMD metric, where

$$\begin{aligned} d_2(s_1, s_2)^2 &= \int dE^2 \left|S_1^{-1}(E^2) - S_2^{-1}(E^2)\right|^2 \\ &= \int dE^2 \left(S_1^{-1}(E^2)^2 + S_2^{-1}(E^2)^2 - 2S_1^{-1}(E^2)S_2^{-1}(E^2)\right) . \end{aligned} \tag{1}$$

Here, the inverse cumulative spectral function can be evaluated from

$$S^{-1}(E^2) = \int_0^{\omega_{\max}} d\omega\, \Theta\left(E^2 - S(\omega)\right) , \tag{2}$$

where $S(\omega)$ is the familiar cumulative spectral function. With this representation, we can evaluate the integrals with integrands that are squares of the inverse cumulative spectral functions:

$$\begin{aligned} \int dE^2\, S^{-1}(E^2)^2 &= \int dE^2 \int d\omega \int d\omega'\, \Theta\left(E^2 - S(\omega)\right)\Theta\left(E^2 - S(\omega')\right) \\ &= 2\int dE^2 \int d\omega \int d\omega'\, \Theta\left(E^2 - S(\omega)\right)\Theta\left(S(\omega) - S(\omega')\right) \\ &= 2\int d\omega \int d\omega'\, \left(E_{\text{tot}}^2 - S(\omega)\right)\Theta\left(S(\omega) - S(\omega')\right) . \end{aligned} \tag{3}$$

Now, because $S(\omega)$ is monotonic,

$$\Theta\left(S(\omega) - S(\omega')\right) = \Theta\left(\omega - \omega'\right) , \tag{4}$$

and so we can integrate over ω' to produce

$$\begin{aligned} \int dE^2\, S^{-1}(E^2)^2 &= 2\int_0^{\omega_{\max}} d\omega\, \omega\left(E_{\text{tot}}^2 - S(\omega)\right) \\ &= E_{\text{tot}}^2\omega_{\max}^2 - 2\int_0^{\omega_{\max}} d\omega\, \omega\, S(\omega) . \end{aligned} \tag{5}$$

Then, we can simplify this expression with integration by parts and so, we finally find

$$\int dE^2\, S^{-1}(E^2)^2 = \int_0^{\omega_{\max}} d\omega\, \omega^2\, s(\omega)\,, \tag{6}$$

just the second moment of the spectral function.

This final integral can then be evaluated directly from the expression for the cumulative spectral EMD:

$$S(\omega) = \sum_i E_i^2 + \sum_{i<j} 2E_iE_j\Theta(\omega-\omega_{ij})\,, \tag{7}$$

and so

$$2\int_0^{\omega_{\max}} d\omega\, \omega\, S(\omega) = E_{\text{tot}}^2\omega_{\max}^2 - \sum_{i<j} 2E_iE_j\,\omega_{ij}^2\,. \tag{8}$$

Therefore, the integral of the square of the inverse cumulative spectral function is just a two-point energy correlation function

$$\int dE^2\, S^{-1}(E^2)^2 = \sum_{i<j} 2E_iE_j\omega_{ij}^2\,. \tag{9}$$

where the sum over all particles $i<j$ in the event.

Then, the squared $p=2$ spectral EMD can be expressed as

$$d_2(s_1,s_2)^2 = \sum_{i<j\in\mathcal{E}_1} 2E_iE_j\omega_{ij}^2 + \sum_{i<j\in\mathcal{E}_2} 2E_iE_j\omega_{ij}^2 - 2\int dE^2\, S_1^{-1}(E^2)S_2^{-1}(E^2)\,. \tag{10}$$

Moving on to the mixed term, we will evaluate it from the explicit form of the inverse cumulative spectral function, $S^{-1}(E^2)$. To do this, we will introduce some notation. We can write the cumulative spectral function as

$$S(\omega) = \sum_i E_i^2 + \sum_{i<j} 2E_iE_j\Theta(\omega-\omega_{ij}) \equiv \sum_i E_i^2 + \sum_{\substack{n\\ \omega_n<\omega_{n+1}}} (2EE)_n\Theta(\omega-\omega_n)\,. \tag{11}$$

In the sum on the right, ω_n is one of the $\binom{N}{2}$ pairwise angles between the N particles in the event and $(2EE)_n$ is its corresponding squared energy weight. In the sum, we have ordered the angular factors so that $\omega_n<\omega_{n+1}$. Using this notation, the inverse cumulative spectral function takes the form:

$$S^{-1}(E^2) = \sum_{\substack{n\\ \omega_n<\omega_{n+1}}} \omega_n\,\Theta\left(\sum_i E_i^2 + \sum_{m\leq n}(2EE)_m - E^2\right)\Theta\left(E^2 - \sum_i E_i^2 - \sum_{m<n}(2EE)_m\right)\,. \tag{12}$$

The proof of this statement follows simply by sketching the functional form of the cumulative spectral function and then rotating the image by 90°.

