

MIT Open Access Articles

SPECTER: efficient evaluation of the spectral EMD

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Gambhir, R., Larkoski, A.J. & Thaler, J. SPECTER: efficient evaluation of the spectral EMD. J. High Energ. Phys. 2024, 219 (2025).

As Published: [https://doi.org/10.1007/JHEP12\(2024\)219](https://doi.org/10.1007/JHEP12(2024)219)

Publisher: Springer Berlin Heidelberg

Persistent URL: <https://hdl.handle.net/1721.1/157951>

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of use: Creative Commons Attribution



SPECTER: efficient evaluation of the spectral EMD

Rikab Gambhir ^{a,b}, Andrew J. Larkoski ^{c,d} and Jesse Thaler ^{a,b}

^a*Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.*

^b*The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, Cambridge, MA, U.S.A.*

^c*Department of Physics and Astronomy, University of California, Los Angeles, CA 90095, U.S.A.*

^d*Mani L. Bhaumik Institute for Theoretical Physics, University of California, Los Angeles, CA 90095, U.S.A.*

E-mail: rikab@mit.edu, larkoa@gmail.com, jthaler@mit.edu

ABSTRACT: The Energy Mover’s Distance (EMD) has seen use in collider physics as a metric between events and as a geometric method of defining infrared and collinear safe observables. Recently, the Spectral Energy Mover’s Distance (SEMD) has been proposed as a more analytically tractable alternative to the EMD. In this work, we obtain a closed-form expression for the Riemannian-like $p = 2$ SEMD metric between events, eliminating the need to numerically solve an optimal transport problem. Additionally, we show how the SEMD can be used to define event and jet shape observables by minimizing the distance between events and parameterized energy flows (similar to the EMD), and we obtain closed-form expressions for several of these observables. We also present the SPECTER framework, an efficient and highly parallelized implementation of the SEMD metric and SEMD-derived shape observables as an analogue of the previously-introduced SHAPER for EMD-based computations. We demonstrate that computing the SEMD with SPECTER can be up to a thousand times faster than computing the EMD with standard optimal transport libraries.

KEYWORDS: Jets and Jet Substructure, Specific QCD Phenomenology

ARXIV EPRINT: [2410.05379](https://arxiv.org/abs/2410.05379)

Contents

1	Introduction	1
2	The $p = 2$ Spectral EMD	3
2.1	Review of the Spectral EMD	3
2.2	Closed-form evaluation	5
2.3	Choice of ground metric	8
2.4	Comparison to the EMD and degeneracies	8
3	The SPECTER framework	11
3.1	Algorithm for the Spectral EMD	11
3.2	Runtime benchmarks	13
4	Spectral shapes	15
4.1	General aspects	15
4.2	Spectral prong shapes	17
4.3	Spectral jet shapes	19
4.4	Spectral event shapes	22
4.5	... and more!	26
5	Empirical studies	26
5.1	Datasets	27
5.2	Pairwise distances between events and jets	28
5.3	Hearing jets spRing and sProng	30
5.4	Event shapes	32
6	Conclusion	37
A	More about sThrust: chord lengths and energy fractions	39
B	Closed-form expressions for spectral observables	41
B.1	Closed-form jet spRinginess	41
B.2	Closed-form spLineliness	42
B.3	Closed-form spIsotropy	42
B.4	Closed-form event spEquatorialness	43

1 Introduction

Events from particle physics collisions live in a space rich with structure, whose dimensionality, topology, curvature, and other geometric properties encode the myriad physical phenomena that can be observed. A first step to quantifying the features of this space is to define a metric distance between events, and further ideally in an infrared and collinear (IRC) safe way to render the parametrization of the space calculable in perturbative quantum field theory. The first IRC-safe metric on the space of collider events was constructed in ref. [1] and called the

Energy Mover’s Distance (EMD), which quantified the minimal energy cost of deforming one event into another. The EMD has since sparked significant interest, with numerous other metrics proposed or exploiting properties of the EMD in various contexts [2–15]. This includes the SHAPER framework [16], which can be used to construct and evaluate generalized EMD-based event and jet shapes. EMD-based observables have also received heightened interest in the experimental community — one such observable, the event Isotropy [5, 17], was recently measured at the Large Hadron Collider (LHC) by the ATLAS collaboration [18].

The EMD is formulated as the Wasserstein metric [19–21] of an optimal transport problem in which energy deposits from one event are moved around the detector to reproduce the energy deposits of another event. An idealized particle physics detector lives on the celestial sphere, so this optimal transport problem is defined in two dimensions. Because objects in two or higher dimensions lack a natural ordering, there is no general closed-form expression for the EMD as a function of event constituents. Thus, generically, this distance must be evaluated through some numerical optimization. This irreducible numerical computation obscures analytical understanding of the EMD, except in the simplest of event configurations, and further can suffer from slow evaluation, though there are now efficient codes for optimal transport in two or more dimensions [16, 22].

With these points as motivation, the Spectral EMD (SEMD) was introduced recently [23]. Instead of optimal transport on the celestial sphere, the SEMD is defined through optimal transport between spectral function representations of events [24–28]. Spectral functions are one-dimensional distributions that encode pairwise particle angles and energy products. As such, the spectral representation is automatically invariant to isometries of the events being compared. Further, a theorem of ref. [29] ensures that the spectral representation can be, with probability 1, mapped to a unique distribution of particles on the celestial sphere, up to isometries. This result establishes that the SEMD is not only an approximation to the EMD or only approximately a metric, but is indeed a genuine metric between collision events satisfying identity of indiscernibles, symmetry, and the triangle inequality. Crucially, because points in one dimension are well ordered, optimal transport in one dimension can be expressed as an integral of the difference of the spectral quantile functions (inverse cumulative spectral functions), which can correspondingly be evaluated very efficiently.

Nevertheless, ref. [23] mostly focused on the $p = 1$ Wasserstein metric of the spectral representation. While simple and motivated historically by the Monge-Kantorovich problem [30] and the Earth Mover’s Distance from computer vision [31–34], direct evaluation of this integral between arbitrary events is challenging because its integrand is expressed as the absolute value of a difference of functions. Thus, in practice, the $p = 1$ SEMD is, like the ordinary EMD, most easily numerically evaluated. Further, $p = 1$ Wasserstein metrics lack properties that, as physicists, we have come to expect of metrics, namely, their direct interpretability as differential line elements with a natural Riemannian structure. Simply modifying the definition of the SEMD to a $p = 2$ Wasserstein metric endows the SEMD with many nice properties familiar from Riemannian metrics [4, 35–37].

In this paper, we flesh out the $p = 2$ SEMD metric, establish its remarkable properties, and introduce a computational framework called SPECTER for its efficient evaluation. First and foremost, the $p = 2$ SEMD can be expressed exactly in closed-form between any two

events as represented as point clouds on the celestial sphere. This represents the *first* metric on collider events that can be expressed with no approximations or need for numerical evaluation, and correspondingly renders theoretical calculations potentially feasible. Novel event and jet shape observables can be constructed from the SEMD as defined by the minimal distance from an event of interest to a desired energy flow configuration as in refs. [16, 38], and essentially becomes a problem in geometric probability [39–41], formulated as shape-line picking. As such, many observables constructed from the SEMD enjoy a closed-form expression as well. Moreover, we argue that SEMD and derived observables capture the same QCD jet physics as do the EMD and its observables, potentially enabling high-speed precision studies of QCD, while being complementary to the EMD for non-QCD physics.

To facilitate these computations, we construct the SPECTER framework, available at <https://github.com/rikab/SPECTER>. SPECTER is to the SEMD what SHAPER is to the EMD: an efficient program for evaluation of the $p = 2$ SEMD on events from data (real or simulated), and for the construction and evaluation of *all* spectral event and jet shape observables.¹ Due to the closed-form expression and well-ordering in one dimension, SPECTER evaluates the SEMD between events hundreds or even thousands of times faster than comparable codes can evaluate the EMD. On an ordinary GPU architecture, SPECTER can perform 1 million metric evaluations per second, making it far more feasible to compute distances between all pairs of events (to compute correlation dimensions [42]), perform gradient descent (for parameter optimization [16]), or experimental studies (potentially even at the trigger level).

The organization of this paper is as follows. In section 2, we review the spectral representation of events, evaluate the $p = 2$ SEMD, and discuss its properties. In section 3, we introduce the SPECTER framework and discuss the algorithm to efficiently evaluate the SEMD between events with N particles in only $\mathcal{O}(N^2 \log N)$ time compared to $\mathcal{O}(N^3)$ for the EMD, and demonstrate performance through some runtime benchmarks. In section 4, we review how to construct event and jet shape observables from a metric formulation, and present a collection of observables defined from the SEMD useful for identifying structure on both complete events as well as on individual jets. In section 5, we study the SEMD and derivative observables on simulated events, providing comparisons where possible to the EMD. We conclude in section 6, and present some more details of ground metric choices and closed-form expressions in the appendices.

2 The $p = 2$ Spectral EMD

In this section, we provide a review of the spectral function, define the general SEMD metric from it, and then specialize to the Riemannian $p = 2$ case, for which the SEMD metric can be written in closed form. Because the explicit form of the $p = 2$ SEMD metric is new, we present a detailed discussion of its derivation, but point the reader to ref. [23] for more details about the spectral function and its one-dimensional optimal transport.

2.1 Review of the Spectral EMD

The spectral function $s(\omega)$ is a one-dimensional representation of a particle collision event or jet. It is a distribution of pairwise particle angles ω , weighted by the product of energies

¹Unlike SHAPER, though, SPECTER does not correspond to some tortured acronym.

of the particles in each pair:

$$s(\omega) \equiv \sum_{i,j \in \mathcal{E}} E_i E_j \delta(\omega - \omega_{ij}) = \sum_{i \in \mathcal{E}} E_i^2 \delta(\omega) + \sum_{i < j \in \mathcal{E}} 2E_i E_j \delta(\omega - \omega_{ij}), \quad (2.1)$$

where \mathcal{E} is the set of particles of interest, ω_{ij} is an appropriately-defined angle or distance between particles i and j , and E_i is the energy of particle i . Many plots showing how a spectral function can be constructed from an event or jet will be shown in sections 4 and 5. We also define the cumulative spectral function:

$$S(\omega) = \int_0^\omega d\omega' s(\omega'), \quad (2.2)$$

and the inverse cumulative spectral function $S^{-1}(E^2)$. The spectral function is normalized to integrate to the total squared energy:

$$E_{\text{tot}}^2 = S(\omega_{\text{max}}). \quad (2.3)$$

The spectral function is IRC safe because it is multi-linear in particle energies and the arguments of the δ -functions are purely functions of angle (with no energy weighting). It is also invariant to isometries of the ground space, such as rotations about the colliding beam directions or permutations of particle labels, because it is sensitive only to relative angles. The spectral function almost uniquely defines the positions of particles on the celestial sphere, up to isometries, because both relative angles and particle energies are drawn from continuous distributions on phase space [29].² By “almost uniquely”, we mean that there is 0 probability that a spectral function corresponds to two or more distinct, non-isometric distributions of particles on the celestial sphere. This is necessary to define a metric — the distance between spectral functions should be always zero if and only if they are the same spectral function, and thus the distance between their corresponding events being zero implies they are *almost* always the same event. Note that the existence of degenerate configurations is highly nontrivial, and we discuss this point more in section 2.4.

Then, as the fundamental object that encodes complete information about the set of particles, a metric between two events A, B can be defined through the optimal transport of one event’s spectral function to the other’s. Further, because the spectral function is one-dimensional, the optimal transport cost can be expressed as a one-dimensional integral of the difference between inverse cumulative spectral functions:

$$\text{SEMD}_p(s_A, s_B) \equiv \int_0^{E_{\text{tot}}^2} dE^2 \left| S_A^{-1}(E^2) - S_B^{-1}(E^2) \right|^p, \quad (2.4)$$

for some $p \geq 1$. Technically, to be a metric (and satisfy the triangle inequality between three events A, B , and C), we further need to take the $1/p$ power of the result, but we instead work with this p th power to avoid carrying around clunky roots. $S_A^{-1}(E^2)$ is the inverse cumulative spectral function of event A , for which the cumulative spectral function is:

$$S(\omega) \equiv \int_0^\omega d\omega' s(\omega') = \sum_{i \in \mathcal{E}} E_i^2 + \sum_{i < j \in \mathcal{E}} 2E_i E_j \Theta(\omega - \omega_{ij}). \quad (2.5)$$

²This is in contrast to distributions on a regular lattice, where pairwise distances are likely to be non-unique.

In ref. [23], the $p = 1$ metric was studied in detail, which corresponds to the historical Monge-Kantorovich distance or Earth Mover's Distance [30] of the more general class of Wasserstein metrics [19–21]. In the $p = 1$ case only, the inverse cumulative spectral functions can be replaced by the cumulative spectral functions themselves:

$$\text{SEMD}_{p=1}(s_A, s_B) \equiv \int_0^{\omega_{\max}} d\omega |S_A(\omega) - S_B(\omega)|. \quad (2.6)$$

For this reason, the $p = 1$ metric was previously studied in detail, while $p > 1$ metrics (which require the inverse cumulative spectral function to be defined) were not. We next demonstrate, however, that this restriction was neither necessary nor especially convenient because we can rather easily exactly evaluate the SEMD integral for $p = 2$.

2.2 Closed-form evaluation

In this work, we focus exclusively on the $p = 2$ SEMD:

$$\begin{aligned} \text{SEMD}_{p=2}(s_A, s_B) &= \int_0^{E_{\text{tot}}^2} dE^2 \left| S_A^{-1}(E^2) - S_B^{-1}(E^2) \right|^2 \\ &= \int_0^{E_{\text{tot}}^2} dE^2 \left(S_A^{-1}(E^2)^2 + S_B^{-1}(E^2)^2 - 2S_A^{-1}(E^2)S_B^{-1}(E^2) \right). \end{aligned} \quad (2.7)$$

We assume that the two events A and B have the same total energy E_{tot} . (As discussed in ref. [23], this can be accomplished by adding any deficit energy to the less energetic event's spectral function a distance ω_R away, where ω_R is a chosen angular scale.) Here, the inverse cumulative spectral function can be evaluated from:

$$S^{-1}(E^2) = \int_0^{\omega_{\max}} d\omega \Theta(E^2 - S(\omega)). \quad (2.8)$$

For two events or jets A, B that each contain a discrete number of particles, the SEMD can be evaluated exactly in closed form. We consider each term in the integrand of eq. (2.7) separately, and then add them together appropriately to evaluate the spectral metric.

We begin by evaluating terms whose integrands are squares of the inverse cumulative spectral functions. To do these integrals, we make a change of variables from squared energy to cumulative spectral function:

$$E^2 = S(\omega). \quad (2.9)$$

We then find:

$$\int_0^{E_{\text{tot}}^2} dE^2 S^{-1}(E^2)^2 = \int_0^{\omega_{\max}} d\omega \omega^2 s(\omega), \quad (2.10)$$

which is just the second moment of the spectral function. This moment is trivial to evaluate because the angular dependence in the spectral function is in δ -functions:

$$\int_0^{E_{\text{tot}}^2} dE^2 S^{-1}(E^2)^2 = \sum_{i < j \in \mathcal{E}} 2E_i E_j \omega_{ij}^2, \quad (2.11)$$

where the sum runs over all particles $i < j$ in the event \mathcal{E} . This is just a two-point energy correlation function observable [43, 44] evaluated on the event of interest.

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

$$S(\omega) = \sum_{i \in \mathcal{E}} E_i^2 + \sum_{i < j \in \mathcal{E}} 2E_i E_j \Theta(\omega - \omega_{ij}) \equiv \sum_{i \in \mathcal{E}} E_i^2 + \sum_{\substack{n \in \mathcal{E}^2 \\ \omega_n < \omega_{n+1}}} (2EE)_n \Theta(\omega - \omega_n). \quad (2.12)$$

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. The notation $n \in \mathcal{E}^2 \equiv \mathcal{E} \times \mathcal{E}$ means that n represents a pair of particles from event \mathcal{E} . In the sum, we have ordered the angular factors so that $\omega_n < \omega_{n+1}$. We then define *inclusive* and *exclusive* cumulative distributions up to a peak $n \in \mathcal{E}^2$ as:

$$S^+(\omega_n) \equiv \lim_{\epsilon \rightarrow 0} S(\omega_n + \epsilon) = \sum_{i \in \mathcal{E}} E_i^2 + \sum_{\substack{n \geq m \in \mathcal{E}^2 \\ \omega_m < \omega_{m+1}}} (2EE)_m, \quad (2.13)$$

$$S^-(\omega_n) \equiv \lim_{\epsilon \rightarrow 0} S(\omega_n - \epsilon) = \sum_{i \in \mathcal{E}} E_i^2 + \sum_{\substack{n > m \in \mathcal{E}^2 \\ \omega_m < \omega_{m+1}}} (2EE)_m. \quad (2.14)$$

Note that $S^+(\omega_n)$ is the cumulative spectral function including the peak at $\omega = \omega_n$, while $S^-(\omega_n)$ does not include the peak at ω_n . Using this notation, the inverse cumulative distribution $S^{-1}(E^2)$ can be expressed as:

$$S^{-1}(E^2) = \sum_{\substack{n \in \mathcal{E}^2 \\ \omega_n < \omega_{n+1}}} \omega_n \Theta\left(S^+(\omega_n) - E^2\right) \Theta\left(E^2 - S^-(\omega_n)\right). \quad (2.15)$$

The proof of this statement follows simply by sketching the functional form of the cumulative spectral function as in figure 1, reflecting about the diagonal, and building the resulting function out of Θ -functions.

To express the mixed term, is it further convenient to define the quantity:

$$\mathcal{S}_{n\ell} \equiv \min\left[S_A^+(\omega_n), S_B^+(\omega_\ell)\right] - \max\left[S_A^-(\omega_n), S_B^-(\omega_\ell)\right], \quad (2.16)$$

where A and B refer to two *different* spectral functions, indexed by n and ℓ respectively. Integrating the product of spectral functions defined by eq. (2.15) and keeping careful track of the Θ -functions, the cross term integral in the SEMD can be written as:

$$\int_0^{E_{\text{tot}}^2} dE^2 S_A^{-1}(E^2) S_B^{-1}(E^2) = \sum_{\substack{n \in \mathcal{E}_A^2, \ell \in \mathcal{E}_B^2 \\ \omega_n < \omega_{n+1} \\ \omega_\ell < \omega_{\ell+1}}} \omega_n \omega_\ell \text{ReLU}(\mathcal{S}_{n\ell}), \quad (2.17)$$

where

$$\text{ReLU}(x) \equiv x \Theta(x), \quad (2.18)$$

is the Rectified Linear Unit function [45].

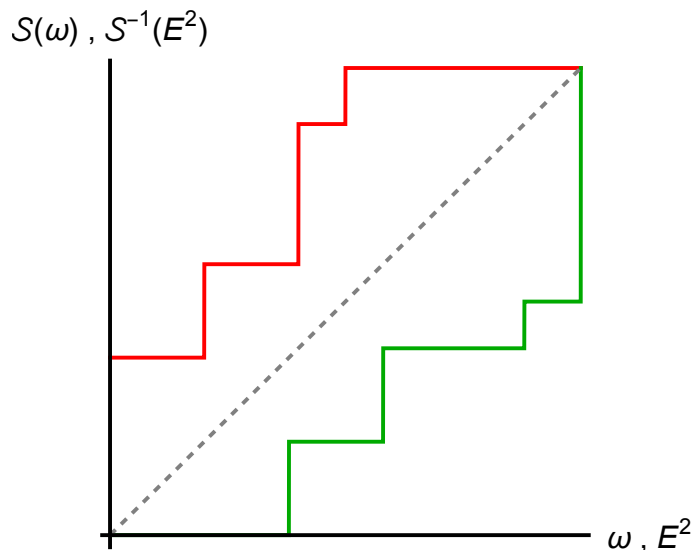


Figure 1. Illustration of the relationship between the cumulative spectral function and its inverse. The cumulative spectral function $S(\omega)$ (red) takes the value of a squared energy E^2 at every angle ω , while the inverse cumulative spectral function $S^{-1}(E^2)$ (green) takes the value of an angle at every squared energy.

Summing up the squared and mixed contributions yields a compact form for the SEMD:

$$\text{SEMD}_{\beta,p=2}(s_A, s_B) = \sum_{i < j \in \mathcal{E}_A} 2E_i E_j \omega_{ij}^2 + \sum_{i < j \in \mathcal{E}_B} 2E_i E_j \omega_{ij}^2 - 2 \sum_{\substack{n \in \mathcal{E}_A^2, \ell \in \mathcal{E}_B^2 \\ \omega_n < \omega_{n+1} \\ \omega_\ell < \omega_{\ell+1}}} \omega_n \omega_\ell \text{ReLU}(\mathcal{S}_{n\ell}). \quad (2.19)$$

We emphasize that this is an *exact* and closed-form expression for the $p = 2$ SEMD between any two sets of discrete points. Moreover, eq. (2.19) is extremely fast to evaluate; we will see in section 3 that it is possible to evaluate it in $\mathcal{O}(N^2 \log N)$ time using the highly efficient, differentiable, and parallelizable SPECTER framework. This is in contrast to the $p = 1$ SEMD [23], the ordinary EMD, or approximations to Wasserstein such as sliced-Wasserstein [46] or Sinkhorn [47], which either have no closed-form expression, are expensive to evaluate, or are inexact. Lastly, the structure of eq. (2.19) is valid for any choice of ground metric ω . In fact, it is not even necessary that ω be a proper metric at all — it suffices for ω_{ij} to be any symmetric matrix such that $\omega_{ii} = 0$.

