MIT Open Access Articles

Jet observables without jet algorithms

Massachusetts Institute of Technology

Jet observables without jet algorithms

Daniele Bertolini, Tucker Chan[†] **and Jesse Thaler**

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

E-mail: danbert@mit.edu, jthaler@mit.edu

ABSTRACT: We introduce a new class of event shapes to characterize the jet-like structure of an event. Like traditional event shapes, our observables are infrared/collinear safe and involve a sum over all hadrons in an event, but like a jet clustering algorithm, they incorporate a jet radius parameter and a transverse momentum cut. Three of the ubiquitous jet-based observables — jet multiplicity, summed scalar transverse momentum, and missing transverse momentum — have event shape counterparts that are closely correlated with their jet-based cousins. Due to their "local" computational structure, these jet-like event shapes could potentially be used for trigger-level event selection at the LHC. Intriguingly, the jet multiplicity event shape typically takes on non-integer values, highlighting the inherent ambiguity in defining jets. By inverting jet multiplicity, we show how to characterize the transverse momentum of the n-th hardest jet without actually finding the constituents of that jet. Since many physics applications do require knowledge about the jet constituents, we also build a hybrid event shape that incorporates (local) jet clustering information. As a straightforward application of our general technique, we derive an event-shape version of jet trimming, allowing event-wide jet grooming without explicit jet identification. Finally, we briefly mention possible applications of our method for jet substructure studies.

[†]Deceased.

Contents

1 Introduction

When quarks and gluons are produced in high energy particle collisions, they undergo a process of showering and hadronization, and the resulting final state can be organized in terms of clusters of hadrons called jets. Jets play a key role at experiments like the Large Hadron Collider (LHC), both for testing standard model (SM) physics and for searching for new phenomena beyond the SM. At present, most jet studies at the LHC are based on jets identified with a jet algorithm [1, 2]. Algorithms such as anti-k_T [3] cluster final state hadrons into jet objects, whose four-momenta are then used as inputs for subsequent analyses. An alternative approach is provided by event shape observables, which are functions involving all final state hadrons in a collision event. Event shapes were extensively used for precision tests of quantum chromodynamics (QCD) at e^+e^- colliders [4–8], and various event shapes have been proposed and used at hadron colliders [9–13].

In this paper, we will blur the distinction between jet algorithms and event shapes by constructing jet-like event shapes. These event shapes incorporate a jet-like radius R as well as a jet-like transverse momentum cut $p_{T\text{cut}}$, and they can be viewed as counterparts to some of the most commonly used jet-based observables. While these event shapes do not involve any kind of clustering procedure, they are correlated with their jet-based cousins and yield comparable information about the jet-like structure of an event. In this paper,

we will mainly discuss jet-like event shapes, but the generalization to subjet-like jet shapes is straightforward, with potential applications in jet substructure studies [14, 15].

We will start by constructing three jet-like event shapes that mirror the three inclusive jet observables — jet multiplicity, summed scalar transverse momentum, and missing transverse momentum — that appear ubiquitously in jet studies at both the trigger and analysis levels. For example, we will construct the jet multiplicity event shape as

$$\widetilde{N}_{\text{jet}}(p_{T\text{cut}}, R) = \sum_{i \in \text{event}} \frac{p_{Ti}}{p_{Ti,R}} \Theta(p_{Ti,R} - p_{T\text{cut}}), \tag{1.1}$$

where $p_{Ti,R}$ is the transverse momentum contained in a cone of radius R around particle i. Our technique for building jet-like event shapes can be generalized to a broad class of inclusive jet observables, namely observables built as a sum over all jets in an event.

We will then show how to manipulate these event shapes to characterize individual jets. By inverting $\widetilde{N}_{\text{jet}}$, we can characterize the p_T of the n-th hardest jet without explicitly identifying the set of hadrons that form that jet. Of course, for practical jet studies, one often wants to know the actual constituents of a jet. Since our jet-like event shapes do not have a natural clustering interpretation, we develop a hybrid method that incorporates local jet clustering into an "event shape density". The integral over this density gives the corresponding event shape, but the density distribution itself has spikes in the direction of candidate jet axes.

A perhaps surprising application of our method is for jet grooming [16–19]. Jet grooming methods aim to mitigate the effects of jet contamination from initial state radiation, underlying event, and pileup by removing soft wide-angle radiation from a jet. In the case of pileup, one can use jet grooming in concert with area subtraction techniques [20–22]. Here, we show how jet trimming [19] can be recast as an event shape. Our method is equivalent to assigning a weight to every particle in the event of

$$w_i = \Theta\left(\frac{p_{Ti,R_{\text{sub}}}}{p_{Ti,R}} - f_{\text{cut}}\right) \Theta(p_{Ti,R} - p_{T\text{cut}}). \tag{1.2}$$

This "shape trimming" method involves the same f_{cut} and R_{sub} parameters as the original "tree trimming" procedure, but does not require the explicit identification of jets or subjets.