With this expression, we can then explicitly evaluate the integral that remains in the $p = 2$ spectral EMD. We have

$$
\begin{aligned}
&\int dE^2\, S_1^{-1}(E^2) S_2^{-1}(E^2) \qquad (13)\\
&= \sum_{\substack{n\in\mathcal{E}_1^2,\, l\in\mathcal{E}_2^2\\ \omega_n<\omega_{n+1}\\ \omega_l<\omega_{l+1}}} \omega_n\omega_l \int dE^2\, \Theta\left(\sum_{i\in\mathcal{E}_1} E_i^2 + \sum_{m\leq n\in\mathcal{E}_1^2}(2EE)_m - E^2\right)\Theta\left(E^2 - \sum_{i\in\mathcal{E}_1}E_i^2 - \sum_{m<n\in\mathcal{E}_1^2}(2EE)_m\right)\\
&\qquad \times \Theta\left(\sum_{i\in\mathcal{E}_2} E_i^2 + \sum_{k\leq l\in\mathcal{E}_2^2}(2EE)_k - E^2\right)\Theta\left(E^2 - \sum_{i\in\mathcal{E}_2}E_i^2 - \sum_{k<l\in\mathcal{E}_2^2}(2EE)_k\right).
\end{aligned}
$$

Here, the notation that $n \in \mathcal{E}_1^2 \equiv \mathcal{E}_1 \times \mathcal{E}_1$; that is, n represents a pair of particles from event $\mathcal{E}_1$. The integrand is simple enough that it can be evaluated exactly, with

$$
\begin{aligned}
&\int dE^2\, S_1^{-1}(E^2) S_2^{-1}(E^2) \qquad (14)\\
&= \sum_{\substack{n\in\mathcal{E}_1^2,\, l\in\mathcal{E}_2^2\\ \omega_n<\omega_{n+1}\\ \omega_l<\omega_{l+1}}} \omega_n\omega_l \left(\min\left[\sum_{i\in\mathcal{E}_1} E_i^2 + \sum_{m\leq n\in\mathcal{E}_1^2}(2EE)_m, \sum_{i\in\mathcal{E}_2} E_i^2 + \sum_{k\leq l\in\mathcal{E}_2^2}(2EE)_k\right]\right.\\
&\qquad\qquad \left. - \max\left[\sum_{i\in\mathcal{E}_1} E_i^2 + \sum_{m< n\in\mathcal{E}_1^2}(2EE)_m, \sum_{i\in\mathcal{E}_2} E_i^2 + \sum_{k< l\in\mathcal{E}_2^2}(2EE)_k\right]\right)\\
&\qquad \times \Theta\left(\min\left[\sum_{i\in\mathcal{E}_1} E_i^2 + \sum_{m\leq n\in\mathcal{E}_1^2}(2EE)_m, \sum_{i\in\mathcal{E}_2} E_i^2 + \sum_{k\leq l\in\mathcal{E}_2^2}(2EE)_k\right]\right.\\
&\qquad\qquad \left. - \max\left[\sum_{i\in\mathcal{E}_1} E_i^2 + \sum_{m< n\in\mathcal{E}_1^2}(2EE)_m, \sum_{i\in\mathcal{E}_2} E_i^2 + \sum_{k< l\in\mathcal{E}_2^2}(2EE)_k\right]\right).
\end{aligned}
$$

Note that if $\mathcal{E}_1 = \mathcal{E}_2$ then this result agrees with Eq. 9.

Then, putting it all together, the complete, closed form expression for the $p = 2$ spectral EMD metric is:

$$
\begin{aligned}
d_2(s_1, s_2)^2 = &\sum_{i<j\in\mathcal{E}_1} 2E_iE_j\omega_{ij}^2 + \sum_{i<j\in\mathcal{E}_2} 2E_iE_j\omega_{ij}^2 \qquad (15)\\
&- 2\sum_{\substack{n\in\mathcal{E}_1^2,\, l\in\mathcal{E}_2^2\\ \omega_n<\omega_{n+1}\\ \omega_l<\omega_{l+1}}} \omega_n\omega_l \left(\min\left[\sum_{i\in\mathcal{E}_1} E_i^2 + \sum_{m\leq n\in\mathcal{E}_1^2}(2EE)_m, \sum_{i\in\mathcal{E}_2} E_i^2 + \sum_{k\leq l\in\mathcal{E}_2^2}(2EE)_k\right]\right.
\end{aligned}
$$

$$
\left. - \max\left[\sum_{i\in\mathcal{E}_1} E_i^2 + \sum_{m<n\in\mathcal{E}_1^2}(2EE)_m, \sum_{i\in\mathcal{E}_2} E_i^2 + \sum_{k<l\in\mathcal{E}_2^2}(2EE)_k\right]\right)
$$

$$
\times\,\Theta\left(\min\left[\sum_{i\in\mathcal{E}_1} E_i^2 + \sum_{m\leq n\in\mathcal{E}_1^2}(2EE)_m, \sum_{i\in\mathcal{E}_2} E_i^2 + \sum_{k\leq l\in\mathcal{E}_2^2}(2EE)_k\right]\right.
$$

$$
\left. - \max\left[\sum_{i\in\mathcal{E}_1} E_i^2 + \sum_{m<n\in\mathcal{E}_1^2}(2EE)_m, \sum_{i\in\mathcal{E}_2} E_i^2 + \sum_{k<l\in\mathcal{E}_2^2}(2EE)_k\right]\right)\,.
$$

Without loss of generality, we can assume both events have the same total energy E_{tot}, and so this can equivalently be expressed as

$$
\begin{aligned}
d_2(s_1,s_2)^2 &= \sum_{i<j\in\mathcal{E}_1} 2E_iE_j\omega_{ij}^2 + \sum_{i<j\in\mathcal{E}_2} 2E_iE_j\omega_{ij}^2 \qquad (16)\\
&- 2\sum_{\substack{n\in\mathcal{E}_1^2,\, l\in\mathcal{E}_2^2\\ \omega_n<\omega_{n+1}\\ \omega_l<\omega_{l+1}}} \omega_n\omega_l \left(\min\left[\sum_{n\leq m\in\mathcal{E}_1^2}(2EE)_m, \sum_{l\leq k\in\mathcal{E}_2^2}(2EE)_k\right] - \max\left[\sum_{n<m\in\mathcal{E}_1^2}(2EE)_m, \sum_{l<k\in\mathcal{E}_2^2}(2EE)_k\right]\right)\\
&\times\,\Theta\left(\min\left[\sum_{n\leq m\in\mathcal{E}_1^2}(2EE)_m, \sum_{l\leq k\in\mathcal{E}_2^2}(2EE)_k\right] - \max\left[\sum_{n<m\in\mathcal{E}_1^2}(2EE)_m, \sum_{l<k\in\mathcal{E}_2^2}(2EE)_k\right]\right)\,.
\end{aligned}
$$