For events A and B each with N particles, the first two terms in this sum have $\mathcal{O}(N^2)$ terms to evaluate. The cross term is a sum over both events, and naively appears to have $\mathcal{O}(N^4)$ terms. However, the structure of $\Theta(\mathcal{S}_{n\ell})$ within the ReLU function ensures that almost all of these terms are zero, and that only $\mathcal{O}(N^2)$ actually contribute to the cross term. This term takes the form:

$$\Theta(\mathcal{S}_{n\ell}) = \Theta\left(S_A^+(\omega_n) \geq S_B^+(\omega_\ell) > S_A^-(\omega_n)\right) \text{ OR } \Theta\left(S_B^+(\omega_\ell) \geq S_A^+(\omega_n) > S_B^-(\omega_\ell)\right). \quad (2.20)$$

Here, we have slightly abused the step function notation, with $\Theta(\text{TRUE}) = 1$ and $\Theta(\text{FALSE}) = 0$. These two inequalities can be computed efficiently in $\mathcal{O}(N^2 \log N)$ time and makes the SEMD much easier to calculate, as we explored further in section 3.

2.3 Choice of ground metric

The distance metric ω depends on the geometry being considered. For jets at a hadron collider, we consider a local rectangular patch in the rapidity-azimuth (y, ϕ) plane and use the Euclidean metric:³

$$\omega_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}. \quad (2.21)$$

In this case, we use the particle’s transverse momentum p_T in place of its energy E , but for the sake of notational simplicity, we continue to use E everywhere in this paper.

For full events in a spherical detector, we use the spherical arc length (angle θ) metric:

$$\omega_{ij} = |\theta_{ij}| = \left| \cos^{-1} \left(1 - \frac{p_i \cdot p_j}{E_i E_j} \right) \right|, \quad (2.22)$$

where throughout this paper we assume that particle momentum four-vectors p_i, p_j are massless.

There are many other possible geometries and metrics one could have chosen instead, such as cylindrical geometries, chord lengths rather than arc lengths, or even arbitrary warpings of the above metrics, but we stick to these two for the remainder of this paper for simplicity. These two metrics in particular are relatively simple to work with and make it easy to relate the SEMD to the EMD and other well-known observables. We briefly discuss using the spherical chord length rather than arc length in appendix A.

We note that the SEMD and eq. (2.19) are valid and interpretable even outside the context of collider physics. For example, in images, the metric ω might be a pixel distance, and the energy E might be pixel intensities. In general, the SEMD is a valid metric on the space of *arbitrary* weighted distributions — in collider physics, an event (jet) is a distribution of points x_i with weights E_i ($p_{T,i}$) and pairwise distances ω_{ij} . These distributions need not even be discrete: the definition of the SEMD given in eq. (2.4) is valid even if s_A or s_B are continuous, though it can no longer be evaluated using the exact expression in eq. (2.19) and must be approximated using sampling.

2.4 Comparison to the EMD and degeneracies

The $p = 2$ SEMD bears many similarities to the ordinary $\beta = 2$ EMD, and their differences can be illuminating. The $\beta = 2$ EMD between two events \mathcal{E} and \mathcal{E}' , assuming they have the same total energy E_{tot} , is given by [1]:⁴

$$\text{EMD}(\mathcal{E}, \mathcal{E}') = \min_{\pi_{ij} \geq 0} \sum_{ij} \pi_{ij} \omega_{ij}^2, \text{ such that } \sum_i \pi_{ij} = \frac{E'_j}{E_{\text{tot}}} \text{ and } \sum_j \pi_{ij} = \frac{E_i}{E_{\text{tot}}}, \quad (2.23)$$

where i indexes particles in \mathcal{E} and j indexes particles in \mathcal{E}' . Both the EMD and SEMD exhibit the same scaling: bilinear in particle energies and quadratic in distance scales, and are expected to be similar in the soft/collinear limits of QCD [23]. In spite of this, their behavior can sometimes differ wildly, especially for “near-degenerate” events.

³This differs from the convention of ref. [16], which for $\beta = 2$ has an overall factor of $\frac{1}{2}$ in the metric.

⁴For the rest of this paper, whenever we refer to the ordinary EMD, we always mean the $\beta = 2$ EMD.

To set the stage, we first review a case where the SEMD and EMD are expected to agree exactly. Consider a generic jet \mathcal{E} and a special jet \mathcal{E}' consisting of a single particle, such that the center-of-energy $x_0 = \frac{1}{E_{\text{tot}}} \sum_i E_i x_i$ of both jets is the same, where $x_i = (y_i, \phi_i)$ is the detector coordinate of particle i and we use the Euclidean metric of eq. (2.21).⁵ The EMD is straightforward to calculate, since the optimal transport plan is to simply transfer all of \mathcal{E} to the single point of \mathcal{E}' :

$$\text{EMD}(\mathcal{E}, \mathcal{E}') = \frac{1}{2} \frac{\sum_i E_i (x_i - x_0)^2}{\sum_i E_i} = \frac{1}{2} \frac{\sum_{ij} E_i E_j (x_i - x_j)^2}{E_{\text{tot}}^2}, \quad (2.24)$$

where in the last step, we have used that $x_0 = \frac{1}{E} \sum_j E_j x_j$ is the center-of-energy. Note that while the EMD naively appears to be linear in energy, it is actually quadratic once the center-of-energy is accounted for. One can show that the $\beta = 2$ EMD between any two localized jets is minimized whenever they share their center-of-energy — in this case, the minimization is equivalent to the $\beta = 2$ recoil-free jet angularity [48, 49], the $\beta = 2$ 1-subjettiness [50], or the yet-to-be-defined 1-sPronginess (see section 4.2.1 for more details).

We can also calculate the SEMD of this configuration directly using eq. (2.19):

$$\text{SEMD}(s_{\mathcal{E}}, s_{\mathcal{E}'}) = \sum_{i,j} E_i E_j (x_i - x_j)^2. \quad (2.25)$$

Up to the total energy normalization, which we may set to 1, and an unimportant extra factor of 2 (which is due to the $\frac{1}{\beta}$ convention in ref. [16]), these agree. Moreover, both metrics are IRC safe, and therefore continuous on the space of events. This implies that for jets \mathcal{E}' that are *almost* one particle, e.g. one particle plus a soft or collinear emission, $\text{EMD}(\mathcal{E}, \mathcal{E}')$ and $\text{SEMD}(s_{\mathcal{E}}, s_{\mathcal{E}'})$ should be *almost* the same. Both the EMD and SEMD scale bilinearly in energy and scale with $(x_i - x_j)^2$, so one should expect them to be similar.

On the other hand, there are cases where the ordinary EMD and the spectral SEMD are very different. Consider the case where \mathcal{E} is an equilateral triangle of three particles, with one particle having an energy of $\frac{2}{3}$ units, and the other two each having energies of $\frac{1}{6}$ units. Without loss of generality, choose the most energetic particle to be at the origin, and one of the two others to be at $(1, 0)$. Then, choose \mathcal{E}' to consist of two particles, each of energy $\frac{1}{2}$ units, with one at the origin and the other at $(1, 0)$. Clearly, the energy flows of \mathcal{E} and \mathcal{E}' are different, and so the EMD will be nonzero. However, this specific configuration happens to be a rare degenerate case, with both \mathcal{E} and \mathcal{E}' having the exact same spectral function:

$$s_{\mathcal{E}}(\omega) = s_{\mathcal{E}'}(\omega) = \frac{1}{2} \delta(\omega) + \frac{1}{2} \delta(\omega - 1), \quad (2.26)$$

and therefore the SEMD is zero. That is, these two configurations that are very different in “energy flow space” become the same event in “spectral space”.

Even though fully degenerate configurations are just a measure-0 subset of all events, this quirk has implications for the full space of events. Since both metrics are IRC safe, events that are *almost* degenerate will be separated by a large distance in energy flow space, but will be separated by an *almost* zero distance in spectral space. The topology induced

⁵The center-of-energy is not well-defined for a full event on the celestial sphere, but is well-defined for a jet living on a rapidity-azimuth patch.

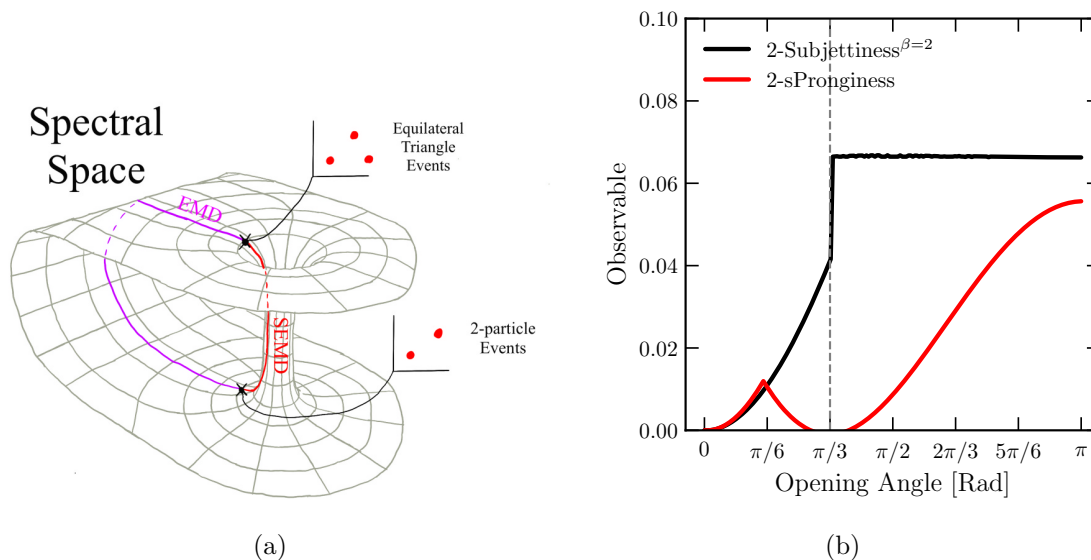


Figure 2. (a) Diagram showing how spectral space induces a different topology on the space of events, as equilateral triangles are closer to 2-particle events in spectral space than in full event space. (b) The minimum EMD (black) and minimum SEMD (red) of an isosceles triangle to all possible 2-particle events (also known as the 2-subjettiness and 2-sPronginess, respectively, defined further in section 4.2.2). The isosceles triangle’s opening angle is varied. The point of the triangle where the angle varies has an energy of $2/3$ and the other two points each have an energy of $1/6$.

by the SEMD is genuinely different than the topology induced by the EMD on the space of events. Observables defined using the EMD are IRC safe in the sense that the EMD changes very little under a soft or collinear splitting. Observables defined using the SEMD satisfy a slightly stronger condition: the SEMD changes very little under a soft or collinear splitting or a hard splitting that happens to be nearly degenerate.

In figure 2, we show the minimum SEMD and EMD between isosceles triangles and the set of all two particle events, also known as the 2-sPronginess (see section 4.2.2) and the 2-subjettiness [50], respectively. The energies of the three particles are as described above, with the apex having an energy of $\frac{2}{3}$, and the opening angle θ is varied. The SEMD and EMD agree locally near the $\theta = 0$ configuration, where the event is essentially two particles, and both exhibit the expected quadratic scaling as the opening angle increases. However, near $\theta = \frac{\pi}{3}$, which is the equilateral triangle configuration, the two sharply diverge, as the event appears to be 3 particles in event space but 2 particles in spectral space. The agreement between the SEMD and EMD has been “spoiled” by the presence of this degeneracy, where otherwise the naive scaling would suggest that the SEMD should continue to grow quadratically alongside the EMD.

Given that the EMD and SEMD exhibit similar energy and distance scaling, their differences can be “blamed” on these topological differences. How often, then, do these differences occur? The simplest possible degenerate configuration is the equilateral triangle configuration described above. Generically, any set of three particles whose distances from each other are all $\mathcal{O}(1)$ of each other and whose relative energies are also all $\mathcal{O}(1)$ will be “almost” degenerate, and these terms in the spectral function will almost collapse to two points.

In a light quark or gluon-initiated QCD jet, this is unlikely to occur: emissions of the hard core are strongly ordered in both the emission angle and energy — that is, QCD jets tend to have 1 hard prong. Even if a degenerate triangle does occur between three particles of similar energy, their contribution to the spectral function will be highly energy-suppressed relative to the harder component. In contrast, a top-quark-initiated jet tends to form three hard subprongs in an approximately triangular formation. Therefore, one should expect that the EMD and SEMD should be similar between pairs of QCD jets, but be significantly different between pairs of top jets. This is indeed the case; the empirical study in section 5.2 shows that the EMD and SEMD is indeed similar for QCD jets, but differs on other phase spaces.

3 The SPECTER framework

SPECTER is a numeric framework for computing the $p = 2$ SEMD between events using eq. (2.19), and for computing observables derived from the SEMD via gradient descent as described further in section 4. SPECTER is implemented in JAX [51], a Python library for differentiable and compilable computing. In section 3.1, we discuss the algorithm to efficiently compute eq. (2.19). Then, in section 3.2, we demonstrate the speed of the SPECTER framework with empirical studies. This section is self-contained, and readers less focused on the implementation of the SEMD can safely skip to section 4.

3.1 Algorithm for the Spectral EMD

Given an event \mathcal{E} with N massless particles, we first compute its spectral representation $s_{\mathcal{E}}$. We assume that all events are normalized with a total energy $E_{\text{tot}} = 1$, and that we can efficiently compute a distance matrix ω_{ij} , computed using either eq. (2.21) or eq. (2.22).⁶ The spectral representation is a $2 \times (1 + \binom{N}{2})$ array of values, with the first column representing the list of distances ω_{ij} and the second column representing the list of energies $(2EE)_{ij}$. The second dimension represents the list of all pairs of particles, plus an additional entry for the $i = j$ pairs. We index these pairs by n , and we 0-index, with the index $n = 0$ referring to the $i = j$ pairs.

For $i > j$, the matrix of distances ω_{ij} are then sorted in ascending order into a single list ω_n and assigned indices $n = 1, \dots, \binom{N}{2}$. The special value $\omega_0 = 0$ is assigned for $n = 0$. These values then populate the first column of the $s_{\mathcal{E}}$ array. Then, for $n \geq 1$ corresponding to the pair (i, j) , the value $(2EE)_{ij} = 2E_i E_j$ is computed and populate the second column, with the special value $(2EE)_0 = \sum_i E_i^2$.

Importantly, every step of this operation is fully differentiable (even the sorting, as long as the indices are tracked). The entire operation can be run in $\mathcal{O}(N^2 \log N)$ time: the N^2 is due to the pairs, and the logarithm to the sorting and searching operations. Note that storing the spectral representation of an event requires $\mathcal{O}(N^2)$ in memory, unlike an ordinary event which requires only $\mathcal{O}(N)$ in memory. For large N , approaching as high as $N \sim 1000$ for heavy ion collision experiments, these memory requirements per event can become a bottleneck of the SPECTER algorithm.

⁶In principle, nothing prevents the use of arbitrary ω_{ij} in the algorithm, but for simplicity we will stick with the two prescribed metrics.

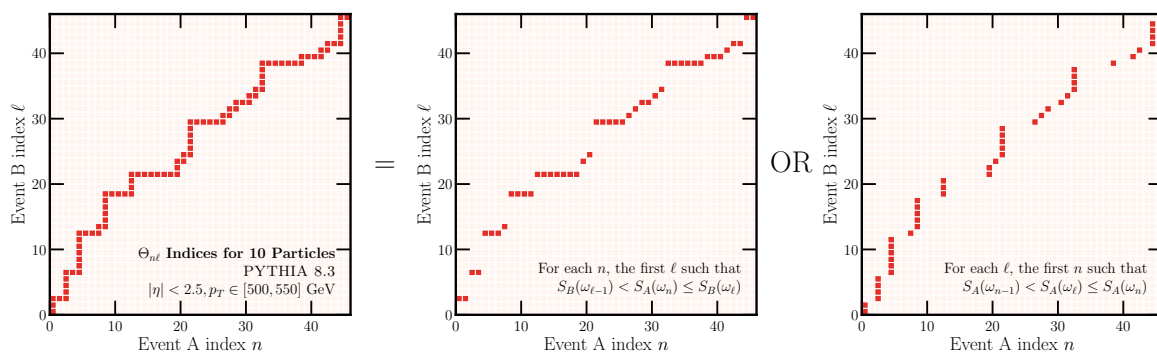


Figure 3. (Left) The indices (n, ℓ) for which $\Theta(\mathcal{S}_{n\ell})$, as defined in eq. (2.16), is nonzero for a typical pair of events with 10 particles. The bright red points are (n, ℓ) pairs where $\Theta(\mathcal{S}_{n\ell}) = 1$. (Middle, Right) This can be computed as the logical OR of eq. (3.2), which consist solely of searches within sorted lists and are primitive operations within JAX.

We will find it convenient to process many events at once (a *batch*), each with a different number of particles N . To deal with this and keep all of our arrays rectangular, we choose a fixed $N_{\text{particles}}$.⁷ Events with fewer particles than this are 0-padded (which will not change the energy flow \mathcal{E} by IRC safety), and events with more particles than this have their least energetic particles removed (which is *not* IRC safe).⁸

Given two events \mathcal{E}_A and \mathcal{E}_B , and their spectral representations constructed as described above, we now wish to compute eq. (2.19) efficiently. The first two terms of eq. (2.19) are straightforward given the spectral representations $s_{\mathcal{E}}$, as they take the form:

$$\begin{aligned} \text{Term 1} + \text{Term 2} &= \int d\omega \omega^2 (s_A(\omega) + s_B(\omega)) \\ &= \sum_{n=0} \omega_n^2 (2EE)_{A,n} + \sum_{\ell=0} \omega_\ell^2 (2EE)_{B,n}, \end{aligned} \quad (3.1)$$

which are simple sums with $\mathcal{O}(N^2)$ terms, and they are also straightforward to differentiate through. The cross term of eq. (2.19) is more daunting, as the sum naively has $\mathcal{O}(N^4)$ terms to evaluate. However, this sum is sparse — almost all terms are 0, and only $\mathcal{O}(N^2)$ terms will survive the $\Theta(\mathcal{S}_{n\ell})$ function defined in eq. (2.20). An example of this is shown in figure 3, which shows $\Theta(\mathcal{S}_{n\ell})$ for a typical pair of events. Only a sparse subset of (n, ℓ) terms will contribute to the sum. Moreover, we can efficiently pre-compute which indices (n, ℓ) will survive the $\Theta(\mathcal{S}_{n\ell})$ function by taking advantage of the fact that spectral functions are one-dimensional objects that can be sorted.⁹

For a given (n, ℓ) pair, eq. (2.20) implies $\Theta(\mathcal{S}_{n\ell})$ is 1 if and only if:

$$S_B(\omega_{\ell-1}) < S_A(\omega_n) \leq S_B(\omega_\ell) \quad \text{OR} \quad S_A(\omega_{n-1}) < S_B(\omega_\ell) \leq S_A(\omega_n), \quad (3.2)$$

⁷In principle, this is not necessary and one can instead use “awkward arrays” [52] to avoid the need for rectangular data structures, but for simplicity this is not implemented in v1.0.0 of SPECTER.

⁸To avoid this, events should be clustered in an IRC-safe way to fewer than $N_{\text{particles}}$ ahead of time.

⁹This is a concrete realization of the conventional wisdom that having sortable distances is why optimal transport is easier in one dimension than in higher dimensions.

where $S(\omega)$ is the cumulative spectral function corresponding to $s(\omega)$. These relationships are simple searches through sorted lists of the form: for each n , the inequality at the left of eq. (3.2) returns an ℓ satisfying the inequality, and similarly, for each ℓ , the inequality at the right of eq. (3.2) returns an n satisfying the inequality. This operation is implemented as an efficient and parallelized primitive in many Python libraries — in JAX, it is the `searchsorted` function, which returns the *first* ℓ or n satisfying the inequalities. We need to combine the inequalities with a logical OR to find all such ℓ or n , which is demonstrated in figure 3. For each n , the search operation runs in only $\mathcal{O}(\log N)$ since $S(\omega_\ell)$ is sorted in ℓ by construction, and so for the entire event it only takes $\mathcal{O}(N^2 \log N)$ to find all indices that contribute to the cross term. The logical OR structure implies that there can be *at most* $2 \binom{N}{2} < N^2$ pairs of indices, which allows us to set an upper bound to the allocated memory, though in practice this limit is rarely saturated. Note that the $n = 0$ and $\ell = 0$ terms can be skipped, since $\omega_0 = 0$ and therefore does not contribute to the sum.

With the list of indices (n, ℓ) for which $\Theta(\mathcal{S}_{n\ell})$ is nonzero in hand, the cross term in eq. (2.19) is now faster to evaluate, as we only need to sum $\mathcal{O}(N^2)$ terms. The indices do not need to be recomputed when calculating derivatives, so time is in fact saved if the same Θ function is used for both the SEMD and its derivatives with respect to $s_{\mathcal{E}}$. Thus, the entire SEMD computation can be performed exactly in $\mathcal{O}(N^2 \log N)$ time and $\mathcal{O}(N^2)$ memory, in a way that is easily amenable to parallelization and differentiation.

It is important to emphasize that this is an *exact* algorithm to compute the SEMD. No approximations have been made, numerical or otherwise. This is important to emphasize since when evaluating spectral shape observables (see section 4 for more details), there are two approximations made: first, continuous events are approximated as discrete, so that this algorithm may be applied; second, evaluating shape observables requires gradient descent over the shape parameters, which is an inexact minimization procedure. However, the SEMD evaluation itself is exact — this is in contrast to SHAPER, where on top of these two approximations, the EMD computation is also an approximation using the debiased Sinkhorn divergence [47, 53–56].