There are a number of potential applications for these jet-like event shapes. At the trigger level, they offer a "local" way to characterize the gross properties of an event. By local, we mean that the event shape is defined as a sum over regions of interest of radius R, without needing global clustering information. This local structure allows for efficient parallel computation of the event shape.[1] If desired, one could even include (local) pileup suppression at the trigger level by incorporating (local) trimming. At the analysis level, these event shapes offer an alternative way to characterize jets in regions of phase space where jets are overlapping. In particular, whereas standard jet algorithms always give an integer value for the jet multiplicity N_{jet}, the corresponding event shape $\widetilde{N}_{\text{jet}}$ in eq. (1.1) typically returns a non-integer value, reflecting the inherent ambiguity in defining jets. At minimum, one can use these event shapes to test the robustness of standard jet selection

---

[1]We thank David Strom for pointing out this possibility to us.

criteria, since a cut on the jet-like event shape should give similar results to a cut on jet objects for the same value of R and $p_{T\text{cut}}$. Ultimately, one would like to study the analytic properties of these jet-like event shapes in perturbative QCD, though such studies are beyond the scope of this paper.

It is worth noting that our approach shares some of the same goals and features as other jet-like methods. For defining jet observables through event shapes, there has been previous work showing how to construct effective jet clustering procedures via optimization of event shapes [23], most recently in taking N-jettiness [24] and minimizing over the choice of jet axes [25]. The difference here is that the jet-like event shapes do not have an obvious clustering interpretation. There are also methods that cast jet finding as a more general optimization problem [26–35], often with a probabilistic interpretation of an event. The difference here is that we (uniquely) assign an event shape value to each event. A set of variables that avoids explicit jet clustering are energy correlation functions [36], which can characterize an event's structure without reference to even a jet axis (in contrast to N-jettiness), though different correlation functions are needed for different jet multiplicities. The difference here is that we need not specify the jet multiplicity of interest, though we do need to choose the jet radius R and threshold $p_{T\text{cut}}$. Finally, for giving a global characterization of an event, there has been recent work to describe the jet-like nature of an event by summing over the contributions of large radius jets [37–39], though these observables make explicit use of tree-like recursive jet algorithms. The difference here is that we can achieve a similar global characterization through an inclusive sum over all particles in an event.

The rest of this paper is organized as follows. We define event shapes for inclusive jet observables in section 2 and perform Monte Carlo studies to demonstrate the correlations present with their jet-based cousins. We then show in section 3 how to manipulate and modify these event shapes to characterize the properties of individual jets, in particular how to find the jet constituents using a hybrid event shape density with a "winner-take-all" recombination scheme. We describe our shape trimming technique in section 4 and show how it is closely correlated with ordinary tree trimming. We suggest possible generalization of our method in section 5 and draw conclusions in section 6. All of the event shapes described in this paper are available as an add-on to FASTJET 3 [42] as part of the FASTJET contrib project (http://fastjet.hepforge.org/contrib/).

2 Event shapes for inclusive jet observables

Jet multiplicity (N_{jet}), summed scalar transverse momentum (H_T), and missing transverse momentum ($\not{p}_T$) are three of the most ubiquitous observables used to globally characterize an event with jets in the final state. Given jets identified through some jet algorithm with

characteristic radius R, they are defined as

$$N_{\mathrm{jet}}(p_{T\mathrm{cut}}, R) = \sum_{\mathrm{jets}} \Theta(p_{T\mathrm{jet}} - p_{T\mathrm{cut}}), \tag{2.1}$$

$$H_T(p_{T\mathrm{cut}}, R) = \sum_{\mathrm{jets}} p_{T\mathrm{jet}} \, \Theta(p_{T\mathrm{jet}} - p_{T\mathrm{cut}}), \tag{2.2}$$

$$\not{p}_T(p_{T\mathrm{cut}}, R) = \left| \sum_{\mathrm{jets}} \vec{p}_{T\mathrm{jet}} \, \Theta(p_{T\mathrm{jet}} - p_{T\mathrm{cut}}) \right|, \tag{2.3}$$

where $\vec{p}_{T\mathrm{jet}}$ is the transverse momentum measured with respect to the beam axis, $p_{T\mathrm{jet}} = |\vec{p}_{T\mathrm{jet}}|$, and $p_{T\mathrm{cut}}$ is the p_T threshold for the analysis.[2] We have made the arguments $p_{T\mathrm{cut}}$ and R explicit in anticipation of the discussion in section 3. These three observables are part of a broader class of inclusive jet observables

$$\mathcal{F}(p_{T\mathrm{cut}}, R) = \sum_{\mathrm{jets}} \mathcal{F}_{\mathrm{jet}} \, \Theta(p_{T\mathrm{jet}} - p_{T\mathrm{cut}}), \tag{2.4}$$

where $\mathcal{F}_{\mathrm{jet}} = f(\{p_j^\mu\}_{j \in \mathrm{jet}})$ depends on the kinematics of the individual jet constituents.