Note that the computational complexity is effectively all encoded in the large cross term. As there are of order N^2 pairs of particles in an event of N total particles, the cost of evaluating the first two terms,

$$
d_2(s_1,s_2)^2 \supset \sum_{i<j\in\mathcal{E}_1} 2E_iE_j\omega_{ij}^2 + \sum_{i<j\in\mathcal{E}_2} 2E_iE_j\omega_{ij}^2\,, \tag{17}
$$

is $\mathcal{O}(N^2)$. For the cross term, we have to first sort the order-N^2 pairwise angles, which in general naively costs $\mathcal{O}(N^2\log N)$, but can be significantly sped up with parallelization. Then, actually evaluating the cross term once all angles are sorted only costs $\mathcal{O}(N^2)$ time, if you step through the sorted angles in increasing order.

1.1 General Analysis for Distance Between Discrete and Continuous Distributions

It will be useful to construct the distance between a jet/event with a discrete number of particles in it and a continuous spectral function distribution to measure the continuous shape-e-ness of a distribution of particles. We have already evaluated the square of the inverse cumulative spectral functions in both the discrete and continuous cases, so we just need to evaluate the mixed integral. We have

$$\int dE^2\, S^{-1}(E^2)\, S_{\text{cont}}^{-1}(E^2) = \int d\omega \int dE^2\, S^{-1}(E^2)\, \Theta\left(E^2 - S_{\text{cont}}(\omega)\right)\,, \tag{18}$$

where $S_{\text{cont}}(\omega)$ is the cumulative spectral function of the continuous distribution. We can perform the integral over E^2 from the expression for the discrete inverse cumulative spectral function, where we find

$$\begin{aligned}\int dE^2\, S^{-1}(E^2)\, S_{\text{cont}}^{-1}(E^2) = \sum_{\substack{n\in\mathcal{E}^2\\ \omega_n<\omega_{n+1}}} \omega_n \int d\omega\, \Theta(\cdot) \\ \times \left(E_{\text{tot}}^2 - \sum_{n<m\in\mathcal{E}^2}(2EE)_m - \max\left[E_{\text{tot}}^2 - \sum_{n\leq m\in\mathcal{E}^2}(2EE)_m, S_{\text{cont}}(\omega)\right]\right)\,,\end{aligned} \tag{19}$$

where $\Theta(\cdot)$ is shorthand for the constraint that the integrand be positive. There are then two cases that can be expanded out, where

$$\begin{aligned}\int dE^2\, S^{-1}(E^2)\, S_{\text{cont}}^{-1}(E^2) &= \sum_{\substack{n\in\mathcal{E}^2\\ \omega_n<\omega_{n+1}}} \omega_n(2EE)_n \int d\omega\, \Theta\left(E_{\text{tot}}^2 - \sum_{n\leq m\in\mathcal{E}^2}(2EE)_m - S_{\text{cont}}(\omega)\right) \\ &+ \sum_{\substack{n\in\mathcal{E}^2\\ \omega_n<\omega_{n+1}}} \omega_n \int d\omega \left(E_{\text{tot}}^2 - \sum_{n<m\in\mathcal{E}^2}(2EE)_m - S_{\text{cont}}(\omega)\right) \\ &\times \Theta\left(E_{\text{tot}}^2 - \sum_{n<m\in\mathcal{E}^2}(2EE)_m - S_{\text{cont}}(\omega)\right) \Theta\left(S_{\text{cont}}(\omega) - E_{\text{tot}}^2 + \sum_{n\leq m\in\mathcal{E}^2}(2EE)_m\right)\,.\end{aligned} \tag{20}$$

While there is still an explicit integral in this general form that must be done, no functional inversions are required so its numerical evaluation is pretty simple.

2 Some Observables

From this closed-form expression for the spectral EMD, we can define various observables that are sensitive to different kinematic configurations.

2.1 1-Pronginess

First, we consider sensitivity to radiation off of a single, hard core. To do this, we define the observable as the spectral distance from the jet of interest $\mathcal{E}$ to the jet that consists of a single particle $\mathcal{E}_1$. Assuming that the jets have the same total energy E, the cumulative spectral function of $\mathcal{E}_1$ is especially simple:

$$S(\mathcal{E}_1) = E^2\,. \tag{21}$$

Therefore, the distance between these jets is

$$d_2(\mathcal{E},\mathcal{E}_1)^2 = \sum_{i<j\in\mathcal{E}} 2E_iE_j\omega_{ij}^2\,. \tag{22}$$

This is just a two-point energy correlation function.