3.2 Runtime benchmarks

In figures 4 and 5, we show the runtime of SPECTER to calculate the SEMD between two events with N particles. For comparison, we also show the runtime for calculating the ordinary EMD using the approximate Sinkhorn algorithm via SHAPER, and an exact calculation using the Python Optimal Transport library (POT) [22]. Figure 4 shows the runtimes when running on a compute cluster GPU (specifically, a NVIDIA A100 with 16 GB of memory) and figure 5 shows the runtimes when running on an ordinary laptop CPU (specifically, one of the authors’ Dell Inspiron 16 with an Intel i7-1255Ux12 CPU with 8 GB of memory), which together span the typical use cases. We include curves for several different batch sizes to show the impact of parallelization: 1, 10, 100, 1000, and 10000 pairs of events at once in figure 4, and 1, 10, and 100 pairs of events at once in figure 5. Note that some of the curves for SHAPER and SPECTER truncate early: this is due to memory limitations. In particular, because JAX arrays are indexed using `int32`, and SPECTER requires $\mathcal{O}(N^2)$ memory, events with more than $2^{16} \approx 30000$ particles cannot be handled.

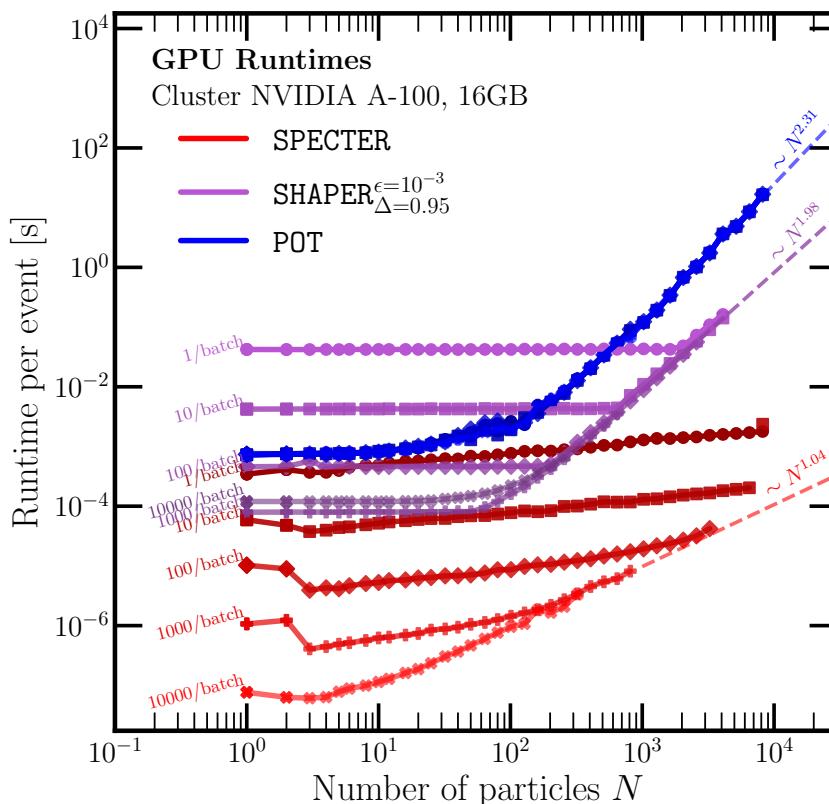


Figure 4. Runtimes for various (S)EMD algorithms as a function of the number of particles N in each event, as run on a NVIDIA A100 GPU with 16 GB of memory. The SEMD, as evaluated by SPECTER, is shown in red. The ordinary EMD, as estimated by the Sinkhorn algorithm with SHAPER and in full with the Python Optimal Transport library, are shown in purple and blue, respectively. All lines eventually scale as $N^{\sim 2}$ or better.

On the GPU, SPECTER is *significantly* faster than the ordinary EMD calculation, achieving speeds that are hundreds to thousands of times faster than SHAPER or POT. For events with ~ 100 particles, which is typical in jet physics, it takes only 0.001 seconds to compute the SEMD between 10000 pairs of events. The corresponding calculation of the EMD on the same hardware using SHAPER would take ~ 1 second (and it would only be an estimate), and nearly 100 seconds using POT. Even on a CPU, SPECTER is faster than either SHAPER or POT at low N , and is approximately just as fast at moderate and large N .

All algorithms tend towards $\sim \mathcal{O}(N^2)$ at large N . However, especially on the GPU (figure 4), the runtime at $N \lesssim 100$ is dominated by an overhead component, and in this regime the scaling is significantly better than $\mathcal{O}(N^2)$. In fact, an $\mathcal{O}(N)$ or smaller overhead seems to completely dominate the SPECTER runtimes across the entire range before being cutoff due to memory limitations, which becomes more apparent with larger batch sizes. It is possible that this is not saturated in our trials, and that even faster per-event runtimes might be achieved with larger batch sizes. In all cases, the compile time is not considered: this is the long-run per-event runtime. Both SHAPER and SPECTER take less than one minute to compile on both the laptop and GPU hardware.

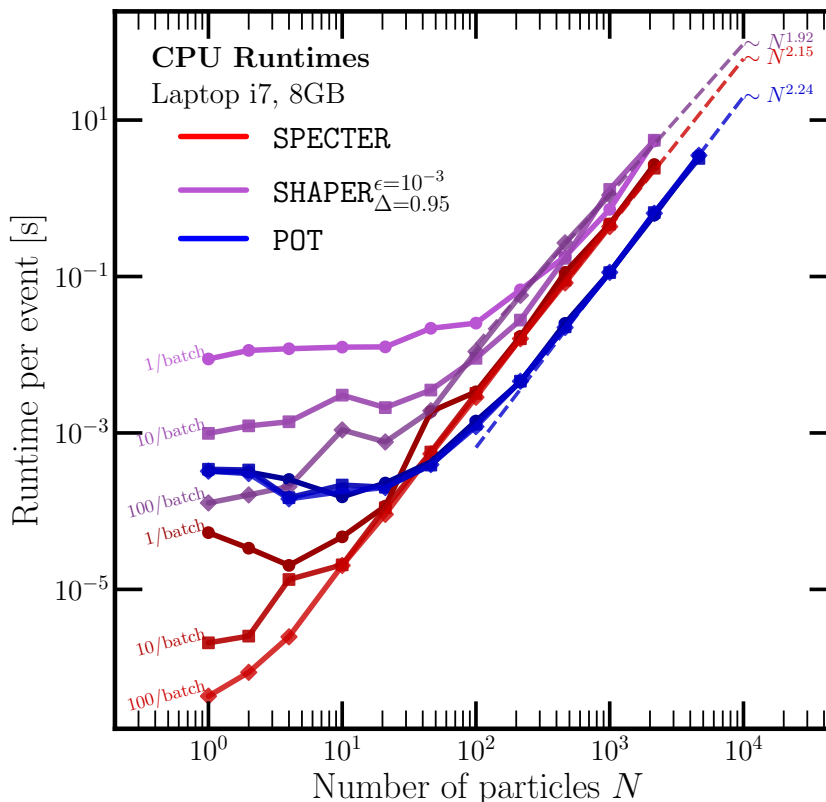


Figure 5. The same as figure 4, but as run on an ordinary laptop CPU with 8 GB of memory. Each of the three lines corresponds to batch sizes of 1, 10, and 100 respectively. All lines eventually scale as N^2 .

4 Spectral shapes

Having defined the $p = 2$ SEMD and presented the SPECTER algorithm for efficient evaluation, we now show how the SEMD can be used to define novel *spectral* event and jet shape observables. These are analogous to observables defined using the ordinary EMD [1, 16], but with the added benefit of being much easier to evaluate on events due to the speed of SPECTER. In some cases, it is even possible to define observables with tractable closed-form expressions, in contrast to the EMD where it is difficult to find closed-form expressions in all but the most trivial cases. This makes it possible to define shape observables that both probe event geometry *and* can be evaluated extremely quickly, and in some cases practically instantaneously with an exact closed-form expression.

4.1 General aspects

Following the philosophy of ref. [16], *any* metric on the space of events can be used to define an infinite family of event and jet shape observables. The SEMD is no different in this regard. A “shape” can be defined as a parameterized energy flow $\mathcal{E}_{\text{Shape}}^{(\theta)}$ on detector space, which then induces a corresponding spectral function $s_{\text{Shape}}^{(\theta)}$. Then, for an event or jet with energy flow \mathcal{E} and corresponding spectral function $s_{\mathcal{E}}$, we can define a *spectral shape observable* $\mathcal{O}_{\text{Shape}}(s_{\mathcal{E}})$

and the corresponding *spectral shape parameters* $\theta_{\text{Shape}}(s_{\mathcal{E}})$:¹⁰

$$\mathcal{O}_{\text{Shape}}(s_{\mathcal{E}}) = \min_{\mathcal{E}_{\text{Shape}}^{(\theta)}} \text{SEMD}(s_{\mathcal{E}}, s_{\text{Shape}}^{(\theta)}) \quad (4.1)$$

$$\theta_{\text{Shape}}(s_{\mathcal{E}}) = \underset{\mathcal{E}_{\text{Shape}}^{(\theta)}}{\text{argmin}} \text{SEMD}(s_{\mathcal{E}}, s_{\text{Shape}}^{(\theta)}). \quad (4.2)$$

Given the algorithm in section 3.1 for efficiently computing the SEMD in a differentiable way, one can numerically approximate the (arg)minimum via gradient descent. This is done essentially identically to SHAPER [16].

Note that even though we use the SEMD, as defined on spectral densities, we still perform the optimization over detector space events $\mathcal{E}_{\text{Shape}}^{(\theta)}$ rather than directly over parameterized spectral functions $s_{\text{Shape}}^{(\theta)}$. This is to ensure that the final parameters correspond to a physically realizable event — not all spectral densities correspond to physical values of energy or particle positions. For instance, the spectral density:

$$s(\omega) = E_{\text{tot}}^2 \delta(\omega - \omega_0), \quad (4.3)$$

for nonzero ω_0 can only come from a non-physical event with two particles, one with energy $E_1 = \frac{E_{\text{tot}}}{2}(1 + i)$ and the other with energy $E_2 = \frac{E_{\text{tot}}}{2}(1 - i)$, and therefore one cannot directly optimize over ω_0 if one wants to avoid these “ghost events”. However, this is still technically a valid optimization to perform, and the observables \mathcal{O} and θ obtained this way are perfectly well-defined if one is willing to accept the minimization over non-physical ghost events, though we do not do so here.

Given this setup, all one must do to define a shape is to specify a manifold \mathcal{M} of parameterized energy flows $\mathcal{E}_{\text{Shape}}^{(\theta)}$ with coordinates θ . This is exactly identical to how observables are defined using SHAPER as described in section 3 of ref. [16], and indeed all the same manifolds of energy flows can be carried over. Shapes are defined by first specifying the parameterization manifold $\mathcal{M}_{\text{Shape}} \ni \theta$, and then providing a sampling procedure to sample N weighted points based on those parameters. For example, one can define the “spectral Ringiness” (or “spRinginess”) of an event, which is parameterized by a positive radius ($R \in \mathcal{M}_{\text{Ring}} = \mathbb{R}_+$), plus a function to sample N points uniformly on a circle of radius R each with weight $1/N$. As with SHAPER, multiple primitive shapes can be composed together to form composite shapes that probe more complex geometries.

The isometry-invariance of the SEMD leads to a small subtlety when parameterizing shapes. Because the SEMD is invariant under translation and rotation of either event, parameters that affect the overall translational or rotational degrees of freedom will carry no information. For example, while the ordinary Ringiness observable needs an overall translational degree of freedom to specify the location of the circle, the spRinginess does not. The same observable definition and sampling can be used for both, but the spectral observable will be completely agnostic to the unneeded parameter. The rest of this section is dedicated to discussing examples of spectral shape observables.

¹⁰Note on naming conventions: we use the argument of the function to distinguish the EMD-based observables and parameters from the SEMD-based ones with $\mathcal{O}_{\text{Shape}}(\mathcal{E})$ and $\theta_{\text{Shape}}(\mathcal{E})$ for the ordinary EMD, and $\mathcal{O}_{\text{Shape}}(s_{\mathcal{E}})$ and $\theta_{\text{Shape}}(s_{\mathcal{E}})$ for the SEMD. When naming shapes, we use “s” or “sp” in front to signal the spectral version.

4.2 Spectral prong shapes

The most basic possible spectral shape observables we can define are the N -sPronginess observables, which are the spectral equivalent of the $\beta = 2$ N -(sub)jettiness [50, 57, 58] observables.¹¹ These observables probe how much an event looks like N prongs. These observables can be defined either on local patches of the rapidity-azimuth plane for jets using the metric in eq. (2.21) (corresponding to N -subjettiness), or over the entire detector for full events using the metric in eq. (2.22) (corresponding to N -jettiness with no beam parameter).¹²

4.2.1 1-sPronginess

The simplest spectral shape observable is the 1 -sPronginess, which probes how much the event or jet looks like it has 1 prong. This is accomplished by calculating the (spectral) distance to an event with just a single particle, given by the energy flow density:

$$\mathcal{E}_{1\text{-Prong}}(x) = E_{\text{tot}}\delta(x - x_0), \quad (4.4)$$

where E_{tot} is the total energy of the particle and x_0 is its position in the detector. However, this position does not matter due to the translational invariance of the metric, for which the relevant object is the spectral representation:

$$s_{1\text{-Prong}}(\omega) = E_{\text{tot}}^2\delta(\omega). \quad (4.5)$$

Note that there are no free parameters in the spectral function — in spectral space, there is only *one* single particle event up to energy normalization. Given this, we can exactly compute the 1-sPronginess observable for an event $\mathcal{O}_{1\text{-Prong}}(s_{\mathcal{E}})$, as the SEMD between $s_{\mathcal{E}}$ and $s_{1\text{-Prong}}$ by using eq. (2.19):

$$\mathcal{O}_{1\text{-Prong}}(s_{\mathcal{E}}) = \sum_{i < j \in \mathcal{E}} 2E_i E_j \omega_{ij}^2. \quad (4.6)$$

In this form, the 1-sPronginess is actually identical to many other well-known observables. For example, the 1-sPronginess is identical to the recoil-free $\beta = 2$ angularity [48, 49], which is itself identical to the $\beta = 2$ 1-subjettiness [50] for $\omega = \Delta R$ as in eq. (2.21).

4.2.2 N -sPronginess

The first *nontrivial* spectral observable is the N -sPronginess, for $N > 1$, which measures how much an event \mathcal{E} looks like it has N prongs. These are characterized by calculating the spectral distance to events with N particles:

$$\mathcal{E}_{N\text{-Prong}}^{(E_i, x_i)}(x) = \sum_i^N E_i \delta(x - x_i), \quad (4.7)$$

¹¹We call it the N -sPronginess because “ N -ssubjettiness” [sic] would have been a terrible name.

¹²It is possible to define a beam parameter by using unbalanced spectral functions, by introducing a scale parameter ω_R to re-balance the spectral functions as described in eq. (2.13) of ref. [23].

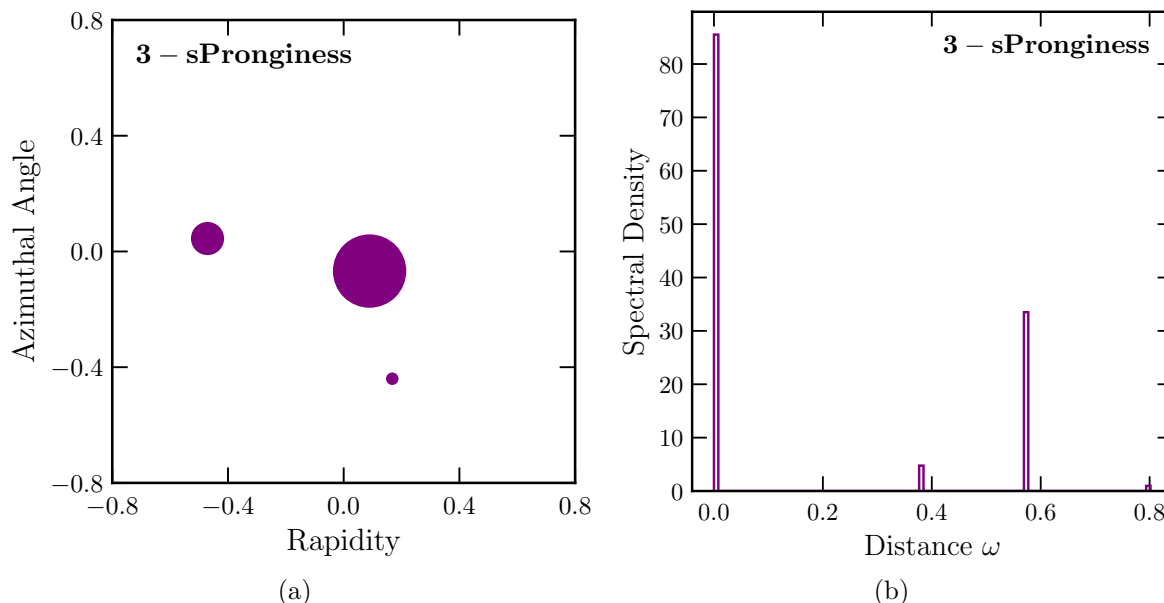


Figure 6. The (a) detector representation from eq. (4.7) and (b) spectral representation from eq. (4.8) of an idealized 3-(s)Prong event, in a patch of the rapidity-azimuth plane. Each dot is a particle, with its size representing its energy weight.

where E_i are the particle energies, constrained to be positive and sum to E_{tot} , and x_i are the particle positions. The corresponding spectral density is:

$$s_{N\text{-Prong}}^{(E_i, x_i)}(\omega) = \sum_{i \in \mathcal{E}_{N\text{-Prong}}} E_i^2 \delta(\omega) + \sum_{i < j \in \mathcal{E}_{N\text{-Prong}}} 2E_i E_j \delta(\omega - \omega_{ij}), \quad (4.8)$$

where ω_{ij} is the distance between x_i and x_j , as defined using either eq. (2.21) or eq. (2.22). Note that due to translational invariance, one of the x_i degrees of freedom is redundant. Additionally, there is an overall redundant rotational degree of freedom between the x_i 's. An example of a 3-(s)Prong event and its spectral representation on the rapidity-azimuth plane is shown in figure 6.

Given this, we can define the N -sProngness observable $\mathcal{O}_{N\text{-Prong}}(s_{\mathcal{E}})$ and corresponding spectral parameters $\theta_{N\text{-Prong}}(s_{\mathcal{E}}) = (E_i, x_i)$ via eqs. (4.1) and (4.2),¹³ which can be numerically evaluated using the SPECTER algorithm. For a given choice of (E_i, x_i) parameters, the SEMD can be exactly evaluated using eq. (2.19), though the minimization over the (E_i, x_i) is nontrivial. To choose initial values of E_i and x_i in an IRC-safe way, we simply apply k_T -clustering [59, 60] to obtain N subjects. The positivity and sum rule constraints on E_i can be enforced during gradient descent using the simplex projection algorithm [61–63].

As discussed in section 2.4, the N -sProngness for $N > 1$ is nontrivial in that it is distinctly different than its corresponding ordinary EMD observable, the N -subjettiness, which is defined using the same parameterized energy flows. The N -sProngness probes slightly different physics than the N -subjettiness: while a low value of 2-subjettiness implies a jet looks like it has two hard subprongs, a low value of 2-sProngness implies that a jet looks like it has two hard subprongs *or* has three hard subprongs in an equilateral triangle formation.

¹³Up to the redundant translational and rotational degrees of freedom.

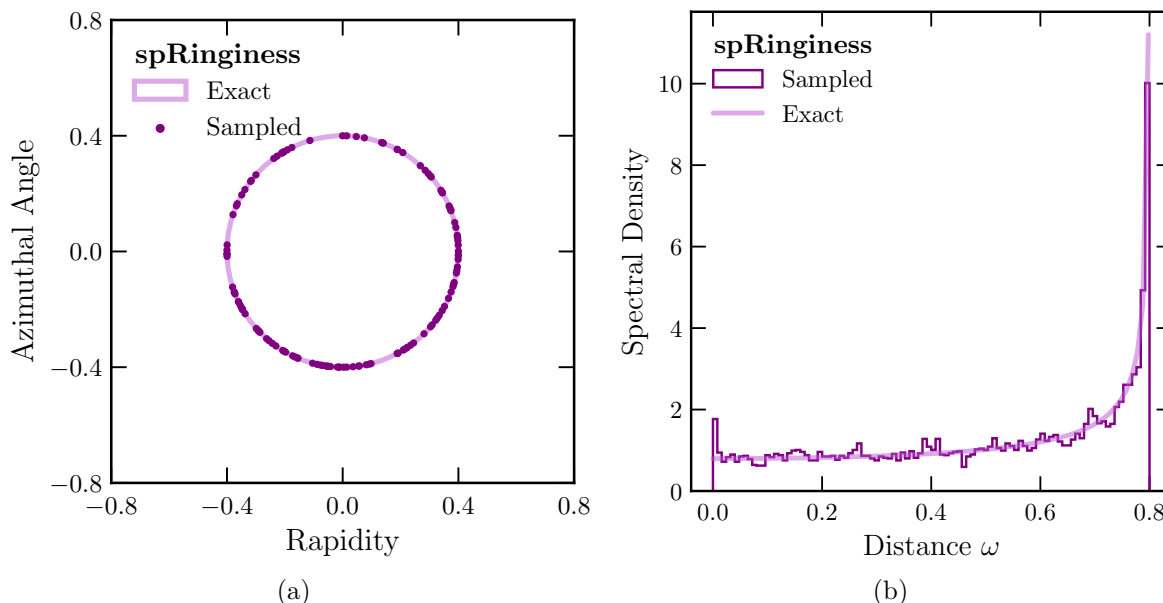


Figure 7. The (a) detector representation from eq. (4.9) and (b) spectral representation from eq. (4.10) of an idealized ring jet, in a patch of the rapidity-azimuth plane. Each dot is a sampled particle, with its size representing its energy weight, which happens to be uniform.