2.2 2-Pronginess

For sensitivity to jets with two hard cores of radiation, we can calculate the distance between a jet of interest $\mathcal{E}$ and a jet with two hard particles, $\mathcal{E}_2$, and then minimize the distance over the phase space variables of jet $\mathcal{E}_2$. The cumulative spectral function for this reference jet is

$$S(\mathcal{E}_2) = (E^2 - 2E_1E_2) + 2E_1E_2\Theta(\omega-\omega_{12})\,. \tag{23}$$

This form makes it clear that the relevant phase space variables are the angle ω_{12} and the product of particle energies, $2E_1E_2$. From the closed-form of the spectral EMD, the distance between these events is

$$\begin{aligned} d_2(\mathcal{E},\mathcal{E}_2)^2 = &\sum_{i<j\in\mathcal{E}} 2E_iE_j\omega_{ij}^2 + 2E_1E_2\omega_{12}^2 \\ &- 2\sum_{\substack{n\in\mathcal{E}^2\\ \omega_n<\omega_{n+1}}} \omega_n\omega_{12}\Bigg(\min\left[\sum_{i\in\mathcal{E}}E_i^2 + \sum_{m\leq n\in\mathcal{E}^2}(2EE)_m, E^2\right] \\ &\qquad\qquad - \max\left[\sum_{i\in\mathcal{E}_1}E_i^2 + \sum_{m<n\in\mathcal{E}_1^2}(2EE)_m, E^2-2E_1E_2\right]\Bigg) \\ &\times\Theta\Bigg(\min\left[\sum_{i\in\mathcal{E}}E_i^2 + \sum_{m\leq n\in\mathcal{E}^2}(2EE)_m, E^2\right] \\ &\qquad\qquad - \max\left[\sum_{i\in\mathcal{E}}E_i^2 + \sum_{m<n\in\mathcal{E}^2}(2EE)_m, E^2-2E_1E_2\right]\Bigg)\,. \end{aligned} \tag{24}$$

Note that, because the total energy is E, the upper bound in the parentheses is always set by the jet of interest. That is,

$$\sum_{i\in\mathcal{E}}E_i^2 + \sum_{m\leq n\in\mathcal{E}^2}(2EE)_m = E^2 - \sum_{m>n\in\mathcal{E}^2}(2EE)_m \leq E^2\,. \tag{25}$$

The distance then becomes

$$d_2(\mathcal{E},\mathcal{E}_2)^2 = \sum_{i<j\in\mathcal{E}} 2E_iE_j\omega_{ij}^2 + 2E_1E_2\omega_{12}^2 \tag{26}$$

$$
\begin{aligned}
&- 2 \sum_{\substack{n\in\mathcal{E}^2\\ \omega_n<\omega_{n+1}}} \omega_n\omega_{12}\Bigg(E^2 - \sum_{m>n\in\mathcal{E}^2}(2EE)_m \\
&\qquad\qquad - \max\left[E^2 - \sum_{m\geq n\in\mathcal{E}^2}(2EE)_m, E^2-2E_1E_2\right]\Bigg) \\
&\qquad \times \Theta\Bigg(E^2 - \sum_{m>n\in\mathcal{E}^2}(2EE)_m \\
&\qquad\qquad - \max\left[E^2 - \sum_{m\geq n\in\mathcal{E}^2}(2EE)_m, E^2-2E_1E_2\right]\Bigg)\,.
\end{aligned}
$$

There are then two possibilities for the remaining constraints:

$$
\begin{aligned}
d_2(\mathcal{E},\mathcal{E}_2)^2 = &\sum_{i<j\in\mathcal{E}} 2E_iE_j\omega_{ij}^2 + 2E_1E_2\omega_{12}^2 \\
&- 2\sum_{\substack{n\in\mathcal{E}^2\\ \omega_n<\omega_{n+1}}} \omega_n\omega_{12}\Bigg[(2EE)_n\,\Theta\left(2E_1E_2 - \sum_{m\geq n\in\mathcal{E}^2}(2EE)_m\right) \\
+ &\left(2E_1E_2 - \sum_{m>n\in\mathcal{E}^2}(2EE)_m\right)\Theta\left(\sum_{m\geq n\in\mathcal{E}^2}(2EE)_m - 2E_1E_2\right)\Theta\left(2E_1E_2 - \sum_{m>n\in\mathcal{E}^2}(2EE)_m\right)\Bigg]\,.
\end{aligned}
\tag{27}
$$

To minimize this distance and to therefore define a 2-pronginess observable, we take derivatives with respect to the phase space coordinates and set them equal to 0. For the angle ω_{12}, the derivative requires

$$
\begin{aligned}
&\frac{d}{d\omega_{12}} d_2(\mathcal{E},\mathcal{E}_2)^2 = 0 \\
&\to 2E_1E_2\omega_{12} = \sum_{\substack{n\in\mathcal{E}^2\\ \omega_n<\omega_{n+1}}} \omega_n\Bigg[(2EE)_n\,\Theta\left(2E_1E_2 - \sum_{m\geq n\in\mathcal{E}^2}(2EE)_m\right) \\
&\qquad + \left(2E_1E_2 - \sum_{m>n\in\mathcal{E}^2}(2EE)_m\right)\Theta\left(\sum_{m\geq n\in\mathcal{E}^2}(2EE)_m - 2E_1E_2\right)\Theta\left(2E_1E_2 - \sum_{m>n\in\mathcal{E}^2}(2EE)_m\right)\Bigg]\,.
\end{aligned}
\tag{28}
$$

In principle, we would also take derivatives with respect to $2E_1E_2$ and demand that they vanish to establish the minimum distance observable. However, while first derivatives with respect to $2E_1E_2$ exist, second derivatives do not, because there is $2E_1E_2$ dependence in Θ-functions. As such, points where the derivative wrt $2E_1E_2$ vanish are not guaranteed to be minima. So, only derivatives wrt ω_{12} can be used to constrain the location of the minimum.

So, using Eq. 28, note that the distance simplifies to

$$\min_{\omega_{12},2E_1E_2} d_2(\mathcal{E},\mathcal{E}_2)^2 = \sum_{i<j\in\mathcal{E}} 2E_iE_j\omega_{ij}^2 - 2E_1E_2\omega_{12}^2\,. \tag{29}$$

The angle ω_{12}, in terms of $2E_1E_2$, is also

$$\begin{aligned}\omega_{12} = \sum_{\substack{n\in\mathcal{E}^2\\ \omega_n<\omega_{n+1}}} \omega_n \left[\frac{(2EE)_n}{2E_1E_2}\,\Theta\left(2E_1E_2 - \sum_{m\geq n\in\mathcal{E}^2}(2EE)_m\right)\right.\\ \left. +\left(1-\sum_{m>n\in\mathcal{E}^2}\frac{(2EE)_m}{2E_1E_2}\right)\Theta\left(\sum_{m\geq n\in\mathcal{E}^2}(2EE)_m - 2E_1E_2\right)\Theta\left(2E_1E_2-\sum_{m>n\in\mathcal{E}^2}(2EE)_m\right)\right].\end{aligned} \tag{30}$$

So, to evaluate Eq. 29, we scan over physical values of $2E_1E_2$, where, for a jet with two particles, $0 \leq 2E_1E_2 \leq E^2/2$, evaluate ω_{12} according to Eq. 30, and find the corresponding minimum value of $d_2(\mathcal{E},\mathcal{E}_2)^2$.

I think that the value of $2E_1E_2$ can further be expressed in a nice way. This energy needs to be the total squared energy of the spikes located at large values of ω; namely,

$$2E_1E_2 = \sum_{n\leq m\in\mathcal{E}^2}(2EE)_m\,, \tag{31}$$

for some position n. That is, the spikes at large ω all need to be transported to the angle ω_{12}, while the spikes with index $m < n$ are transported to $\omega = 0$. Then, with this construction, the position ω_{12} simplifies to **(correct? –ajl)**

$$\omega_{12} = \sum_{\substack{n\leq m\in\mathcal{E}^2\\ \omega_m<\omega_{m+1}}} \omega_m\frac{(2EE)_m}{2E_1E_2}\,, \tag{32}$$

simply the mean spike location in this large-ω region. Then, the 2-pronginess observable can be expressed as

$$\min_{\omega_{12},2E_1E_2} d_2(\mathcal{E},\mathcal{E}_2)^2 = \min_n\left[\sum_{i<j\in\mathcal{E}} 2E_iE_j\omega_{ij}^2 - \frac{\left(\sum_{\substack{n\leq m\in\mathcal{E}^2\\ \omega_m<\omega_{m+1}}}\omega_m(2EE)_m\right)^2}{\sum_{n\leq m\in\mathcal{E}^2}(2EE)_m}\right]. \tag{33}$$

A jet with only two particles has a maximum value for $2E_1E_2$; namely, $2E_1E_2 \leq E^2/2$ which is maximized when each particle has energy equal to half of the jet's energy. If $2E_1E_2 > E^2/2$, then necessarily this is a degenerate configuration with multiple pairs of particles with the same angular separation. So, to prohibit obvious degenerate configurations, we must further enforce an upper bound on $2E_1E_2$, so

$$\min_{\omega_{12},2E_1E_2} d_2(\mathcal{E},\mathcal{E}_2)^2 = \min_{\substack{n \\ \sum\limits_{n\leq m\in\mathcal{E}^2}(2EE)_m\leq E^2/2}} \left[\sum_{i<j\in\mathcal{E}} 2E_iE_j\omega_{ij}^2 - \frac{\left(\sum\limits_{\substack{n\leq m\in\mathcal{E}^2\\ \omega_m<\omega_{m+1}}}\omega_m(2EE)_m\right)^2}{\sum\limits_{n\leq m\in\mathcal{E}^2}(2EE)_m}\right]\,. \tag{34}$$

2.3 N-Pronginess

(partition problem: for N-pronginess, there are $\binom{N}{2}$ peaks that divide the interval $[0,\omega_{\max}]$ into $\binom{N}{2}+1$ regions. Region 1, the leftmost region, needs to be transported to the origin, region 2 to the its mean angle, etc. Then, choose the boundaries to minimize the total transportation cost. –ajl)

3 Geometric Probability

(main reference: "Geometric Probability" by H. Solomon –ajl)

3.1 Event Isotropy

Consider a continuous energy distribution on the celestial sphere that is perfectly isotropic. Such a configuration of energy consists of an infinite number of particles, $N\to\infty$. The cumulative spectral function of angular distances ω on the sphere is

$$S_{\text{iso}}(\omega) = \frac{1-\cos\omega}{2}E_{\text{tot}}^2\,, \tag{35}$$

for a total energy E_{tot}. The inverse cumulative spectral function is then

$$S_{\text{iso}}^{-1}(E^2) = \cos^{-1}\left(1-2\frac{E^2}{E_{\text{tot}}^2}\right)\,. \tag{36}$$

The integral of its square is

$$\int_0^{E_{\text{tot}}^2} dE^2\, S_{\text{iso}}^{-1}(E^2)^2 = \frac{\pi^2-4}{2}\,E_{\text{tot}}^2\,. \tag{37}$$

Next, we need to integrate this inverse cumulative spectral function against the inverse cumulative spectral function for the event of interest. We have

$$\int dE^2\, S^{-1}(E^2)S_{\text{iso}}^{-1}(E^2) = \sum_{\substack{n\in\mathcal{E}^2\\ \omega_n<\omega_{n+1}}}\omega_n\int dE^2\,\cos^{-1}\left(1-2\frac{E^2}{E_{\text{tot}}^2}\right) \tag{38}$$

$$\times\, \Theta\left(\sum_i E_i^2 + \sum_{m\leq n}(2EE)_m - E^2\right)\Theta\left(E^2 - \sum_i E_i^2 - \sum_{m<n}(2EE)_m\right).$$

(calculate the inverse cumulative spectral function for uniform energy distribution on celestial sphere –ajl)