4.3 Spectral jet shapes

Building off of the N -sPronginess, we can build more nontrivial observables that probe the structure of localized jets. In this subsection, we define several novel spectral jet shape observables and discuss their properties. For all of the observables in this subsection, we take the ground metric ω to be the Euclidean metric in eq. (2.21).

4.3.1 Jet spRinginess

We define the jet *spRinginess* observable, in analogy with the Ringiness observables of ref. [16]. A ring-shaped event with radius R is given by the energy flow density:

$$\mathcal{E}_{\text{Jet Ring}}^{(R)}(x) = \frac{E_{\text{tot}}}{2\pi R} \delta(|x| - R). \quad (4.9)$$

The radius R is an explicit parameter that will later have to be optimized. The corresponding spectral representation is:

$$s_{\text{Jet Ring}}^{(R)}(\omega) = \frac{E_{\text{tot}}^2}{\pi} \frac{2}{\sqrt{4R^2 - \omega^2}}. \quad (4.10)$$

Unlike the N -sPronginess of the previous subsection, this is a continuous rather than discrete distribution. However, one can approximate the continuous distribution by simply sampling a large number N of points in a circle of radius R , each with energy $\frac{E_{\text{tot}}}{N}$. Then, all one has to do is use the SPECTER algorithm to compute the SEMD between the event of interest \mathcal{E} and the discrete event approximating the sphere, and use gradient descent to find the optimal R . There are a variety of ways to choose the initial value of R , such as the distance between the leading two subjects, but for simplicity we choose to initialize $R = 0$

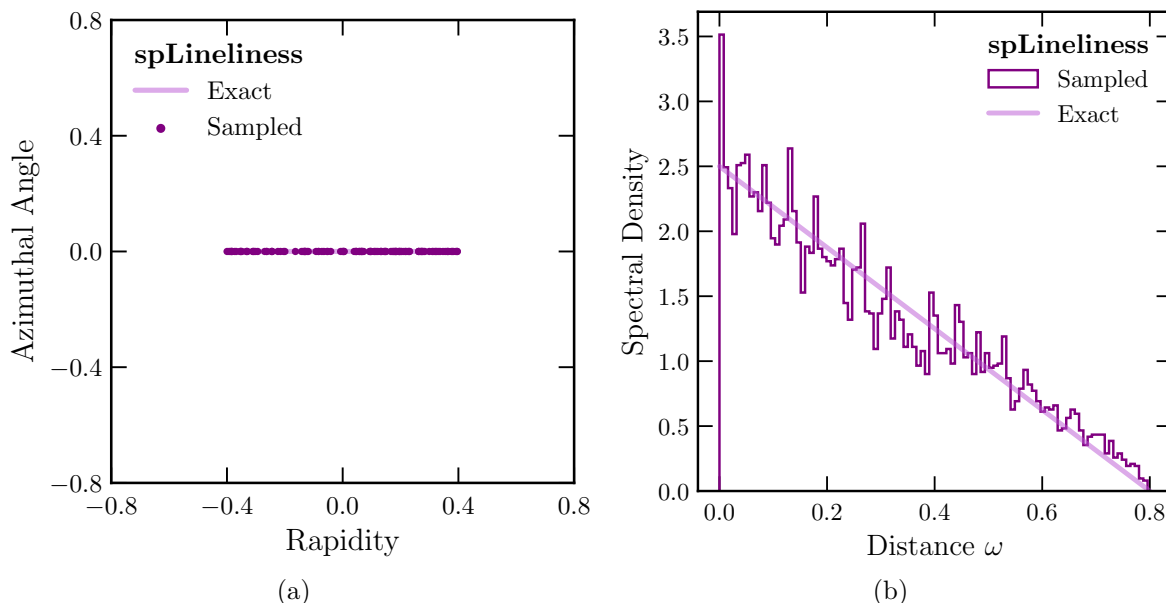


Figure 8. The (a) detector representation from eq. (4.13) and (b) spectral representation from eq. (4.14) of an idealized line jet, in a patch of the rapidity-azimuth plane. Each dot is a sampled particle, with its size representing its energy weight, which happens to be uniform.

so that the observable can be treated as a perturbation to the 1-sPronginess, similar to the choices of initialization for SHAPER. Unlike the ordinary jet Ringiness, however, there is little danger of getting stuck in a local minimum, since we will see extremely shortly in eq. (4.12) that the SEMD is convex in R . An example ring is shown in figure 7. In figure 7(b), there is a peak in the sampled ring distribution at $\omega = 0$. The $\omega = 0$ bin receives contributions from the individual particle energies squared — in the fully continuous distribution, these are infinitesimal quantities, but in a uniform sample with N particles, this peak will go as $N \times \frac{1}{N^2}$. This peak is a generic feature of sampling continuous shapes.

The jet spRinginess is our first example of a nontrivial observable with a fully closed-form expression: by using the cumulative spectral function and eq. (2.7), we can actually completely bypass the SPECTER algorithm and sampling just described and exactly evaluate the jet spRinginess and R_{opt} (see appendix B.1 for details). The optimal radius $R_{\text{opt}}(s_{\mathcal{E}})$ of an event \mathcal{E} with spectral representation $s_{\mathcal{E}}$ is given by:

$$R_{\text{opt}} = \frac{2}{\pi} \sum_{\substack{n \in \mathcal{E}^2 \\ \omega_n < \omega_{n+1}}} \omega_n \left[\cos \left(\frac{\pi}{2E_{\text{tot}}^2} S^-(\omega_n) \right) - \cos \left(\frac{\pi}{2E_{\text{tot}}^2} S^+(\omega_n) \right) \right]. \quad (4.11)$$

The spRinginess, $\mathcal{O}_{\text{Jet Ring}}(s_{\mathcal{E}})$, of the event is then given by:

$$\mathcal{O}_{\text{Jet Ring}}(s_{\mathcal{E}}) = \min_R \text{SEMD}_{\beta,p=2} \left(s_{\mathcal{E}}, s_{\text{Jet Ring}}^{(R)} \right) = \sum_{i < j \in \mathcal{E}} 2E_i E_j \omega_{ij}^2 - 2E_{\text{tot}}^2 R_{\text{opt}}^2. \quad (4.12)$$

4.3.2 Jet spLineliness

Similar to jet spRinginess, which probes how ring-like a jet is, we can also ask how line-like a jet is to define the jet *spLineliness*. Here, we compare to line-segment-like events with a length L :

$$\mathcal{E}_{\text{Jet Line}}^{(L)}(x) = \frac{E_{\text{tot}}}{L} \delta(x^{(2)} - 0) \Theta(0 < x^{(1)} < L), \quad (4.13)$$

where $x^{(1)}$ and $x^{(2)}$ are the first and second detector coordinates, respectively (e.g., in rapidity-azimuth space, $x^{(1)}$ is the rapidity and $x^{(2)}$ is the azimuthal angle). Up to isometries, this describes all possible line-segment-like jets. The corresponding spectral representation is:

$$s_{\text{Jet Line}}^{(L)}(\omega) = \left(\frac{2}{L} - \frac{2\omega}{L^2} \right) E_{\text{tot}}^2. \quad (4.14)$$

An example line is shown in figure 8.

Just like the jet spRinginess, both the jet spLineliness and the optimal parameter L_{opt} can be numerically estimated with the SPECTER algorithm, by sampling points on a line segment of length L and performing gradient descent over L to minimize the SEMD.¹⁴ Similarly, it is possible to evaluate the spLineliness and optimal length in closed form to bypass the numeric estimate (see appendix B.2 for details):

$$L_{\text{opt}}(s\mathcal{E}) = \frac{6}{E_{\text{tot}}^2} \sum_{\substack{n \in \mathcal{E}^2 \\ \omega_n < \omega_{n+1}}} \omega_n \left[S^+(\omega_n) + \frac{2}{3E_{\text{tot}}} \left(E_{\text{tot}}^2 - S^+(\omega_n) \right)^{3/2} \right. \\ \left. - S^-(\omega_n) - \frac{2}{3E_{\text{tot}}} \left(E_{\text{tot}}^2 - S^-(\omega_n) \right)^{3/2} \right]. \quad (4.15)$$

$$\mathcal{O}_{\text{Jet Line}}(s\mathcal{E}) = \sum_{i < j \in \mathcal{E}} 2E_i E_j \omega_{ij}^2 - \frac{L_{\text{opt}}^2}{6} E_{\text{tot}}^2. \quad (4.16)$$

4.3.3 Jet sDiskiness

The last jet shape observable we present is the jet *sDiskiness*, which asks how much like a uniform disk a given jet looks like. The corresponding disk-like events are given by:

$$\mathcal{E}_{\text{Jet Disk}}^{(R)}(x) = \frac{E_{\text{tot}}}{\pi R^2} \Theta(|x| < R), \quad (4.17)$$

where R is a free parameter determining the radius of the disk. The jet sDiskiness and R_{opt} can be estimated using the SPECTER algorithm by sampling points uniformly on a disk with radius R , and then performing gradient descent to minimize the SEMD with respect to R . An example disk is shown in figure 9.

However, unlike the above cases, a closed form expression for $\mathcal{O}_{\text{Jet Disk}}(s\mathcal{E})$ or $R_{\text{opt}}(s\mathcal{E})$ seems intractable. The spectral density of a disk is given by:¹⁵

$$s_{\text{Jet Disk}}^{(R)}(\omega) = \frac{4\omega E_{\text{tot}}^2}{\pi R^2} \cos^{-1} \frac{\omega}{2R} - \frac{2\omega^2 E_{\text{tot}}^2}{\pi R^3} \sqrt{1 - \frac{\omega^2}{4R^2}}. \quad (4.18)$$

¹⁴SPECTER is built off of and heavily borrows from SHAPER, especially in its sampling procedure for shapes. The original version of SHAPER did not have line segments built in, but it is possible to cheat and sample line segments by using SHAPER's built-in ellipses with $R_1 = L/2$ and R_2 frozen to zero. The fact that the exact and numerical curves in figure 19 match shows that this is fine to do, and shows how shapes can be manipulated in both programs.

¹⁵This spectral density is the solution to the well-known ‘‘ball line picking’’ problem [39–41].

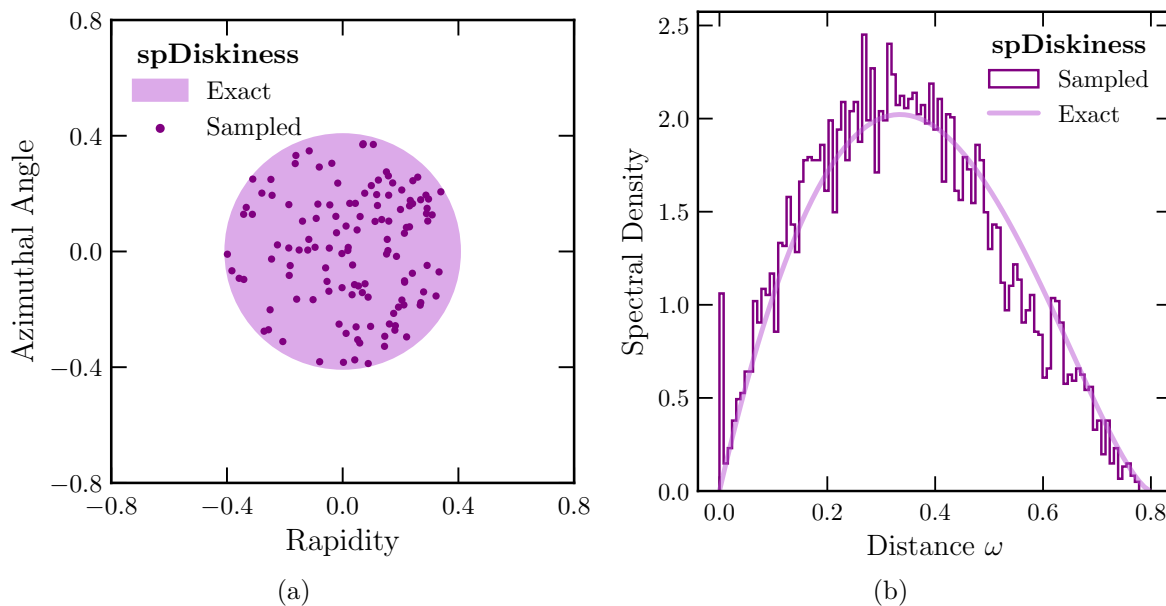


Figure 9. The (a) detector representation from eq. (4.17) and (b) spectral representation from eq. (4.18) of an idealized disk jet, in a patch of the rapidity-azimuth plane. Each dot is a sampled particle, with its size representing its energy weight, which happens to be uniform.

To perform the closed-form SEMD, one must obtain the inverse of the cumulative spectral function, and then integrate it against eq. (2.15). This is a difficult integral — one can attempt to numerically estimate it via Monte Carlo, but this is effectively what the SPECTER algorithm already does by sampling.

4.4 Spectral event shapes

We can also define spectral shape observables over the entire detector space rather than just a localized region. The observables listed in this subsection probe the full event geometry. For these observables, we assume that the ground metric ω is the arc-length metric on the sphere of eq. (2.22).¹⁶ All the previously described observables in sections 4.2 and 4.3 can also be defined on the sphere by appropriately substituting in the correct ground metric ω .

As discussed in section 2.4, the EMD and SEMD of localized events are expected to agree exactly when one of the events has only a single particle. However, when considering a full detector in the center-of-mass frame, events with a single massless particle are unphysical, and events cannot be localized to a single patch. Therefore, for event shapes, it is often convenient to adopt observable-dependent normalizations such that the EMD and SEMD agree on some other physical configuration.

4.4.1 Event sThrust

In analogy with the ordinary event thrust observable [64–66], we can define the event *sThrust*, which probes how back-to-back an event is using the SEMD. The *sThrust* is given by the

¹⁶One can instead choose to use other metrics, such as chord length, to define these observables. In this case, all of these observables are still well defined and qualitatively similar to the arc-length observables, but the closed-form expressions may be different (if they are derivable at all).

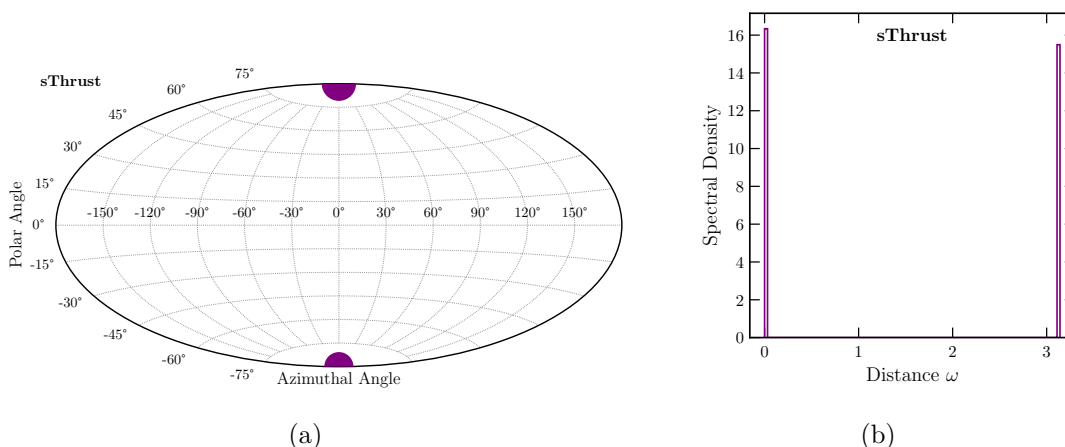


Figure 10. The (a) detector representation from eq. (4.19) and (b) spectral representation from eq. (4.20) of an idealized thrust event with $z \approx 0.45$, on the celestial sphere. Each dot is a particle, with its size representing its energy weight.

minimum spectral distance to a back-to-back 2 particle event with potentially different energies:

$$\mathcal{E}_{\text{Thrust}}^{(z)}(x) = E_{\text{tot}}(z\delta(\theta) + (1-z)\delta(\theta - \pi)), \quad (4.19)$$

where z and $1-z$ parameterize the energy fractions carried by each particle and θ is the usual polar coordinate of the sphere. The corresponding spectral function is

$$s_{\text{Thrust}}^{(z)}(\omega) = (z^2 + (1-z)^2) E_{\text{tot}}^2 \delta(\omega) + 2z(1-z) E_{\text{tot}}^2 \delta(\omega - \pi). \quad (4.20)$$

Note that unlike the case with ordinary Thrust, which when calculated using the EMD requires optimizing over both z and a thrust axis, rotational symmetry removes the need to optimize over a thrust axis for sThrust, and thus only a single optimization over z is required. The sThrust is actually just a special case of the event 2-sPronginess discussed in section 4.2.2, with the particle positions fixed. An example of a thrust configuration is shown in figure 10.

The event sThrust of an event \mathcal{E} is then given by:

$$\mathcal{O}_{\text{Thrust}}(s_{\mathcal{E}}) = \frac{2}{\pi^2} \min_{z \in [0,1]} \text{SEMD}(s_{\mathcal{E}}, s_{\text{Thrust}}^{(z)}), \quad (4.21)$$

with the factor of $\frac{2}{\pi^2}$ chosen such that the sThrust is normalized between 0 and $\frac{1}{2}$, like the ordinary Thrust.¹⁷ This normalization reflects the choice of using arc-length rather than chord length; the ordinary thrust is typically defined through dot products between the particle momentum and the thrust axis (resulting in a chord length) and is thus slightly different, but in the collinear limit these are linearly related by this normalization factor. One can, of course, have used chord lengths to define sThrust instead in closer correspondence with the ordinary Thrust, which we discuss in appendix A.

¹⁷In our convention, sThrust is low for pencil-like events and large for events away from pencil-like. This differs from the “FASTJET” Thrust, which does the opposite. The sThrust really corresponds to $(1 - \text{Thrust})$.

The SEMD between the thrust event and a general event \mathcal{E} is

$$\begin{aligned} \text{SEMD} = \min_{z \in [0,1]} & \sum_{i < j \in \mathcal{E}} 2E_i E_j \omega_{ij}^2 + 2\pi^2 z(1-z) E_{\text{tot}}^2 \\ & - 2\pi \sum_{\substack{n \in \mathcal{E}^2 \\ \omega_n < \omega_{n+1}}} \omega_n \text{ReLU} \left(S^+(\omega_n) - \max \left[\left(z^2 + (1-z)^2 \right) E_{\text{tot}}^2, S^-(\omega_n) \right] \right). \end{aligned} \quad (4.22)$$

The optimization over z can be performed via gradient descent. However, it can be shown that $z = \frac{1}{2}$ is a local extremum of the SEMD (see appendix A for more details), and is thus not an ideal initial point for gradient descent. The value $z = \frac{1}{2}$ is not necessarily the true minimum, merely an extremum, though in practice we find that the true minimum is extremely close to $z = \frac{1}{2}$ in dijet events. It may also be possible to perform the minimization *exactly* in finite time as a distribution-partitioning problem in spectral space, analogous to the hemisphere-picking problem for ordinary Thrust (see ref. [67] for a review of classical and quantum algorithms), though we do not pursue this further here. There is also a simple test for checking if $z = 0.5$ is the true minimum, given by eq. (A.7) in appendix A.

4.4.2 Event spIsotropy

Next, we define the spectral Isotropy (*spIsotropy*) of an event over the celestial sphere. In analogy with the ordinary Isotropy [5], this is defined as the (spectral) distance to the uniform event over the sphere:

$$\mathcal{E}_{\text{Isotropy}}(x) = \frac{E_{\text{tot}}}{4\pi}. \quad (4.23)$$

The corresponding spectral representation can be calculated to be:

$$s_{\text{Isotropy}}(\omega) = \frac{E_{\text{tot}}^2}{2} \sin(\omega). \quad (4.24)$$

Like the continuous observables described in the previous subsection, the uniform event can be approximated by sampling N points on the sphere, each with energy $\frac{E_{\text{tot}}}{N}$. Then, all one has to do is use the SPECTER algorithm to compute the SEMD between the event of interest \mathcal{E} and this discrete event. There are no free parameters in eq. (4.23), so no minimization is necessary. An example of such a sampling is shown in figure 11.

Like the event sThrust, we find it convenient to normalize the event spIsotropy. By convention, we multiply the SEMD value by a factor of $\frac{1}{\pi-2}$, such that the event spIsotropy of a perfectly back-to-back equal energy event is 1, to match the ordinary spherical Isotropy.