3.2 Event Ringiness

For continuous energy distribution on the equator of the celestial sphere, the cumulative spectral function is

$$S_{\text{ring}}(\omega) = \frac{\omega}{\pi}\, E_{\text{tot}}^2\,. \tag{39}$$

The inverse cumulative spectral function is then

$$S_{\text{ring}}^{-1}(E^2) = \pi\, \frac{E^2}{E_{\text{tot}}^2}\,. \tag{40}$$

The integral of the square of this inverse cumulative spectral function is

$$\int_0^{E_{\text{tot}}^2} dE^2\, S_{\text{ring}}^{-1}(E^2)^2 = \frac{\pi^2}{3}\, E_{\text{tot}}^2\,. \tag{41}$$

Next, we need to integrate this inverse cumulative spectral function against the inverse cumulative spectral function for the event of interest. We have

$$\begin{aligned}
\int dE^2\, S^{-1}(E^2) S_{\text{ring}}^{-1}(E^2) &= \sum_{\substack{n\in\mathcal{E}^2\\ \omega_n<\omega_{n+1}}} \omega_n \int dE^2\, \pi\, \frac{E^2}{E_{\text{tot}}^2} \\
&\quad\times \Theta\left(\sum_i E_i^2 + \sum_{m\leq n}(2EE)_m - E^2\right)\Theta\left(E^2 - \sum_i E_i^2 - \sum_{m<n}(2EE)_m\right) \\
&= \frac{\pi}{2}\frac{1}{E_{\text{tot}}^2} \sum_{\substack{n\in\mathcal{E}^2\\ \omega_n<\omega_{n+1}}} \omega_n \left[2E_{\text{tot}}^2(2EE)_n + \left(\sum_{m>n}(2EE)_m\right)^2 + \left(\sum_{m\geq n}(2EE)_m\right)^2\right].
\end{aligned} \tag{42}$$

Then, the event ringiness observable is

$$\begin{aligned}
d_2(\mathcal{E},\mathcal{E}_{\text{ring}})^2 &= \sum_{i<j\in\mathcal{E}} 2E_iE_j\omega_{ij}\left(\omega_{ij} - 2\pi\right) + \frac{\pi^2}{3}\, E_{\text{tot}}^2 \\
&\quad - \frac{\pi}{E_{\text{tot}}^2} \sum_{\substack{n\in\mathcal{E}^2\\ \omega_n<\omega_{n+1}}} \omega_n \left[\left(\sum_{m>n}(2EE)_m\right)^2 + \left(\sum_{m\geq n}(2EE)_m\right)^2\right].
\end{aligned} \tag{43}$$

(finish this –ajl)

3.3 Jet Ringiness

Consider a continuous distribution of energy distributed in the shape of a ring of angular radius R in a jet. Then, the cumulative spectral function of such an energy configuration is

$$S_{\text{ring}}(\omega) = \frac{2E_{\text{tot}}^2}{\pi} \sin^{-1}\left(\frac{\omega}{2R}\right) . \tag{44}$$

The inverse cumulative spectral function is therefore

$$S_{\text{ring}}^{-1}(E^2) = 2R \sin\left(\frac{\pi}{2}\frac{E^2}{E_{\text{tot}}^2}\right) . \tag{45}$$

The integral of its square is

$$\int dE^2\, S_{\text{ring}}^{-1}(E^2)^2 = 2E_{\text{tot}}^2 R^2 . \tag{46}$$

Next, we need to integrate this inverse cumulative spectral function against the inverse cumulative spectral function for the event of interest. We have

$$\begin{aligned}
\int dE^2\, S^{-1}(E^2) S_{\text{ring}}^{-1}(E^2) &= 2R \sum_{\substack{n\in\mathcal{E}^2 \\ \omega_n<\omega_{n+1}}} \omega_n \int dE^2 \, \sin\left(\frac{\pi}{2}\frac{E^2}{E_{\text{tot}}^2}\right) \\
&\quad \times \Theta\left(\sum_i E_i^2 + \sum_{m\leq n}(2EE)_m - E^2\right)\Theta\left(E^2 - \sum_i E_i^2 - \sum_{m<n}(2EE)_m\right) \\
&= \frac{4R}{\pi} E_{\text{tot}}^2 \sum_{\substack{n\in\mathcal{E}^2 \\ \omega_n<\omega_{n+1}}} \omega_n \left[\sin\left(\frac{\pi}{2E_{\text{tot}}^2}\sum_{m\geq n}(2EE)_m\right) - \sin\left(\frac{\pi}{2E_{\text{tot}}^2}\sum_{m>n}(2EE)_m\right)\right] .
\end{aligned} \tag{47}$$

Then, the jet ringiness observable is

$$\begin{aligned}
d_2(\mathcal{E},\mathcal{E}_{\text{ring}})^2 &= \sum_{i<j\in\mathcal{E}} 2E_iE_j\omega_{ij}^2 + 2E_{\text{tot}}^2 R^2 \\
&\quad - \frac{8R}{\pi} E_{\text{tot}}^2 \sum_{\substack{n\in\mathcal{E}^2 \\ \omega_n<\omega_{n+1}}} \omega_n \left[\sin\left(\frac{\pi}{2E_{\text{tot}}^2}\sum_{m\geq n}(2EE)_m\right) - \sin\left(\frac{\pi}{2E_{\text{tot}}^2}\sum_{m>n}(2EE)_m\right)\right] .
\end{aligned} \tag{48}$$

3.4 Jet Lineness

For a continuous energy distribution arranged along a line of length L in a jet, its cumulative spectral function is

$$S_{\text{line}}(\omega) = \left(\frac{2\omega}{L} - \frac{\omega^2}{L^2}\right) E_{\text{tot}}^2 . \tag{49}$$

The inverse cumulative spectral function is then

$$S_{\text{line}}^{-1}(E^2) = \left(1 - \sqrt{1 - \frac{E^2}{E_{\text{tot}}^2}}\right) L\,. \tag{50}$$