The spIsotropy observable can be written exactly in closed-form, allowing one to bypass the SPECTER algorithm for generic observables. By using eq. (2.7) and the exact form of the cumulative spectral function for a uniform sphere (see appendix B.3 for details), one can show:

$$\mathcal{O}_{\text{Isotropy}}(s_{\mathcal{E}}) = \frac{1}{\pi-2} \left(\sum_{i < j \in \mathcal{E}} 2E_i E_j \omega_{ij}^2 + \frac{\pi^2-4}{2} E_{\text{tot}}^2 - 2 \sum_{\substack{n \in \mathcal{E}^2 \\ \omega_n < \omega_{n+1}}} \omega_n [f^+(n) - f^-(n)] \right), \quad (4.25)$$

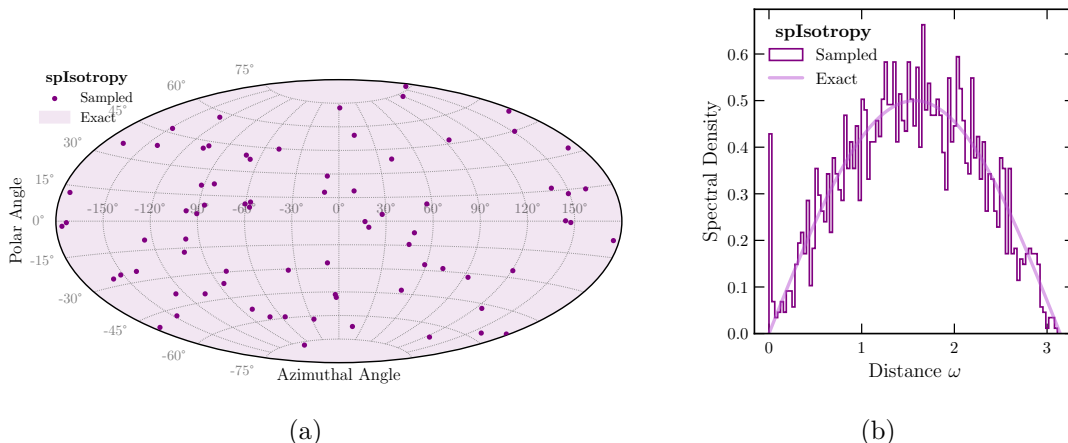


Figure 11. The (a) detector representation from eq. (4.23) and (b) spectral representation from eq. (4.24) of an idealized isotropic event, on the celestial sphere. Each dot is a sampled particle, with its size representing its energy weight, which happens to be uniform.

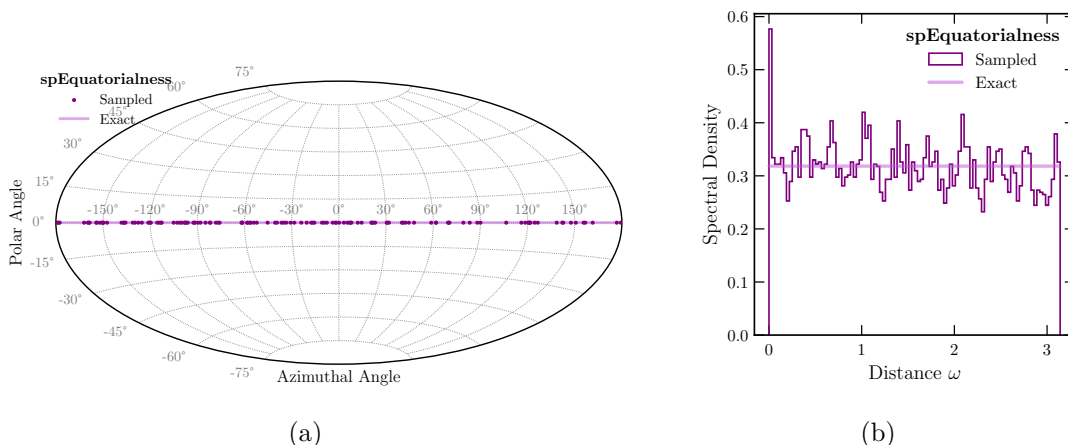


Figure 12. The (a) detector representation from eq. (4.27) and (b) spectral representation from eq. (4.28) of an idealized ring event, on the celestial sphere. Each dot is a sampled particle, with its size representing its energy weight, which happens to be uniform.

where

$$f^\pm(n) = \sqrt{S^\pm(\omega_n)} \sqrt{E_{\text{tot}}^2 - S^\pm(\omega_n)} + S^\pm(\omega_n) \cos^{-1} \left(1 - 2 \frac{S^\pm(\omega_n)}{E_{\text{tot}}^2} \right) - E_{\text{tot}} \sin^{-1} \left(\frac{\sqrt{S^\pm(\omega_n)}}{E_{\text{tot}}} \right). \quad (4.26)$$

4.4.3 Event *spEquatorialness*

On the celestial sphere, we can define the event *spEquatorialness* as the distance of an event to the ring-like event on the equator of the sphere, given by:

$$\mathcal{E}_{\text{Equator}}(x) = \frac{E_{\text{tot}}}{2\pi} \delta \left(\theta - \frac{\pi}{2} \right), \quad (4.27)$$

where θ is the usual polar coordinate of the sphere. Due to rotational invariance, all other great circles on the sphere are the same in spectral space. The corresponding spectral representation is:

$$s_{\text{Equator}}(\omega) = \frac{E_{\text{tot}}^2}{\pi}. \tag{4.28}$$

Just like the event spIsotropy, the event spEquatorialness can be easily estimated by sampling N points uniformly on the equator of the sphere using the SPECTER algorithm. An example of this sampling is shown in figure 12.

However, it is also possible to solve for the SEMD exactly in closed form by using eq. (2.7) and the cumulative spectral function for a ring (see appendix B.4 for details):

$$\begin{aligned} \mathcal{O}_{\text{Equator}}(s_{\mathcal{E}}) = & \sum_{i < j \in \mathcal{E}} 2E_i E_j \omega_{ij}^2 + \frac{\pi^2}{3} E_{\text{tot}}^2 \\ & - \frac{\pi}{E_{\text{tot}}^2} \sum_{\substack{n \in \mathcal{E}^2 \\ \omega_n < \omega_{n+1}}} \omega_n \left[(S^+(\omega_n))^2 - (S^-(\omega_n))^2 \right]. \end{aligned} \tag{4.29}$$

4.5 ... and more!

It is important to emphasize that the above list of shape observables is *not* exhaustive. As with SHAPER, one can use *any* parameterized manifold of events $\mathcal{E}^{(\theta)}$ to define a shape, so long as it is known how to sample from it. This includes more exotic shapes such as ellipses or polygons, but also combinations of shapes such as N -spRinginess, or N -sPronginess plus spIsotropy defined as N prongs on top of a uniform background. All of these are possible within the SPECTER framework.

Many of the observables discussed above were simple enough to admit an exact closed-form expression, bypassing the need to use the SPECTER algorithm at all to estimate them. However, this is not the case for a generic spectral observable — for these, one must use SPECTER to estimate them.

5 Empirical studies

In this section, we use SPECTER to evaluate SEMDs and spectral shape observables on simulated collider events. We begin with a comparison of pairwise (S)EMDs between jets and between full events, to explore the differences between the SEMD and ordinary EMD. Then, we evaluate the spectral shape observables defined in section 4 on both localized jets and global events. These are evaluated using the numeric SPECTER algorithm, and, wherever possible, also in closed-form. For comparison, we show the distribution of the corresponding ordinary EMD-based shapes, as defined in ref. [16] using the SHAPER algorithm for $\beta = 2$.¹⁸

¹⁸To match conventions, we multiply the ordinary EMD as evaluated by SHAPER by a factor of 2, since the metric used in ref. [16] is $\frac{1}{\beta}|x - y|^{\beta/2}$ rather than the $|x - y|^{\beta/2}$ used here. This factor was chosen in ref. [16] so that the gradients of the Sinkhorn divergence have nicer scaling with β and to match onto other metrics in the optimal transport literature.

5.1 Datasets

For the empirical studies in this section, we generate several different sample datasets using PYTHIA 8.3 [68]. The first three datasets consist of localized jets for studying the properties of pairwise (S)EMDs and jet shape observables. The final, fourth dataset consists of full events for the purposes of studying event shape observables.

For each dataset described below, we use 100k samples, with this number chosen because SHAPER (but not SPECTER) is often time-limited. All datasets are publicly available, as described in [Code and data](#) below. The datasets are:

- QCD Jets.** We generate QCD events in proton-proton collisions at $\sqrt{s} = 14$ TeV using PYTHIA’s `HardQCD:a11` process, with a phase space cut of $\hat{p}_T \in [375, 687.5]$ GeV. Multi-parton interactions, initial-state radiation, and final-state radiation are kept on. The events are then clustered into jets using FASTJET 3.3.0 [69] using the anti- k_T algorithm [70] with jet radius $R = 1.0$ (AK10 “fat jets”). Final-state invisible particles, namely neutrinos, are not included in the clustering. Jets are required to satisfy $p_{T,J} \in [500, 550]$ GeV and $\eta_J < |2.5|$. The jets are then saved. If an event has multiple jets meeting all cuts, only one jet is saved, selected randomly. The 125 most energetic particles are saved per jet. These jets are used for the pairwise (S)EMD studies in section 5.2 and for the jet shape studies in section 5.3. An example of a typical jet in this dataset, along with its spectral representation, is shown in figure 13.
- Top Quark Jets.** We generate $t\bar{t}$ events in proton-proton collisions at $\sqrt{s} = 14$ TeV using PYTHIA’s `Top:gg2ttbar` and `Top:qqbar2ttbar` processes, with a phase space cut of $\hat{p}_T \in [375, 687.5]$ GeV. Multi-parton interactions, initial-state radiation, and final-state radiation are kept on. All tops are forced to decay hadronically. Events are then clustered into AK10 jets, exactly as in the QCD jets above. Jets are required to satisfy $p_{T,J} \in [500, 550]$ GeV and $\eta_J < |2.5|$. Additionally, we require that there exists a top quark parton within $\Delta R = 1.0$ of the jet axis. One jet amongst those that satisfy the cuts is randomly selected per event to be saved as a t jet. The 125 most energetic particles are saved per jet. The t jets are used for the pairwise (s)EMD studies in section 5.2.
- Uniform Phase Space “Jets”.** We generate points in uniform phase space using the RAMBO algorithm [71]. Each event has 125 particles with at total $\sqrt{s} = 150$ GeV. These events are then boosted in the x -direction such that their total energy is 500 GeV, which makes the events appear jet-like, but without the singularity structure of QCD. These “jets” are used for the pairwise (S)EMD studies in section 5.2.
- LEP Dijet Events.** We generate $e^+e^- \rightarrow$ hadrons events at $\sqrt{s} = m_Z$ using PYTHIA’s `WeakSingleBoson:ffbar2gmZ` process, in order to simulate LEP-like jets. The intermediate virtual photon or Z is forced to decay to one of the 5 light quarks. Initial-state radiation due to the leptons is turned off. No additional cuts or clustering are applied. An example of a typical event in this dataset, along with its spectral representation, is shown in figure 14.

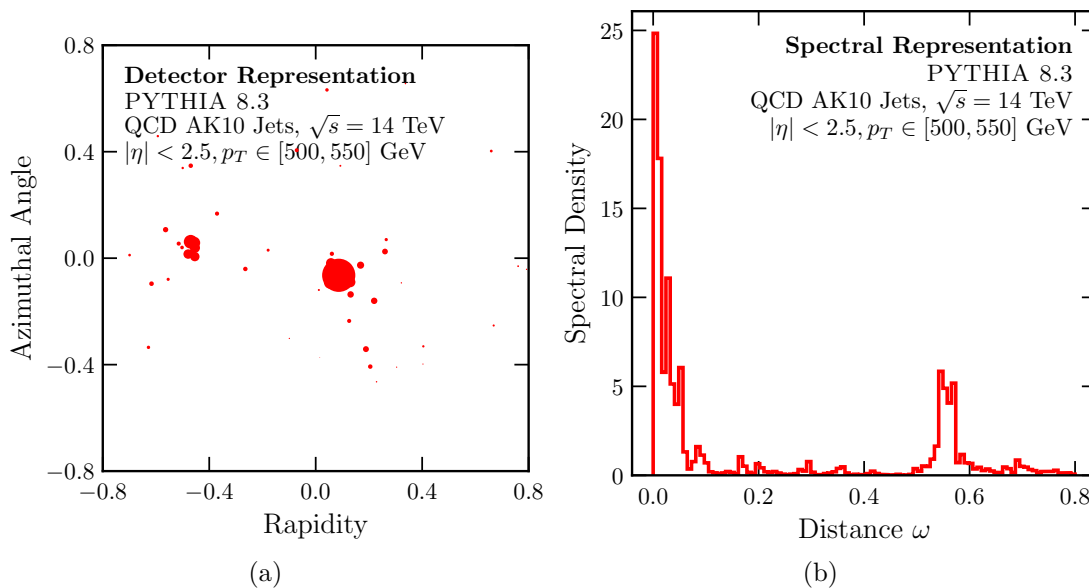


Figure 13. The (a) detector representation and (b) spectral representation of a typical simulated QCD jet, in a patch of the rapidity-azimuth plane.

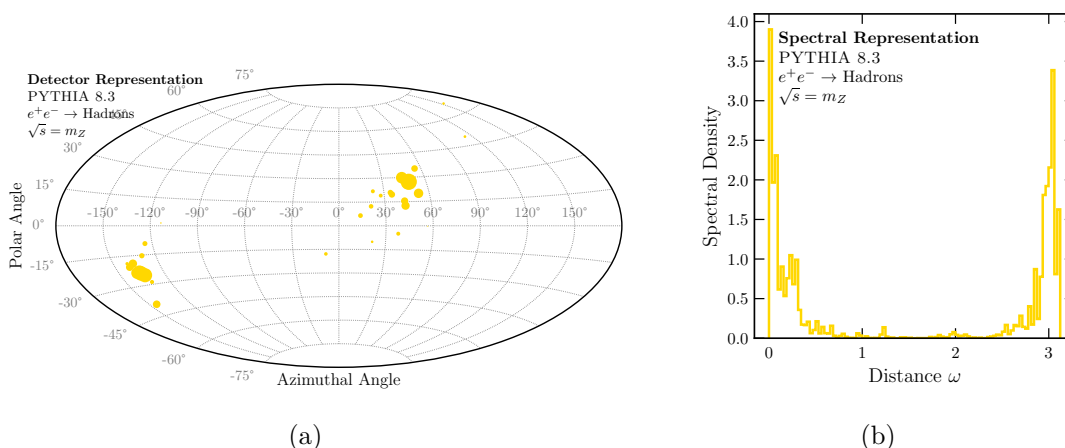


Figure 14. The (a) detector representation and (b) spectral representation of a typical simulated $e^+e^- \rightarrow$ hadrons event, on the celestial sphere.

5.2 Pairwise distances between events and jets

In this subsection, we compute the (spectral) EMD between pairs of jets associated with different underlying processes. As discussed in section 2.4, one should not generally expect the SEMD and ordinary EMD to be the same, though they may be correlated. Whenever an (almost) equilateral triangle configuration occurs between any set of three particles in an event, one should expect a deviation between the SEMD and EMD, as the information probed is genuinely different. These configurations occur whenever the energies of particles are all within an $\mathcal{O}(1)$ number of each other, and their pairwise distances are also all within an $\mathcal{O}(1)$ number of each other. There are two factors we can identify that lead to equilateral triangle configurations:

1. **Number of Particles.** Combinatorially, if an event has N particles, we can form $\binom{N}{3}$ triangles between them. If the energies and positions of particles are uniformly distributed, one should expect that the number of equilateral triangles should scale similarly. In QCD jets, however, emissions off the hard core are strongly ordered in both energy and angle, so forming equilateral triangles with balanced energies involving the hard core is highly suppressed.
2. **Scaling.** Even if there do exist equilateral triangle configurations, if the pairwise $(2E_i E_j)$ or the ω_{ij} scales are small, then the degeneracy will contribute very little to the EMD/SEMD difference. In QCD jets, these are the soft and collinear scales, which are suppressed. In top jets, there is an additional hard scale, m_t/E , to contend with.

Thus, we should expect that as we increase the number of particles in an event, or as we change the event type to include harder-scale physics, that the EMD and SEMD become more different. To show this, we calculate pairwise SEMDs (using SPECTER) and isometry-modded EMDs (using SHAPER) between QCD jets, top quark jets, and uniform phase space, each described in section 5.1. Moving from QCD to top jets to uniform phase space, one should expect that the emissions are less strongly-ordered and that there exist harder and wider scales to form equilateral triangles. Each dataset is clustered into 2, 5, and 25 subjets with the exclusive k_T -algorithm [59, 60] in FASTJET [69].

Unlike the SEMD, the ordinary EMD is not invariant under isometries. To alleviate this and put our comparisons on fairer ground, we use the “Tangent EMD” [72], wherein we “mod out” the isometries by minimizing the EMD over all possible translations and rotations using SHAPER. Precisely, this is given by:

$$\text{Tangent EMD}(\mathcal{E}, \mathcal{E}') = \min_{\theta \in E(2)} \text{EMD}(\mathcal{E}, \mathcal{E}'^{(\theta)}), \tag{5.1}$$

where $E(2)$ is the two-dimensional Euclidean group, and $\mathcal{E}'^{(\theta)}$ is a translated and rotated energy flow parameterized by θ .

The results of this study are shown in figure 15. In the grid of plots, N is increased going downwards and harder physics is shown going rightwards. In both directions, the SEMD to EMD difference grows relative to the baseline $N = 2$ QCD jets in the upper left. Moreover, the EMD is often greater than the SEMD, but this is not always the case. We can make several observations from this plot:

- **SEMD \approx EMD in QCD jets.** Even across a large number of k_T subjets, the EMD and SEMD are very close, especially compared to top jets or uniform phase space. This follows from the above arguments — there are typically no additional hard emissions in QCD jets that can “spoil” the spectral representation with a near-degeneracy.
- **SEMD \lesssim EMD generically.** It seems there is a slight bias for the EMD to be greater than the SEMD rather than the other way around. It is hard to tell if this is a result of imperfect minimization while calculating the Tangent EMD, or a robust feature — one might expect that the existence of near-degeneracies would generically decrease distances in spectral space, though section 7.3 of ref. [23] provides an example where the SEMD can be greater.

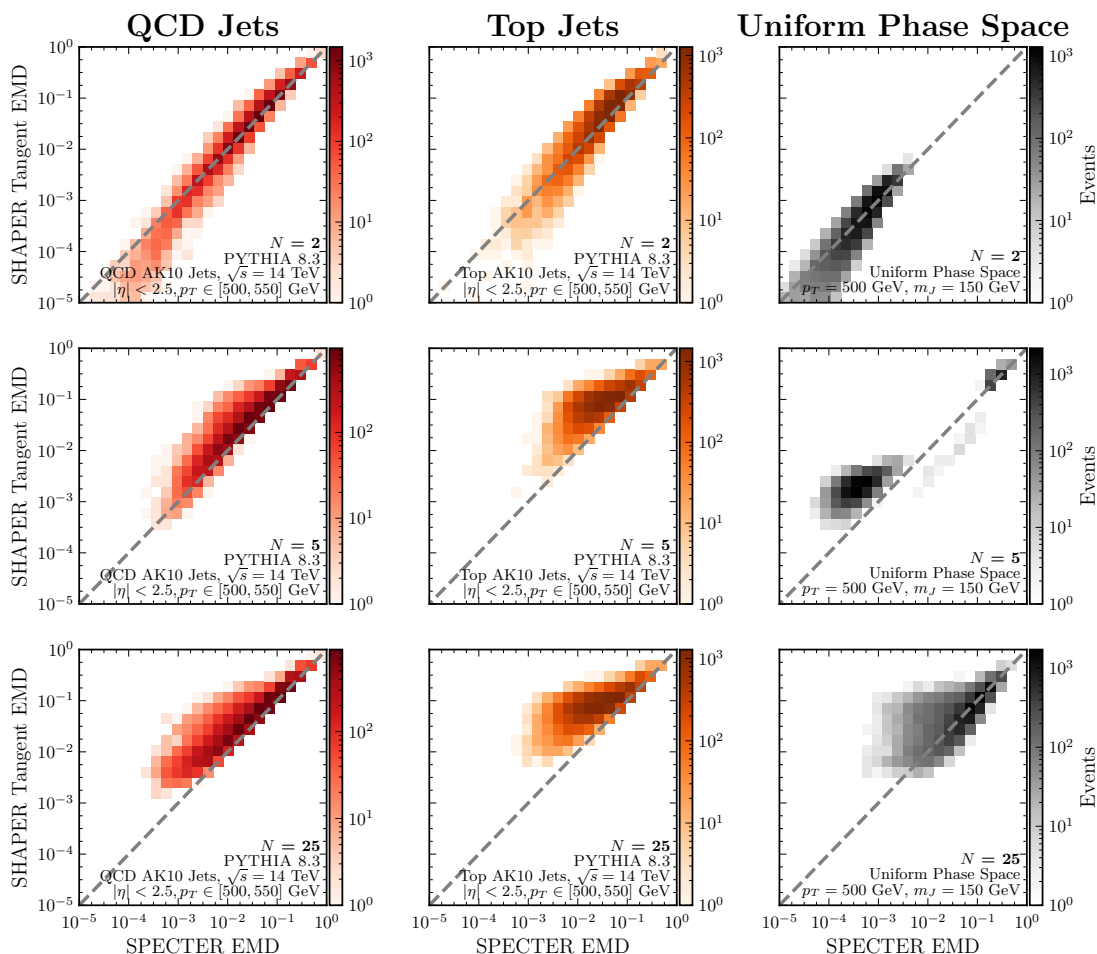


Figure 15. Comparison of the SEMD, as evaluated using SPECTER, to the isometry-modded EMD, as evaluated using SHAPER, on pairs of events. Pairs of QCD jets are used for the left column (red), pairs of top jets are used for the middle column (orange), and pairs of uniform boosted phase space are used for the right column (yellow). Each jet is clustered into subjects before the (S)EMD is computed: 2 subjects in the top row, 5 subjects in the middle row, and 25 subjects in the bottom row. Note that the histogram is in log scale.

- **Differences grow with N_{subjects} .** The relative SEMD to EMD difference of the distributions grows as the number of subjects increase. In particular, for $N = 2$, the two are highly correlated across all three types of jets.

There is clear evidence of a physics-dependent correlation between the EMD and SEMD, potentially intertwined with the collinear and strongly-ordered structure of QCD. There is more to be done to elucidate the relationship between these quantities, which we leave for future work.

5.3 Hearing jets spRing and sProng

In this subsection, we evaluate the jet shape observables defined in section 4 on the QCD jets dataset described in section 5.1. For each observable, we evaluate both the spectral shape observable $\mathcal{O}_{\text{Shape}}(s_{\mathcal{E}})$ and the corresponding spectral shape parameters $\theta_{\text{Shape}}(s_{\mathcal{E}})$ on the

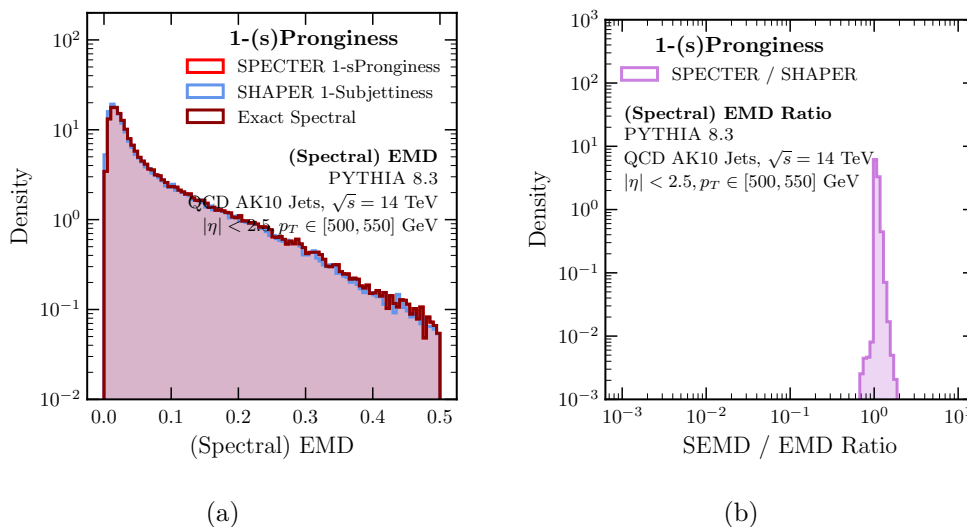


Figure 16. (a) Distributions of the 1-(s)Pronginess observable, as defined in section 4.2.1, evaluated on jets using SPECTER (red) and SHAPER (purple). (b) The corresponding ratio of the sPronginess to the Pronginess.

distribution of events \mathcal{E} using SPECTER. We also show the distributions for the corresponding EMD-based observables defined using the same shapes, as evaluated using SHAPER.

For both programs, we sample $N = 125$ points for each shape. The gradient descent operation is run for 500 epochs with the Adam optimizer with a learning rate of 0.001. Often, convergence is reached by ~ 100 epochs, but we do not do any early stopping to be conservative. SHAPER requires two additional parameters that control the Sinkhorn estimation: we choose $\epsilon = 10^{-3}$ and $\Delta = 0.99$ for results as precise as possible.

In figures 16 and 17, we show the 1-(s)Pronginess, 2-(s)Pronginess, and 3-(s)Pronginess, as defined in sections 4.2.1 and 4.2.2. The spectral 1-sPronginess and the ordinary 1-Pronginess (1-subjettiness) should be identical — indeed, we see that the ratio of the two is close to 1. Deviations from 1 are due to numerics: both SPECTER and SHAPER use a finite gradient descent algorithm (though SPECTER’s gradient descent happens to be irrelevant for 1-sPronginess), and SHAPER uses an approximation of the EMD rather than the exact EMD. The spectral 2- and 3-sPronginess, on the other hand, are significantly less than their ordinary EMD counterparts, and are less correlated.

In figures 18, 19, and 20, we show the jet (sp)Ringiness, the jet (sp)Lineliness, and the jet (s)Diskiness respectively, as defined in section 4.3. Each of these observables is defined with a corresponding length scale parameter (radii for the rings and disks, and lengths for the lines). For each observable, we show the (spectral) EMD (solid) as well as the corresponding length scale parameter (dashed) as calculated using SPECTER and SHAPER. In all cases, the spectral observable is smaller than the corresponding ordinary observable by factor of about 2 to 5. With the exception of the 1-sPronginess, which is expected to exactly match the 1-subjettiness, this ratio is dataset dependent and difficult to predict. For instance, one can construct a dataset of near-equilateral triangles, in which the 2-subjettiness is infinitely larger than the 2-sPronginess, as in figure 2(b). Understanding the origin of this factor of 2 to 5

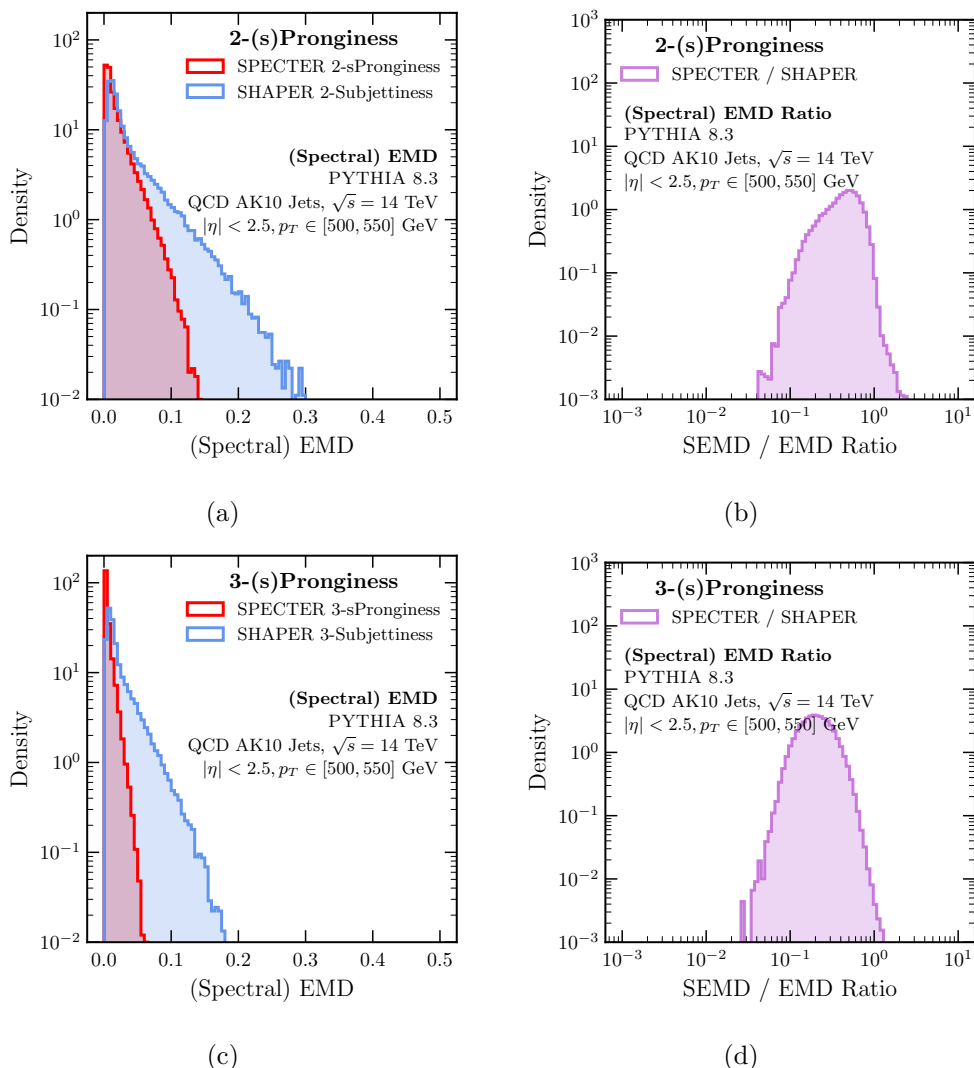


Figure 17. Distributions of the 2- and 3- (s)Pronginess observables, as defined in section 4.2.2, evaluated on jets using SPECTER (red) and SHAPER (purple) in (a) and (c), respectively. The corresponding ratio of the sPronginess to the Pronginess is shown in (b) and (d), respectively.

times would require a full QCD computation, which we leave for future work. Interestingly, the two circular observables achieve essentially the same radius using either the spectral or ordinary metric, but the spLineliness tends to have a length parameter 2 times bigger than the ordinary Lineliness. Notably, all three parameter ratios peak sharply near either 1 or 2, whereas the corresponding (spectral) EMD ratios are broad — changing the metric seems to scale the physical distance between events, which is not unexpected, but does not seem to change the optimal geometry by much.

5.4 Event shapes

In this subsection, we evaluate the event shape observables defined in section 4 on the e^+e^- dataset described in section 5.1, namely the event sThrust, event spIsotropy, and event spRinginess.

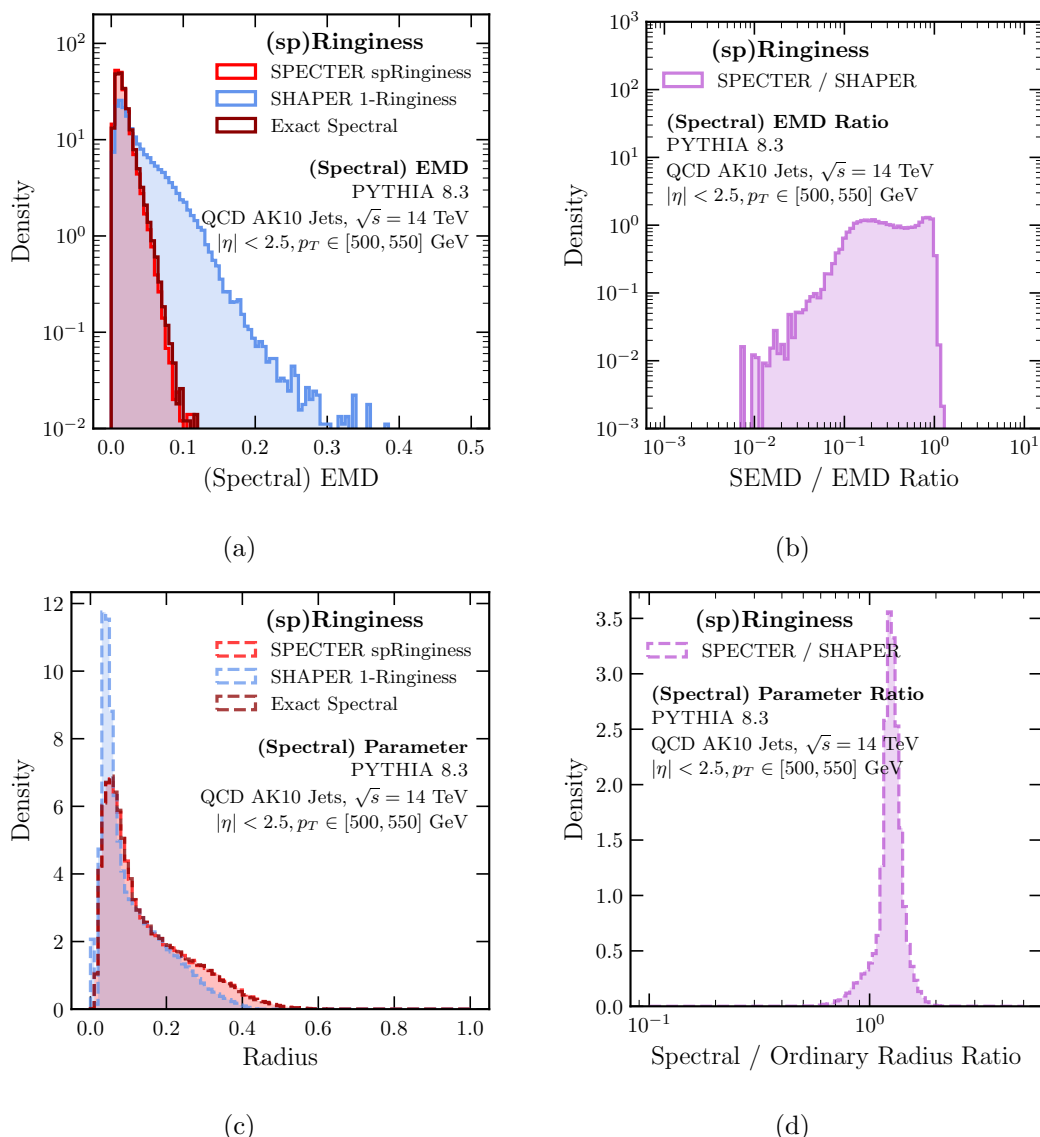


Figure 18. Distributions for the (sp)Ringiness observable, as defined in section 4.3.1, evaluated on jets using SPECTER (red) and SHAPER (purple). The observable value is shown in (a), and the corresponding shape parameter is shown in (c). The ratio of the spRinginess to the Ringiness is shown in (b) and (d).

In figure 21, we show the sThrust, as computed by the SPECTER algorithm, and $(1 - \text{Thrust})$, as calculated using FASTJET 3.3.0. The two distributions agree extremely well near (S)EMD values of 0, with the distribution of ratios sharply peaked at 1, indicating that the sThrust and Thrust behave similarly in the collinear limit, as expected. At larger values of the (S)EMD, however, the two distributions diverge, and the SEMD does not reach these higher values. This is a consequence of the fact that the sThrust and Thrust use different ground metrics. The sThrust uses the arc-length metric of eq. (2.22), but ordinary Thrust corresponds to the chord-length metric, $\omega^2 \sim 1 - \cos(\theta)$. If one were to define sThrust using

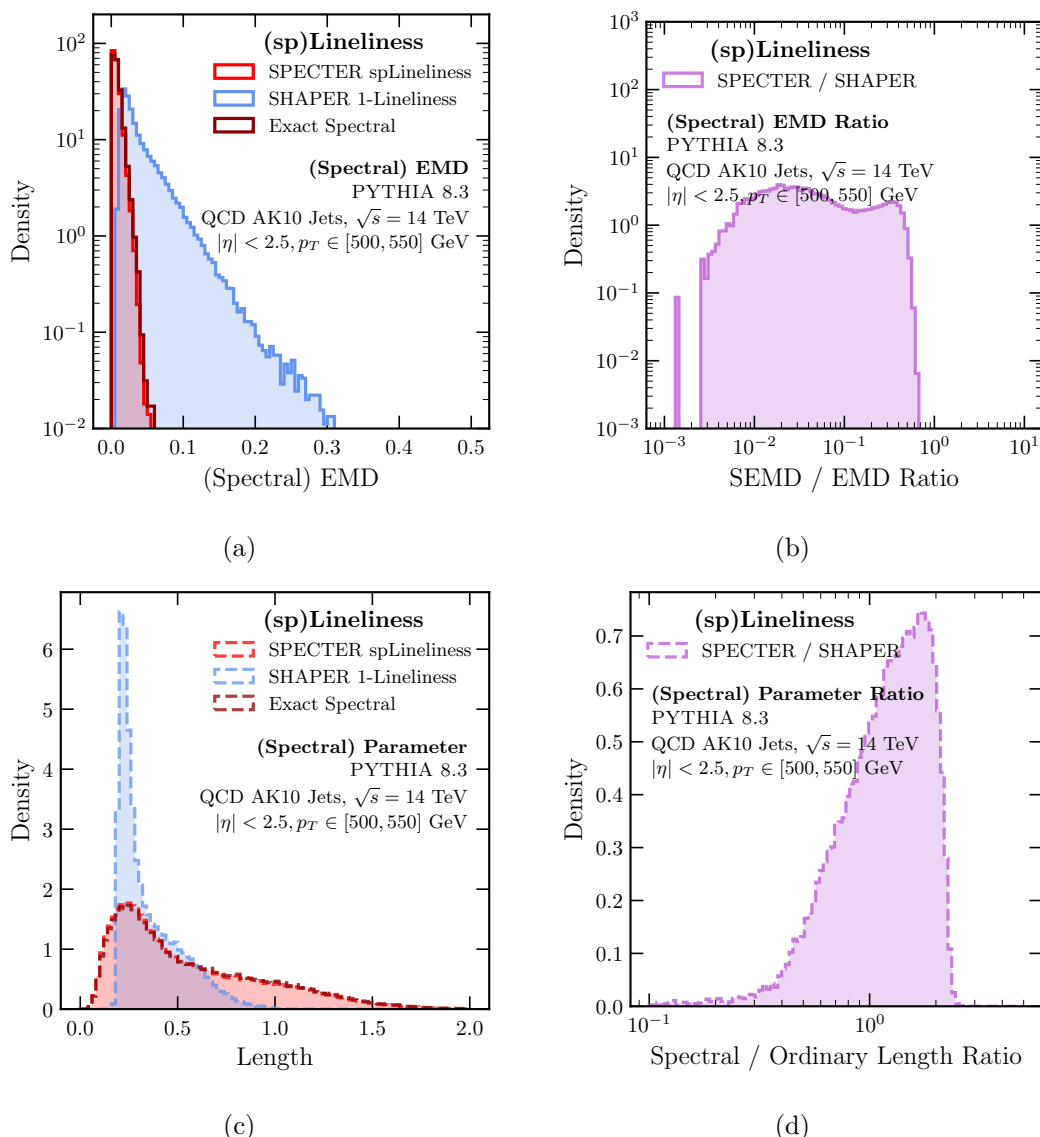


Figure 19. Distributions for the (sp)Lineliness observable, as defined in section 4.3.2, evaluated on jets using SPECTER (red) and SHAPER (purple). The observable value is shown in (a), and the corresponding shape parameter is shown in (c). The ratio of the spLineliness to the Lineliness is shown in (b) and (d).

this metric as well, the sThrust and Thrust distributions become nearly identical, as we explore further in appendix A.

In figure 22, we show the spIsotropy, as computed by the SPECTER algorithm, as well as the exact closed-form expression in eq. (4.25). We also show the ordinary spherical event Isotropy, as calculated using the Python Optimal Transport library [22], rather than SHAPER, since SHAPER in its current form is unable to handle manifolds with local curvature. The ordinary Isotropy is normalized such that the spIsotropy and Isotropy agree on 2-particle events. Dijet events are not isotropic, and indeed we see that both the spIsotropy and

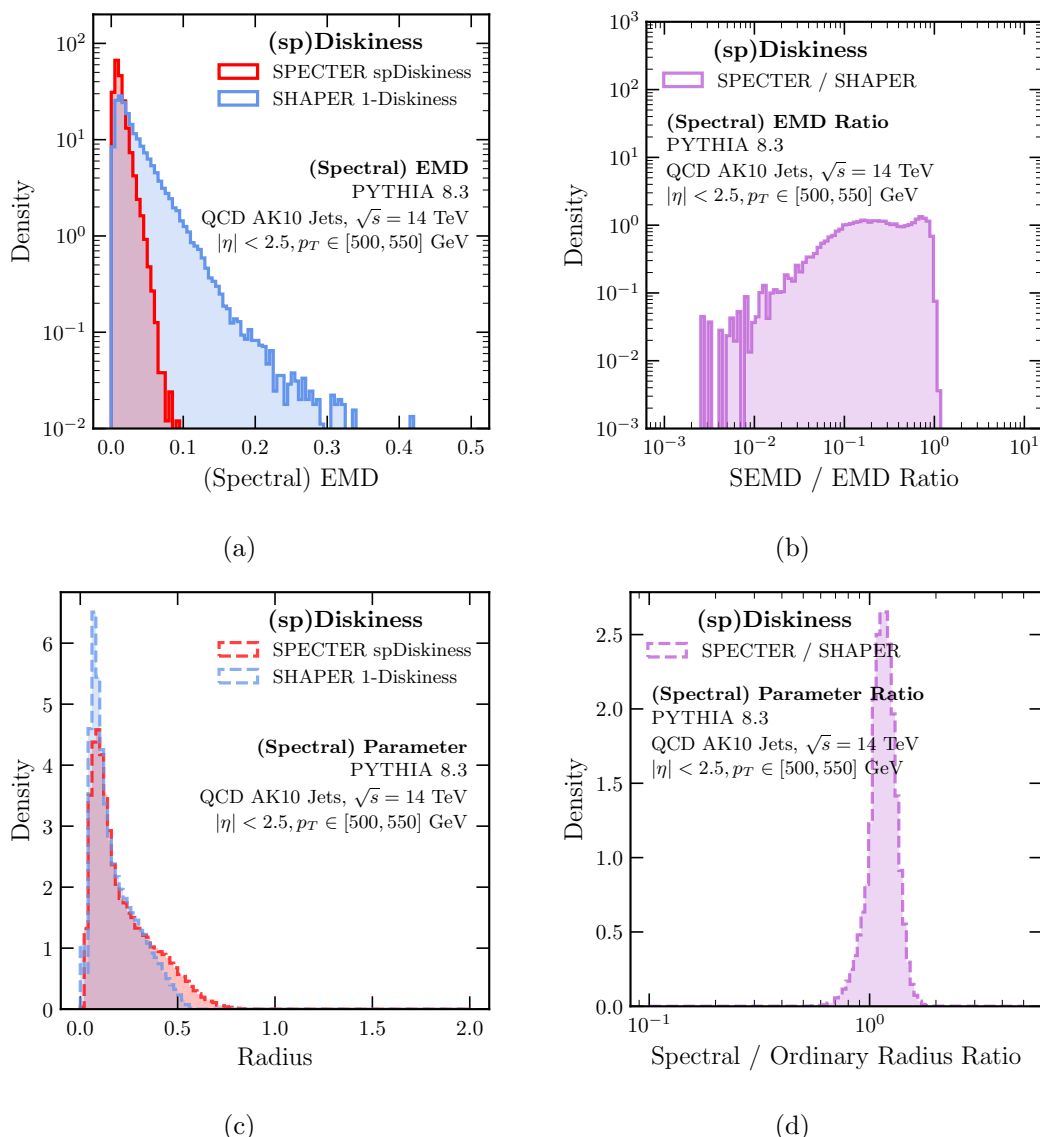


Figure 20. Distributions for the (sp)Diskiness observable, as defined in section 4.3.3, evaluated on jets using SPECTER (red) and SHAPER (purple). The observable value is shown in (a), and the corresponding shape parameter is shown in (c). The ratio of the spDiskiness to the Diskiness is shown in (b) and (d).