Its squared integral is then

$$\int dE^2\, S_{\text{line}}^{-1}(E^2)^2 = \frac{L^2}{6}\, E_{\text{tot}}^2\,. \tag{51}$$

Next, we need to integrate this inverse cumulative spectral function against the inverse cumulative spectral function for the event of interest. We have

$$\begin{aligned}
\int dE^2\, S^{-1}(E^2) S_{\text{line}}^{-1}(E^2) &= L \sum_{\substack{n\in\mathcal{E}^2\\ \omega_n<\omega_{n+1}}} \omega_n \int dE^2 \left(1 - \sqrt{1 - \frac{E^2}{E_{\text{tot}}^2}}\right) \\
&\quad \times \Theta\left(\sum_i E_i^2 + \sum_{m\leq n} (2EE)_m - E^2\right) \Theta\left(E^2 - \sum_i E_i^2 - \sum_{m<n} (2EE)_m\right) \\
&= L \sum_{\substack{n\in\mathcal{E}^2\\ \omega_n<\omega_{n+1}}} \omega_n \left[(2EE)_n + \frac{2}{3E_{\text{tot}}}\left(\sum_i E_i^2 + \sum_{m\leq n}(2EE)_m\right)^{3/2}\right. \\
&\qquad \left. - \frac{2}{3E_{\text{tot}}}\left(\sum_i E_i^2 + \sum_{m<n}(2EE)_m\right)^{3/2}\right].
\end{aligned} \tag{52}$$

Then, the jet lineliness observable is

$$\begin{aligned}
d_2(\mathcal{E},\mathcal{E}_{\text{line}})^2 &= \sum_{i<j\in\mathcal{E}} 2E_iE_j\omega_{ij}^2 + \frac{L^2}{6}\, E_{\text{tot}}^2 \\
&\quad - 2L \sum_{\substack{n\in\mathcal{E}^2\\ \omega_n<\omega_{n+1}}} \omega_n \left[(2EE)_n + \frac{2}{3E_{\text{tot}}}\left(\sum_i E_i^2 + \sum_{m\leq n}(2EE)_m\right)^{3/2}\right. \\
&\qquad \left. - \frac{2}{3E_{\text{tot}}}\left(\sum_i E_i^2 + \sum_{m<n}(2EE)_m\right)^{3/2}\right]
\end{aligned} \tag{53}$$

3.5 Uniform on a Disk/Over Jet Area

The spectral function for uniform radiation on a disk of radius R encodes the energy density of two points on the disk are a distance ω apart. This is a well-known problem, (ball line picking), whose solution is **(refs? –ajl)**

$$s_{\text{disk}}(\omega) = \frac{4\omega E_{\text{tot}}^2}{\pi R^2}\cos^{-1}\frac{\omega}{2R} - \frac{2\omega^2 E_{\text{tot}}^2}{\pi R^3}\sqrt{1 - \frac{\omega^2}{4R^2}}\,. \tag{54}$$

This can be integrated to establish the cumulative spectral function. The integral of the square of the cumulative spectral function is, from Eq. 6,

$$\int dE^2\, S_{\text{disk}}^{-1}(E^2)^2 = \int_0^{2R} d\omega\, \omega^2 s_{\text{disk}}(\omega) = E_{\text{tot}}^2 R^2\,. \tag{55}$$

Next, we need to evaluate the cross term integral. We have

$$\begin{aligned}\int dE^2\, S^{-1}(E^2)\, S_{\text{disk}}^{-1}(E^2) &= \int dE^2\, S^{-1}(E^2) \int d\omega\, \Theta\left(E^2 - S_{\text{disk}}(\omega)\right) \\ &= \sum_{\substack{n\in\mathcal{E}^2\\ \omega_n<\omega_{n+1}}} \omega_n \int_0^{2R} d\omega\, \left(E_{\text{tot}}^2 - \sum_{n<m\in\in\mathcal{E}^2} (2EE)_m \right.\\ &\qquad\left. - \max\left[E_{\text{tot}}^2 - \sum_{n\leq m\in\mathcal{E}^2}(2EE)_m, S_{\text{disk}}(\omega)\right]\right)\Theta(\cdot)\,,\end{aligned} \tag{56}$$

where $\Theta(\cdot)$ is a placeholder for demanding that the integral is positive. There are then two cases, so the integral takes the form

$$\begin{aligned}\int dE^2\, S^{-1}(E^2)\, S_{\text{disk}}^{-1}(E^2) &= \sum_{\substack{n\in\mathcal{E}^2\\ \omega_n<\omega_{n+1}}} \omega_n (2EE)_n \int_0^{2R} d\omega\, \Theta\left(E_{\text{tot}}^2 - \sum_{n\leq m\in\mathcal{E}^2}(2EE)_m - S_{\text{disk}}(\omega)\right) \\ &+ \sum_{\substack{n\in\mathcal{E}^2\\ \omega_n<\omega_{n+1}}} \omega_n \int_0^{2R} d\omega\, \left(E_{\text{tot}}^2 - \sum_{n<m\in\mathcal{E}^2}(2EE)_m - S_{\text{disk}}(\omega)\right) \\ &\times \Theta\left(E_{\text{tot}}^2 - \sum_{n<m\in\mathcal{E}^2}(2EE)_m - S_{\text{disk}}(\omega)\right)\Theta\left(S_{\text{disk}}(\omega) - E_{\text{tot}}^2 + \sum_{n\leq m\in\mathcal{E}^2}(2EE)_m\right)\,.\end{aligned}$$

(do this; can we go any further? –ajl)

3.6 Cylindrical Event Isotropy

The spectral function for radiation uniform on a cylinder with pseudorapidity range η_{max} can be calculated from

$$\begin{aligned}s_{\text{cyl}}(\omega) &= \frac{E_{\text{tot}}^2}{2\pi\eta_{\text{max}}^2}\int_0^{\pi} d\phi \int_{-\eta_{\text{max}}}^{\eta_{\text{max}}} d\eta_1 \int_{-\eta_{\text{max}}}^{\eta_1} d\eta_2\, \delta\left(\omega - \sqrt{\phi^2 + (\eta_1-\eta_2)^2}\right) \\ &= \frac{E_{\text{tot}}^2\,\omega}{2\pi\eta_{\text{max}}^2}\int_0^{2\eta_{\text{max}}} d\Delta\eta \int_{\Delta\eta-\eta_{\text{max}}}^{\eta_{\text{max}}} d\eta_1\, \frac{\Theta(\omega-\Delta\eta)\Theta\left(\pi^2 - \sqrt{\omega^2-\Delta\eta^2}\right)}{\sqrt{\omega^2-\Delta\eta^2}} \\ &= \frac{E_{\text{tot}}^2\,\omega}{2\pi\eta_{\text{max}}^2}\int_0^{2\eta_{\text{max}}} d\Delta\eta\, \frac{2\eta_{\text{max}}-\Delta\eta}{\sqrt{\omega^2-\Delta\eta^2}}\,\Theta(\omega-\Delta\eta)\Theta\left(\pi^2 - \sqrt{\omega^2-\Delta\eta^2}\right)\end{aligned} \tag{57}$$

(this is a rectangular-line picking problem, where the rectangle has dimensions $\pi \times 2\eta_{\max}$ and one point is confined on one of the sides of length $2\eta_{\max}$. –ajl)

(finish this... –ajl)

3.7 Subtleties: Jet Equilateral Triangularness

4 Pile-Up Subtraction

We can define a similar pile-up subtraction method as used by SHAPER.

4.1 1-Pronginess

We would like to measure 1-pronginess on a jet on which there is some soft, uniform radiation over the area of the jet. Our pile-up subtraction scheme will therefore be to find the minimum distance between the spectral function of the jet and the spectral function of an ideal one-prong jet plus the spectral function for soft uniform radiation over the area of the jet. To lowest order in the energy Λ of the contamination, this reference spectral function can be expressed as

$$s_{\text{1-prong+PU}}(\omega) = (E_{\text{tot}}^2 - E_{\text{tot}}\Lambda)\,\delta(\omega) + \frac{2E_{\text{tot}}\Lambda}{R^2}\,\omega\,, \tag{58}$$

where R is the jet radius and E_{tot} is the total energy of the measured jet. This spectral function therefore has a single parameter, Λ. The cumulative spectral function is then

$$S_{\text{1-prong+PU}}(\omega) = E_{\text{tot}}^2 - E_{\text{tot}}\Lambda + \frac{E_{\text{tot}}\Lambda}{R^2}\,\omega^2\,, \tag{59}$$

and therefore its inverse is

$$S^{-1}_{\text{1-prong+PU}}(E^2) = \frac{R}{\sqrt{E_{\text{tot}}\Lambda}}\left(E^2 - E_{\text{tot}}^2 + E_{\text{tot}}\Lambda\right)^{1/2}\Theta(E^2 - E_{\text{tot}}^2 + E_{\text{tot}}\Lambda)\,. \tag{60}$$

The integral of the square of the inverse cumulative distribution is

$$\int dE^2\, S^{-1}_{\text{1-prong+PU}}(E^2)^2 = \frac{E_{\text{tot}}\Lambda R^2}{2}\,. \tag{61}$$

The integral of the inverse cumulative spectral function against the inverse cumulative spectral function of the jet of interest is

$$\int dE^2\, S^{-1}_{\text{1-prong+PU}}(E^2)\, S^{-1}(E^2) = \frac{R}{\sqrt{E_{\text{tot}}\Lambda}} \sum_{\substack{n\in\mathcal{E}^2\\ \omega_n<\omega_{n+1}}} \omega_n \int dE^2\, \left(E^2 - E_{\text{tot}}^2 + E_{\text{tot}}\Lambda\right)^{1/2} \tag{62}$$
$$\times\,\Theta(E^2 - E_{\text{tot}}^2 + E_{\text{tot}}\Lambda)\,\Theta\left(E_{\text{tot}}^2 - \sum_{n<m\in\mathcal{E}^2}(2EE)_m - E^2\right)\Theta\left(E^2 - E_{\text{tot}}^2 + \sum_{n\le m\in\mathcal{E}^2}(2EE)_m\right)$$

$$
\begin{aligned}
&= \frac{2}{3}\frac{R}{\sqrt{E_{\mathrm{tot}}\Lambda}} \sum_{\substack{n\in\mathcal{E}^2\\ \omega_n<\omega_{n+1}}} \omega_n \left(E_{\mathrm{tot}}\Lambda - \sum_{n<m\in\mathcal{E}^2}(2EE)_m\right)^{3/2} \Theta\left(E_{\mathrm{tot}}\Lambda - \sum_{n<m\in\mathcal{E}^2}(2EE)_m\right) \\
&\quad - \frac{2}{3}\frac{R}{\sqrt{E_{\mathrm{tot}}\Lambda}} \sum_{\substack{n\in\mathcal{E}^2\\ \omega_n<\omega_{n+1}}} \omega_n \left(E_{\mathrm{tot}}\Lambda - \sum_{n\leq m\in\mathcal{E}^2}(2EE)_m\right)^{3/2} \Theta\left(E_{\mathrm{tot}}\Lambda - \sum_{n\leq m\in\mathcal{E}^2}(2EE)_m\right) .
\end{aligned}
$$