Isotropy take large values away from zero. As is the case with the jet shape observables, the spIsotropy tends to be strictly less than its ordinary counterpart, with the ratio $\lesssim 1$. Unlike the event sThrust, the spIsotropy differs greatly from its ordinary counterpart, but this is to be expected: the fully isotropic event has a (nearly) highly degenerate spectral function, since it is easy to form (nearly) equilateral triangles from uniformly sampled particles, and thus we are far from the non-degenerate region probed by thrust. We also observe a small bump in the spIsotropy distribution near 0 — we speculate that this may be due to degenerates caused by some additional prongy structure in the dijet events.

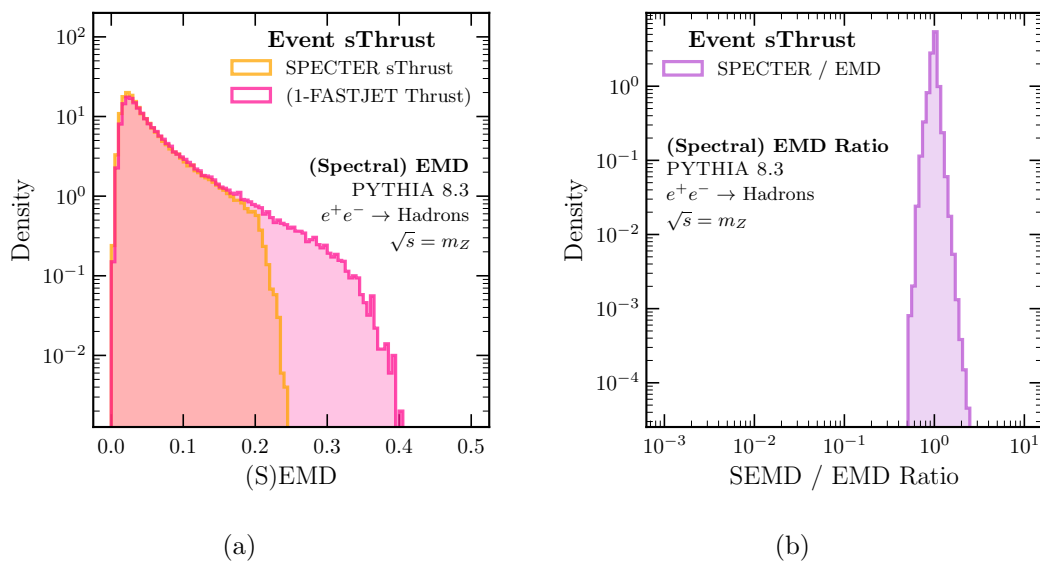


Figure 21. Distributions for the event (s)Thrust observable, as defined in section 4.4.1, evaluated on jets using SPECTER (orange), and the FASTJET 3.3.0 (pink). The observable value is shown in (a) and the ratio of the sThrust to the Thrust is shown in (b). A variant of the sThrust based on chord lengths rather than arc lengths is shown in figure 24.

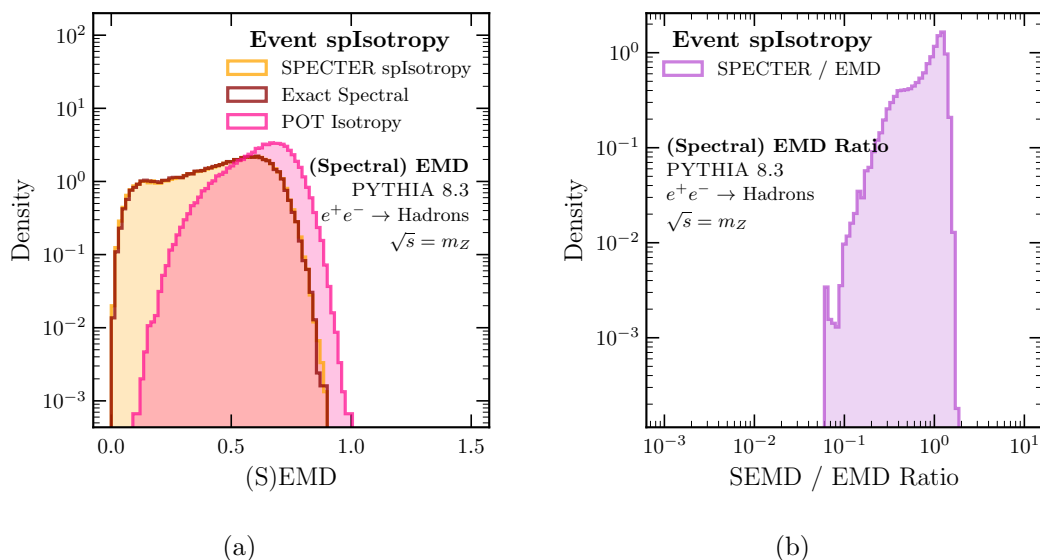


Figure 22. Distributions for the event (sp)Isotropy observable, as defined in section 4.4.2, evaluated on jets using SPECTER (orange) and the Python Optimal Transport (POT) library (pink). The observable value is shown in (a) and the ratio of the spIsotropy to the Isotropy is shown in (b).

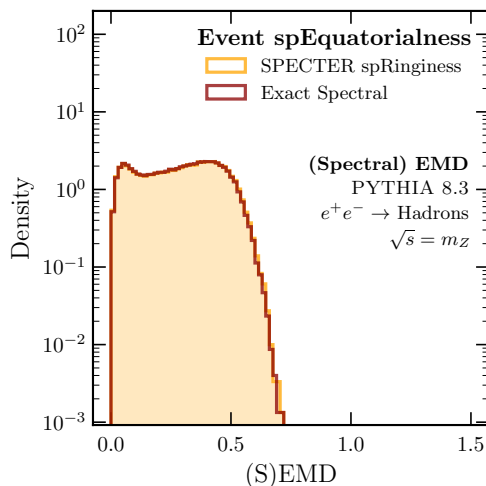


Figure 23. (a) Distribution for the event spEquatorialness observable, as defined in section 4.4.3, evaluated on $e^+e^- \rightarrow$ hadrons events using SPECTER (orange) and exactly using eq. (4.29) (dark red).

Finally, in figure 23, we show the event spEquatorialness distribution as evaluated by SPECTER, and evaluated exactly using eq. (4.29). Here, we are unable to evaluate a corresponding observable using the ordinary EMD, as SHAPER cannot handle local curvature and the Python Optimal Transport library cannot efficiently optimize the rotational degrees of freedom of the ring over the sphere. Interestingly, like the spIsotropy, there is also a bump in the distribution near 0, but even larger. These bumps appear to be a generic feature of spectral event shape observables that potentially probe interesting physics, and we leave their analysis to future work.

6 Conclusion

In this paper, we introduced an efficient algorithm for evaluating the $p = 2$ SEMD. The $p = 2$ SEMD takes on a simple closed form, which enables an exact computation between any pair of discrete events in $\mathcal{O}(N^2 \log N)$ time, and is to our knowledge the only exactly calculable metric on collider events. In addition, the SEMD is *fast*: using the SPECTER framework, which is an efficient and highly parallelized implementation of the SEMD, it is possible to perform SEMD computations at scale, and in practical cases can even go as fast as $\mathcal{O}(N)$. The SEMD is highly correlated with the ordinary EMD, especially for QCD jets, and its fast-to-evaluate, closed-form nature makes it suitable to perform precision studies of the geometry of QCD.

The speed of SPECTER makes previously infeasible analyses on the geometry of collider events possible. It is now feasible, for example, to compute distances between all $\binom{N_{\text{Events}}}{2}$ pairs of events in a dataset. This, for instance, makes it easier to perform correlation dimension studies [42], and makes it possible to extend machine learning scaling law studies far beyond what is possible with the ordinary EMD [73]. The idea of a metric on “theory space”, or Σ MD, has also been proposed [1, 23], but to our knowledge has never been empirically studied due to being prohibitively expensive. The ground metric of the Σ MD is itself the EMD between events, requiring $\binom{N_{\text{Events}}}{2}$ evaluations of the EMD. SPECTER may be able to make

the “Spectral Σ MD” [23] finally tractable, which would potentially enable a novel method to probe the QCD β -function and the anomalous dimensions of other operators.

Moreover, as the SEMD is an IRC-safe metric between events, we have shown how the SHAPER prescription of ref. [16] can be carried over to define a new set of IRC-safe shape observables. Unlike observables defined via the ordinary EMD with SHAPER, however, the spectral observables defined via the SEMD with SPECTER are far easier to evaluate due to SPECTER’s speed and exact metric computation, and in many cases the observables even admit exact closed-form expressions. For instance, while the 2D ordinary event Isotropy does not seem to admit a closed form solution and is expensive to calculate, the event spIsotropy has a closed form enabling easier theoretical study, and is practically instantaneous to evaluate on huge batches of events at once. SPECTER makes the analysis of a huge class of complex jet observables far more computationally feasible, potentially higher upstream in the LHC analysis chain, and is a step closer towards being able to study complex observables at trigger level.

There are a variety of open questions and future directions not addressed in this paper. Now that we have closed-form metrics and even closed-form shapes, an obvious next step is to attempt to compute their distributions in QCD. While ref. [23] made progress towards fixed order calculations for the $p = 1$ SEMD, is it possible to go beyond? The EMD and SEMD seem highly correlated in QCD jets, but not other physics scenarios — while we have speculated that this is due to the degeneracy structure of spectral space, the precise mechanism requires further explanation. Indeed, the difference or ratio between the EMD and SEMD may itself be an interesting observable, as they behave differently depending on the context. Throughout this paper, we have primarily focused on balanced optimal transport, but in principle, it is possible to extend the ideas here (and the SPECTER code) to accommodate unbalanced optimal transport and derived observables. We wish to emphasize that we have only shown a small, *finite* subset of possible shape observables in this paper — the class of observables is truly infinite, and limited only by one’s ability to parameterize arbitrary 2D distributions. Finally, there exists the possibility for the SPECTER code to be faster: it was developed in Python, rather than a fully compiled language, and written by a physicist rather than computer scientist — there is almost certainly potential to be even better!

Code and data. Version 1.0.0 of the SPECTER framework is available at <https://github.com/rikab/SPECTER>. It may also be installed via PyPi using the command `pip install specterpy`. Note that a JAX installation is required for SPECTER to work. Several example and tutorial scripts for operating SPECTER may be found in the subfolder <https://github.com/rikab/SPECTER/tree/main/examples>. Additionally, the code used to perform all of the analyses and produce all of the plots presented in this paper may be found in <https://github.com/rikab/SPECTER/tree/main/studies>.

The datasets described in section 5.1 are also available for use. The SPECTER package can be used to automatically download and load them — see the examples folder. These datasets can also be accessed independently of `specterpy` using the `ParticleLoader` package, hosted at <https://github.com/rikab/ParticleLoader>.

Acknowledgments

We would like to thank Samuel Alipour-fard, Cari Cesarotti, Matt LeBlanc, and Melissa van Beekveld for useful and interesting discussions and comments.

R.G. and J.T. are supported by the National Science Foundation under Cooperative Agreement PHY-2019786 (The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, <http://iaifi.org/>), and by the U.S. DOE Office of High Energy Physics under grant number DE-SC0012567. J.T. is additionally supported by the Simons Foundation through Investigator grant 929241.

A More about sThrust: chord lengths and energy fractions

The “ordinary” Thrust observable [64–66] is based off of the chord length between points on the celestial sphere rather than the arc length. The sThrust also has a parameter to optimize over, the energy fraction z , which unlike other spectral shape observables (such as spRinginess) does not take a simple closed form. In this appendix, we briefly describe what it would look like to use chord length as the basis for the sThrust, rather than the arc length used in section 4.4.1, and discuss the optimization of the z parameter.

If we define the angular measure as

$$\omega_{ij}^2 = \frac{1 - \cos \theta_{ij}}{2}, \quad (\text{A.1})$$

then the spectral function for a perfect, narrow dijet event is

$$s_{\text{Thrust}'}(\omega) = \left(z^2 + (1 - z)^2 \right) E_{\text{tot}}^2 \delta(\omega) + 2z(1 - z) E_{\text{tot}}^2 \delta(\omega - 1), \quad (\text{A.2})$$

on events with total energy E_{tot} in the center-of-mass frame and energies zE_{tot} and $(1 - z)E_{\text{tot}}$ in each hemisphere. With this definition of the angular measure, note that the second moment of the thrust spectral function is

$$\int d\omega \omega^2 s_{\text{Thrust}'}(\omega) = 2z(1 - z) E_{\text{tot}}^2, \quad (\text{A.3})$$

and the second moment of the spectral function for a general event is

$$\sum_{i < j \in \mathcal{E}} 2E_i E_j \omega_{ij}^2 = \frac{1}{2} \sum_{i < j \in \mathcal{E}} 2E_i E_j (1 - \cos \theta_{ij}) = \frac{E_{\text{tot}}^2}{2}, \quad (\text{A.4})$$

in the center-of-mass frame. Therefore, the SEMD between this thrust event and a general event \mathcal{E} is

$$\begin{aligned} \text{SEMD} &= \frac{E_{\text{tot}}^2}{2} + 2z(1 - z) E_{\text{tot}}^2 \\ &\quad - 2 \sum_{\substack{n \in \mathcal{E}^2 \\ \omega_n < \omega_{n+1}}} \omega_n \text{ReLU} \left(S^+(\omega_n) - \max \left[\left(z^2 + (1 - z)^2 \right) E_{\text{tot}}^2, S^-(\omega_n) \right] \right). \end{aligned} \quad (\text{A.5})$$

The chord-length sThrust is then the minimum SEMD over z .

In figure 24, we show the distribution of this chord-based sThrust versus the ordinary Thrust. The distributions are practically identical, and the ratio of the two observables is

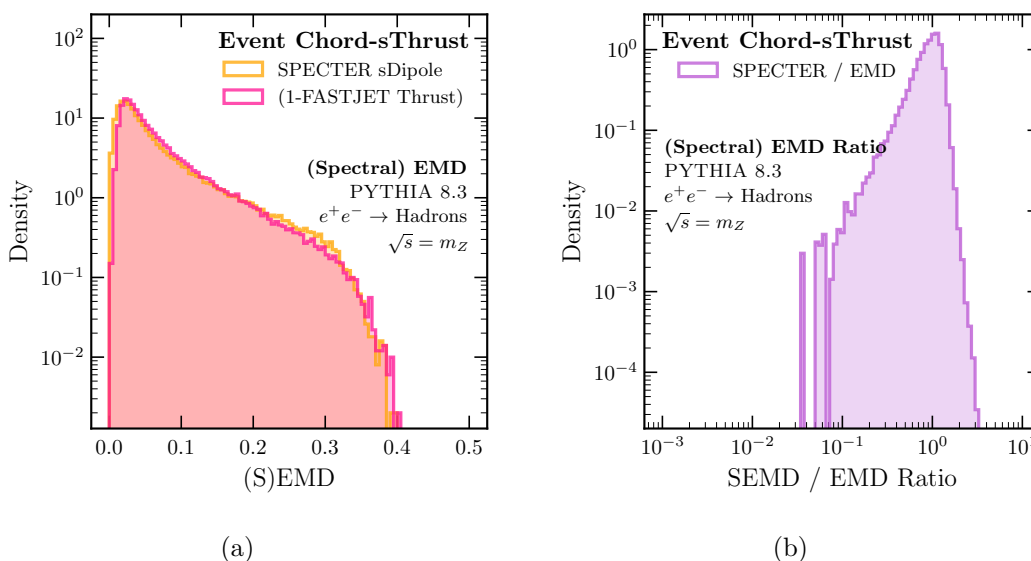


Figure 24. The same as figure 21, but using the chord-based sThrust rather than the arc-length-based sThrust.

sharply peaked at 1. Note that there is some spread around 1 — one might expect this is due to the finite nature of the gradient descent algorithm in finding the optimal z for the sThrust, but gradient descent can only ever overestimate the sThrust — any deviations towards smaller values are due to genuine differences between the two observables. Nevertheless, for these back-to-back dijet events, the chord-based sThrust and Thrust are nearly identical, as expected in the collinear limit. The differences between the two observables seen in figure 21, then, are likely solely due to the difference in ground metric.

For both arc-length sThrust and chord-length sThrust, the value $z = \frac{1}{2}$ is significant, and empirically is almost always the minimum argument in dijet events. Because the spectral thrust event is symmetric under $z \leftrightarrow (1 - z)$, the value $z = \frac{1}{2}$ is a local extremum. We can check if this is a (local) maximum or a minimum with the second derivative (for the following expressions, we use the chord-length sThrust, but analogous statements hold for the arc-length sThrust with additional factors of π):

$$\left. \frac{d^2}{dz^2} \text{SEMD} \right|_{z=1/2} = -4 + 8 \sum_{\substack{n \in \mathcal{E}^2 \\ \omega_n < \omega_{n+1}}} \omega_n \Theta \left(S^+(\omega_n) - \frac{1}{2} \right) \Theta \left(\frac{1}{2} - S^-(\omega_n) \right). \quad (\text{A.6})$$

Therefore, there is a simple test if $z = 1/2$ is the minimum. If

$$\sum_{\substack{n \in \mathcal{E}^2 \\ \omega_n < \omega_{n+1}}} \omega_n \Theta \left(S^+(\omega_n) - \frac{1}{2} \right) \Theta \left(\frac{1}{2} - S^-(\omega_n) \right) > \frac{1}{2}, \quad (\text{A.7})$$

then $z = 1/2$ is a local minimum. In fact, $z = \frac{1}{2}$ is the true *global* minimum. If eq. (A.7) is true, this means that at least half the mass of the event spectral function lies to the right of $\omega = \frac{1}{2}$, which is to be optimally transported to the sThrust spectral function’s peak at

$\omega = 1$. However, this peak can only support a mass of $2z(1-z)$, which maxes out at $\frac{1}{2}$ at $z = \frac{1}{2}$, and so this is the optimal value.

On the other hand, if eq. (A.7) is not true, then the minimum is displaced from $z = 1/2$. In that case, the optimal value of z then corresponds to the amount of squared energy transported from the region where $\omega < 1/2$ to $\omega = 0$, as the solution of

$$z^2 + (1-z)^2 = \frac{1}{E_{\text{tot}}^2} \int_0^{1/2} d\omega s(\omega). \quad (\text{A.8})$$

The precise value of z actually isn't that important, because the optimal transport plan for the SEMD is to move all energy at less than $\omega = 1/2$ to $\omega = 0$, and to move all energy above $\omega = 1/2$ to $\omega = 1$.

B Closed-form expressions for spectral observables

Many of the observables in section 4 admit exact closed-form expressions. In this appendix, we go over the derivations of these expressions.

B.1 Closed-form jet spRinginess

For a uniform energy distribution in the shape of a circle of radius R in a jet with total energy E_{tot} , the cumulative spectral function is

$$S_{\text{Jet Ring}}(\omega) = \frac{2E_{\text{tot}}^2}{\pi} \sin^{-1} \left(\frac{\omega}{2R} \right). \quad (\text{B.1})$$

The inverse cumulative spectral function is then

$$S_{\text{Jet Ring}}^{-1}(E^2) = 2R \sin \left(\frac{\pi}{2} \frac{E^2}{E_{\text{tot}}^2} \right). \quad (\text{B.2})$$

The integral of the square of this is

$$\int_0^{E_{\text{tot}}^2} dE^2 \left(S_{\text{Jet Ring}}^{-1}(E^2) \right)^2 = 2E_{\text{tot}}^2 R^2. \quad (\text{B.3})$$

We also need the integral

$$\begin{aligned} \int dE^2 S^{-1}(E^2) S_{\text{Jet Ring}}^{-1}(E^2) &= 2R \sum_{\substack{n \in \mathcal{E}^2 \\ \omega_n < \omega_{n+1}}} \omega_n \int_{S^-(\omega_n)}^{S^+(\omega_n)} dE^2 \sin \left(\frac{\pi}{2} \frac{E^2}{E_{\text{tot}}^2} \right) \\ &= \frac{4R}{\pi} E_{\text{tot}}^2 \sum_{\substack{n \in \mathcal{E}^2 \\ \omega_n < \omega_{n+1}}} \omega_n \left[\cos \left(\frac{\pi}{2E_{\text{tot}}^2} S^-(\omega_n) \right) - \cos \left(\frac{\pi}{2E_{\text{tot}}^2} S^+(\omega_n) \right) \right]. \end{aligned} \quad (\text{B.4})$$

With these results, we note that the SEMD between a jet and the uniform ring configuration is quadratic and convex in the radius parameter R . As such, there is a unique, global minimum that can be determined and established as the optimal value of the radius. This optimal radius $R_{\text{opt}}(s_{\mathcal{E}})$ is

$$R_{\text{opt}} = \frac{2}{\pi} \sum_{\substack{n \in \mathcal{E}^2 \\ \omega_n < \omega_{n+1}}} \omega_n \left[\cos \left(\frac{\pi}{2E_{\text{tot}}^2} S^-(\omega_n) \right) - \cos \left(\frac{\pi}{2E_{\text{tot}}^2} S^+(\omega_n) \right) \right]. \quad (\text{B.5})$$

These results can then be used to construct the observable of eq. (4.12).

B.2 Closed-form spLineliness

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

$$S_{\text{line}}(\omega) = \left(\frac{2\omega}{L} - \frac{\omega^2}{L^2} \right) E_{\text{tot}}^2. \quad (\text{B.6})$$

The inverse cumulative spectral function is then

$$S_{\text{line}}^{-1}(E^2) = \left(1 - \sqrt{1 - \frac{E^2}{E_{\text{tot}}^2}} \right) L. \quad (\text{B.7})$$

Its squared integral is then

$$\int dE^2 S_{\text{line}}^{-1}(E^2)^2 = \frac{L^2}{6} E_{\text{tot}}^2. \quad (\text{B.8})$$

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

$$\begin{aligned} \int dE^2 S^{-1}(E^2) S_{\text{line}}^{-1}(E^2) &= L \sum_{\substack{n \in \mathcal{E}^2 \\ \omega_n < \omega_{n+1}}} \omega_n \int_{S^-(\omega_n)}^{S^+(\omega_n)} dE^2 \left(1 - \sqrt{1 - \frac{E^2}{E_{\text{tot}}^2}} \right) \\ &= L \sum_{\substack{n \in \mathcal{E}^2 \\ \omega_n < \omega_{n+1}}} \omega_n \left[S^+(\omega_n) + \frac{2}{3E_{\text{tot}}} \left(E_{\text{tot}}^2 - S^+(\omega_n) \right)^{3/2} \right. \\ &\quad \left. - S^-(\omega_n) - \frac{2}{3E_{\text{tot}}} \left(E_{\text{tot}}^2 - S^-(\omega_n) \right)^{3/2} \right]. \end{aligned} \quad (\text{B.9})$$

As we observed with the jet spRinginess, the SEMD between the uniform line configuration and a jet is quadratic and convex in the length parameter L , and so it can be optimized by setting it equal to its unique minimum. These results can then be used to construct the observable of eq. (4.16).

B.3 Closed-form spIsotropy

Consider a continuous energy distribution on the celestial sphere that is perfectly isotropic. The spectral function for this energy distribution can be calculated by fixing a reference particle to be located at the north pole, and the other particle at some point with polar angle θ and azimuthal angle ϕ . The spectral function then follows from

$$s_{\text{iso}}(\omega) = \frac{E_{\text{tot}}^2}{4\pi} \int_0^{2\pi} d\phi \int_0^\pi \sin \theta d\theta \delta(\omega - \theta) = \frac{\sin \omega}{2} E_{\text{tot}}^2, \quad (\text{B.10})$$

for a total energy E_{tot} . The cumulative spectral function of angular distances ω on the sphere is then

$$S_{\text{iso}}(\omega) = \frac{1 - \cos \omega}{2} E_{\text{tot}}^2. \quad (\text{B.11})$$

The inverse cumulative spectral function is then

$$S_{\text{iso}}^{-1}(E^2) = \cos^{-1} \left(1 - 2 \frac{E^2}{E_{\text{tot}}^2} \right). \quad (\text{B.12})$$

The integral of its square is

$$\int_0^{E_{\text{tot}}^2} dE^2 S_{\text{iso}}^{-1}(E^2)^2 = \frac{\pi^2 - 4}{2} E_{\text{tot}}^2. \quad (\text{B.13})$$

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

$$\begin{aligned} \int dE^2 S^{-1}(E^2) S_{\text{iso}}^{-1}(E^2) &= \sum_{\substack{n \in \mathcal{E}^2 \\ \omega_n < \omega_{n+1}}} \omega_n \int_{S^-(\omega_n)}^{S^+(\omega_n)} dE^2 \cos^{-1} \left(1 - 2 \frac{E^2}{E_{\text{tot}}^2} \right) \\ &= \sum_{\substack{n \in \mathcal{E}^2 \\ \omega_n < \omega_{n+1}}} \omega_n \left[\sqrt{S^+(\omega_n)} \sqrt{E_{\text{tot}}^2 - S^+(\omega_n)} + S^+(\omega_n) \cos^{-1} \left(1 - 2 \frac{S^+(\omega_n)}{E_{\text{tot}}^2} \right) \right. \\ &\quad \left. - E_{\text{tot}}^2 \sin^{-1} \left(\frac{\sqrt{S^+(\omega_n)}}{E_{\text{tot}}} \right) - \sqrt{S^-(\omega_n)} \sqrt{E_{\text{tot}}^2 - S^-(\omega_n)} \right. \\ &\quad \left. - S^-(\omega_n) \cos^{-1} \left(1 - 2 \frac{S^-(\omega_n)}{E_{\text{tot}}^2} \right) + E_{\text{tot}}^2 \sin^{-1} \left(\frac{\sqrt{S^-(\omega_n)}}{E_{\text{tot}}} \right) \right]. \end{aligned} \quad (\text{B.14})$$

These results can then be used to construct the expression for the observable of eq. (4.25).

B.4 Closed-form event spEquatorialness

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

$$S_{\text{ring}}(\omega) = \frac{\omega}{\pi} E_{\text{tot}}^2. \quad (\text{B.15})$$

The inverse cumulative spectral function is then

$$S_{\text{ring}}^{-1}(E^2) = \pi \frac{E^2}{E_{\text{tot}}^2}. \quad (\text{B.16})$$

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

$$\int_0^{E_{\text{tot}}^2} dE^2 S_{\text{ring}}^{-1}(E^2)^2 = \frac{\pi^2}{3} E_{\text{tot}}^2. \quad (\text{B.17})$$

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

$$\begin{aligned} \int dE^2 S^{-1}(E^2) S_{\text{ring}}^{-1}(E^2) &= \sum_{\substack{n \in \mathcal{E}^2 \\ \omega_n < \omega_{n+1}}} \omega_n \int_{S^-(\omega_n)}^{S^+(\omega_n)} dE^2 \pi \frac{E^2}{E_{\text{tot}}^2} \\ &= \frac{\pi}{2} \frac{1}{E_{\text{tot}}^2} \sum_{\substack{n \in \mathcal{E}^2 \\ \omega_n < \omega_{n+1}}} \omega_n \left[(S^+(\omega_n))^2 - (S^-(\omega_n))^2 \right]. \end{aligned} \quad (\text{B.18})$$

This can then be used to construct the observable of eq. (4.29).

Data Availability Statement. This article has associated data in a data repository.

Code Availability Statement. This article has associated code in a code repository.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

- [1] P.T. Komiske, E.M. Metodiev and J. Thaler, *Metric Space of Collider Events*, *Phys. Rev. Lett.* **123** (2019) 041801 [[arXiv:1902.02346](https://arxiv.org/abs/1902.02346)] [[INSPIRE](#)].
- [2] A. Mullin et al., *Does SUSY have friends? A new approach for LHC event analysis*, *JHEP* **02** (2021) 160 [[arXiv:1912.10625](https://arxiv.org/abs/1912.10625)] [[INSPIRE](#)].
- [3] A.J. Larkoski and T. Melia, *Covariantizing phase space*, *Phys. Rev. D* **102** (2020) 094014 [[arXiv:2008.06508](https://arxiv.org/abs/2008.06508)] [[INSPIRE](#)].
- [4] T. Cai, J. Cheng, N. Craig and K. Craig, *Linearized optimal transport for collider events*, *Phys. Rev. D* **102** (2020) 116019 [[arXiv:2008.08604](https://arxiv.org/abs/2008.08604)] [[INSPIRE](#)].
- [5] C. Cesarotti and J. Thaler, *A Robust Measure of Event Isotropy at Colliders*, *JHEP* **08** (2020) 084 [[arXiv:2004.06125](https://arxiv.org/abs/2004.06125)] [[INSPIRE](#)].
- [6] T. Cai, J. Cheng, K. Craig and N. Craig, *Which metric on the space of collider events?*, *Phys. Rev. D* **105** (2022) 076003 [[arXiv:2111.03670](https://arxiv.org/abs/2111.03670)] [[INSPIRE](#)].
- [7] M. Crispim Romão et al., *Use of a generalized energy Mover’s distance in the search for rare phenomena at colliders*, *Eur. Phys. J. C* **81** (2021) 192 [[arXiv:2004.09360](https://arxiv.org/abs/2004.09360)] [[INSPIRE](#)].
- [8] S. Tsan et al., *Particle Graph Autoencoders and Differentiable, Learned Energy Mover’s Distance*, in the proceedings of the *35th Conference on Neural Information Processing Systems*, Online Conference, Canada, December 06–14 (2021) [[arXiv:2111.12849](https://arxiv.org/abs/2111.12849)] [[INSPIRE](#)].
- [9] O. Kitouni, N. Nolte and M. Williams, *Finding NEEMo: Geometric Fitting using Neural Estimation of the Energy Mover’s Distance*, [arXiv:2209.15624](https://arxiv.org/abs/2209.15624) [[INSPIRE](#)].
- [10] R. Shenoy et al., *Differentiable Earth mover’s distance for data compression at the high-luminosity LHC*, *Mach. Learn. Sci. Tech.* **4** (2023) 045058 [[arXiv:2306.04712](https://arxiv.org/abs/2306.04712)] [[INSPIRE](#)].
- [11] A. Davis, T. Menzo, A. Youssef and J. Zupan, *Earth mover’s distance as a measure of CP violation*, *JHEP* **06** (2023) 098 [[arXiv:2301.13211](https://arxiv.org/abs/2301.13211)] [[INSPIRE](#)].
- [12] S. Alipour-Fard, P.T. Komiske, E.M. Metodiev and J. Thaler, *Pileup and Infrared Radiation Annihilation (PIRANHA): a paradigm for continuous jet grooming*, *JHEP* **09** (2023) 157 [[arXiv:2305.00989](https://arxiv.org/abs/2305.00989)] [[INSPIRE](#)].
- [13] T. Gaertner and J. Reiten, *Unsupervised learning in the metric space of jets*, [arXiv:2312.06948](https://arxiv.org/abs/2312.06948) [[INSPIRE](#)].
- [14] N. Craig, J.N. Howard and H. Li, *Exploring Optimal Transport for Event-Level Anomaly Detection at the Large Hadron Collider*, [arXiv:2401.15542](https://arxiv.org/abs/2401.15542) [[INSPIRE](#)].
- [15] T. Cai et al., *The phase space distance between collider events*, *JHEP* **09** (2024) 054 [[arXiv:2405.16698](https://arxiv.org/abs/2405.16698)] [[INSPIRE](#)].
- [16] D. Ba et al., *SHAPER: can you hear the shape of a jet?*, *JHEP* **06** (2023) 195 [[arXiv:2302.12266](https://arxiv.org/abs/2302.12266)] [[INSPIRE](#)].

- [17] C. Cesarotti and M. LeBlanc, *A Field Guide to Event-Shape Observables Using Optimal Transport*, [arXiv:2409.13150](#) [INSPIRE].
- [18] ATLAS collaboration, *Measurements of multijet event isotropies using optimal transport with the ATLAS detector*, *JHEP* **10** (2023) 060 [[arXiv:2305.16930](#)] [INSPIRE].
- [19] L.V. Kantorovich, *The mathematical method of production planning and organization*, *Manage. Sci.* **6** (1939) 363.
- [20] L.N. Wasserstein, *Markov processes over denumerable products of spaces describing large systems of automata*, *Probl. Info. Transm.* **5** (1969) 47.
- [21] R.L. Dobrushin, *Prescribing a system of random variables by conditional distributions*, *Theory of Probability & Its Applications* **15** (1970) 458.
- [22] R. Flamary et al., *Pot: Python optimal transport*, *J. Mach. Learn. Res.* **22** (2021) 1.
- [23] A.J. Larkoski and J. Thaler, *A spectral metric for collider geometry*, *JHEP* **08** (2023) 107 [[arXiv:2305.03751](#)] [INSPIRE].
- [24] C.L. Basham, L.S. Brown, S.D. Ellis and S.T. Love, *Energy Correlations in electron-Positron Annihilation: Testing QCD*, *Phys. Rev. Lett.* **41** (1978) 1585 [INSPIRE].
- [25] F.V. Tkachov, *Measuring multi-jet structure of hadronic energy flow or What is a jet?*, *Int. J. Mod. Phys. A* **12** (1997) 5411 [[hep-ph/9601308](#)] [INSPIRE].
- [26] M. Jankowiak and A.J. Larkoski, *Jet Substructure Without Trees*, *JHEP* **06** (2011) 057 [[arXiv:1104.1646](#)] [INSPIRE].
- [27] H. Chen, I. Moutl, X.Y. Zhang and H.X. Zhu, *Rethinking jets with energy correlators: Tracks, resummation, and analytic continuation*, *Phys. Rev. D* **102** (2020) 054012 [[arXiv:2004.11381](#)] [INSPIRE].
- [28] A. Chakraborty, S.H. Lim and M.M. Nojiri, *Interpretable deep learning for two-prong jet classification with jet spectra*, *JHEP* **07** (2019) 135 [[arXiv:1904.02092](#)] [INSPIRE].
- [29] M. Boutin and G. Kemper, *On reconstructing n -point configurations from the distribution of distances or areas*, *Adv. Appl. Math.* **32** (2004) 709.
- [30] G. Monge, *Mémoire sur la théorie des déblais et des remblais*, *Mem. Math. Phys. Acad. Royale Sci.* (1781) 666.
- [31] S. Peleg, M. Werman and H. Rom, *A unified approach to the change of resolution: space and gray-level*, *IEEE Trans. Pattern Anal. Machine Intell.* **11** (1989) 739.
- [32] Y. Rubner, C. Tomasi and L.J. Guibas, *A metric for distributions with applications to image databases*, in *Proceedings of the Sixth International Conference on Computer Vision, ICCV '98*, (U.S.A.), p. 59, IEEE Computer Society, (1998).
- [33] Y. Rubner, C. Tomasi and L.J. Guibas, *The Earth Mover's Distance as a Metric for Image Retrieval*, *Int. J. Comput. Vision* **40** (2000) 99 [INSPIRE].
- [34] O. Pele and M. Werman, *A linear time histogram metric for improved sift matching*, in *ECCV*, (2008), [[DOI:10.1007/978-3-540-88690-7_37](#)].
- [35] J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, [math/0412127](#).
- [36] J. Lott, *Some geometric calculations on Wasserstein space*, [math/0612562](#).
- [37] C. Villani, *Optimal Transport: Old and New*, Grundlehren der mathematischen Wissenschaften, Springer Berlin Heidelberg (2008).

- [38] P.T. Komiske, E.M. Metodiev and J. Thaler, *The Hidden Geometry of Particle Collisions*, *JHEP* **07** (2020) 006 [[arXiv:2004.04159](#)] [[INSPIRE](#)].
- [39] H. Solomon, *Geometric probability*, SIAM (1978).
- [40] D.A. Klain and G.-C. Rota, *Introduction to geometric probability*, Cambridge University Press (1997).
- [41] L.A. Santaló, *Integral geometry and geometric probability*, Cambridge University Press (2004).
- [42] P.T. Komiske, S. Kryhin and J. Thaler, *Disentangling quarks and gluons in CMS open data*, *Phys. Rev. D* **106** (2022) 094021 [[arXiv:2205.04459](#)] [[INSPIRE](#)].
- [43] A. Banfi, G.P. Salam and G. Zanderighi, *Principles of general final-state resummation and automated implementation*, *JHEP* **03** (2005) 073 [[hep-ph/0407286](#)] [[INSPIRE](#)].
- [44] A.J. Larkoski, G.P. Salam and J. Thaler, *Energy Correlation Functions for Jet Substructure*, *JHEP* **06** (2013) 108 [[arXiv:1305.0007](#)] [[INSPIRE](#)].
- [45] A.F. Agarap, *Deep Learning using Rectified Linear Units (ReLU)*, [arXiv:1803.08375](#) [[INSPIRE](#)].
- [46] N. Bonneel, J. Rabin, G. Peyré and H. Pfister, *Sliced and Radon Wasserstein Barycenters of Measures*, *J. Math. Imaging Vision* **51** (2015) 22.
- [47] M. Cuturi, *Sinkhorn distances: Lightspeed computation of optimal transport*, *Adv. Neural Inf. Process. Syst.* **26** (2013).
- [48] C.F. Berger, T. Kucs and G.F. Sterman, *Event shape / energy flow correlations*, *Phys. Rev. D* **68** (2003) 014012 [[hep-ph/0303051](#)] [[INSPIRE](#)].
- [49] C.F. Berger and L. Magnea, *Scaling of power corrections for angularities from dressed gluon exponentiation*, *Phys. Rev. D* **70** (2004) 094010 [[hep-ph/0407024](#)] [[INSPIRE](#)].
- [50] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N-subjettiness*, *JHEP* **03** (2011) 015 [[arXiv:1011.2268](#)] [[INSPIRE](#)].
- [51] J. Bradbury et al., *JAX: composable transformations of Python+NumPy programs*, <http://github.com/google/jax> (2018).
- [52] J. Pivarski et al., *Awkward Array*, (2018) DOI:[10.5281/zenodo.4341376](#).
- [53] R. Sinkhorn, *A relationship between arbitrary positive matrices and stochastic matrices*, *Can. J. Math.* **18** (1966) 303.
- [54] C. Clason, D.A. Lorenz, H. Mahler and B. Wirth, *Entropic regularization of continuous optimal transport problems*, [arXiv:1906.01333](#) [DOI:[10.1016/j.jmaa.2020.124432](#)].
- [55] J. Feydy et al., *Interpolating between Optimal Transport and MMD using Sinkhorn Divergences*, in *The 22nd International Conference on Artificial Intelligence and Statistics*, 2019, pp. 2681–2690, [[arXiv:1810.08278](#)].
- [56] H. Janati, M. Cuturi and A. Gramfort, *Debiased Sinkhorn barycenters*, in *International Conference on Machine Learning*, PMLR (2020), pp. 4692–4701 [[arXiv:2006.02575](#)].
- [57] I.W. Stewart, F.J. Tackmann and W.J. Waalewijn, *N-Jettiness: An Inclusive Event Shape to Veto Jets*, *Phys. Rev. Lett.* **105** (2010) 092002 [[arXiv:1004.2489](#)] [[INSPIRE](#)].
- [58] I.W. Stewart et al., *XCone: N-jettiness as an Exclusive Cone Jet Algorithm*, *JHEP* **11** (2015) 072 [[arXiv:1508.01516](#)] [[INSPIRE](#)].
- [59] S. Catani, Y.L. Dokshitzer, M.H. Seymour and B.R. Webber, *Longitudinally invariant K_t clustering algorithms for hadron hadron collisions*, *Nucl. Phys. B* **406** (1993) 187 [[INSPIRE](#)].

- [60] S.D. Ellis and D.E. Soper, *Successive combination jet algorithm for hadron collisions*, *Phys. Rev. D* **48** (1993) 3160 [[hep-ph/9305266](#)] [[INSPIRE](#)].
- [61] P. Tankala, A. Tasissa, J.M. Murphy and D. Ba, *K-Deep Simplex: Deep Manifold Learning via Local Dictionaries*, [arXiv:2012.02134](#).
- [62] M. Mueller, S. Aeron, J.M. Murphy and A. Tasissa, *Geometric Sparse Coding in Wasserstein Space*, [arXiv:2210.12135](#).
- [63] W. Wang and M. Carreira-Perpiñán, *Projection onto the probability simplex: An efficient algorithm with a simple proof, and an application*, [arXiv:1309.1541](#).
- [64] E. Farhi, *A QCD Test for Jets*, *Phys. Rev. Lett.* **39** (1977) 1587 [[INSPIRE](#)].
- [65] S. Brandt, C. Peyrou, R. Sosnowski and A. Wroblewski, *The principal axis of jets. An attempt to analyze high-energy collisions as two-body processes*, *Phys. Lett.* **12** (1964) 57 [[INSPIRE](#)].
- [66] A. De Rujula, J.R. Ellis, E.G. Floratos and M.K. Gaillard, *QCD Predictions for Hadronic Final States in e^+e^- Annihilation*, *Nucl. Phys. B* **138** (1978) 387 [[INSPIRE](#)].
- [67] A.Y. Wei, P. Naik, A.W. Harrow and J. Thaler, *Quantum Algorithms for Jet Clustering*, *Phys. Rev. D* **101** (2020) 094015 [[arXiv:1908.08949](#)] [[INSPIRE](#)].
- [68] C. Bierlich et al., *A comprehensive guide to the physics and usage of PYTHIA 8.3*, *SciPost Phys. Codeb.* **2022** (2022) 8 [[arXiv:2203.11601](#)] [[INSPIRE](#)].
- [69] M. Cacciari, G.P. Salam and G. Soyez, *FastJet User Manual*, *Eur. Phys. J. C* **72** (2012) 1896 [[arXiv:1111.6097](#)] [[INSPIRE](#)].
- [70] M. Cacciari, G.P. Salam and G. Soyez, *The anti- k_t jet clustering algorithm*, *JHEP* **04** (2008) 063 [[arXiv:0802.1189](#)] [[INSPIRE](#)].
- [71] R. Kleiss, W.J. Stirling and S.D. Ellis, *A New Monte Carlo Treatment of Multiparticle Phase Space at High-energies*, *Comput. Phys. Commun.* **40** (1986) 359 [[INSPIRE](#)].
- [72] O. Pele and B. Taskar, *The tangent earth mover's distance*, in *Geometric Science of Information*, F. Nielsen and F. Barbaresco, eds., Springer Berlin Heidelberg (2013), pp. 397–404.
- [73] J. Batson and Y. Kahn, *Scaling Laws in Jet Classification*, [arXiv:2312.02264](#) [[INSPIRE](#)].