As written, $\mathcal{F}$ is intrinsically tied to a given jet algorithm. Here, we wish to build a corresponding event shape $\widetilde{\mathcal{F}}$ which makes no reference to a clustering procedure. The first step is to effectively replace the sum over jets with a sum over particles, using the fact that

$$1 = \frac{1}{p_{T\mathrm{jet}}} \sum_{i \in \mathrm{jet}} p_{Ti}, \qquad \sum_{\mathrm{jets}} \sum_{i \in \mathrm{jet}} \Rightarrow \sum_{i \in \mathrm{event}}, \tag{2.5}$$

where we now use a more convenient definition $p_{T\mathrm{jet}} \equiv \sum_{i \in \mathrm{jet}} p_{Ti}$ such that the first expression is a strict equality,[3] and the second expression has an implicit restriction to particles i which are part of a jet cluster. The second step is to convert jet measurements into measurements on jet-like cones of radius R around each particle:

$$\mathcal{F}_{\mathrm{jet}} = f(\{p_j^\mu\}_{j \in \mathrm{jet}}) \quad \Rightarrow \quad \mathcal{F}_{i,R} = f(\{p_j^\mu \, \Theta(R - \Delta R_{ij})\}_{j \in \mathrm{event}}), \tag{2.6}$$

$$p_{T\mathrm{jet}} = \sum_{i \in \mathrm{jet}} p_{Ti} \quad \Rightarrow \quad p_{Ti,R} = \sum_{j \in \mathrm{event}} p_{Tj} \, \Theta(R - \Delta R_{ij}), \tag{2.7}$$

where $\Delta R_{ij} = \sqrt{\Delta \eta_{ij}^2 + \Delta \phi_{ij}^2}$ is the distance in the rapidity-azimuth plane, and $p_{Ti,R}$ is the sum of transverse momentum contained in a cone of radius R around particle i. Applying these two steps, we derive the event shape associated with the generic inclusive jet observable in eq. (2.4):

$$\widetilde{\mathcal{F}}(p_{T\mathrm{cut}}, R) = \sum_{i \in \mathrm{event}} \frac{p_{Ti}}{p_{Ti,R}} \mathcal{F}_{i,R} \, \Theta(p_{Ti,R} - p_{T\mathrm{cut}}). \tag{2.8}$$

Because of the weight factor $p_{Ti}/p_{Ti,R}$, this definition avoids double-counting, even though the jet-like cones around each particle are overlapping. As long as the original $\mathcal{F}_{\mathrm{jet}}$ was infrared/collinear safe, then $\widetilde{\mathcal{F}}$ will also be infrared/collinear safe (assuming $p_{T\mathrm{cut}} > 0$). Our general strategy is depicted in figure 1.

---

[2]Typically, $\not{p}_T$ would include non-hadronic objects in the event as well, but we will not need that for the case studies in this paper.

[3]Note that the two definitions $p_{T\mathrm{jet}} \equiv |\vec{p}_{T\mathrm{jet}}|$ vs. $\sum_{i \in \mathrm{jet}} p_{Ti}$ yield the same value for infinitely narrow jets. Instead of p_T, one could accomplish the same goal using the energy relation $1 = (1/E_{\mathrm{jet}}) \sum_{i \in \mathrm{jet}} E_i$.
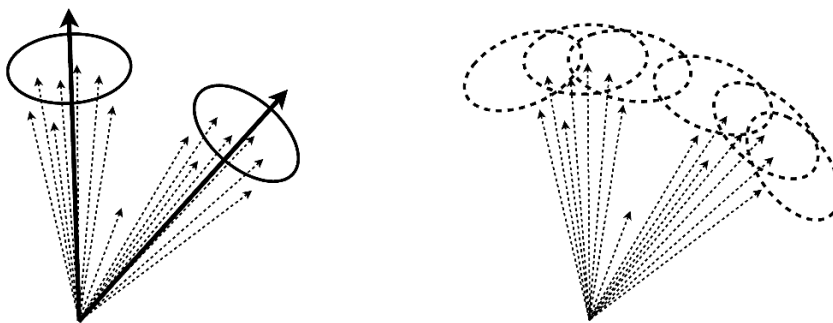
Figure 1: Instead of defining inclusive jet observables by summing over jet regions according to a jet algorithm (left), our event shapes sum over the contributions from cones of radius R centered on each particle i (right). The weight factor $p_{Ti}/p_{Ti,R}$ in eq. (2.8) avoids double-counting despite overlapping cones. For infinitely narrow jets separated by more than R, the two methods yield the same result.

Following this logic, we define the following jet-like event shapes corresponding to $N_{\rm jet}$, H_T, and $\not{p}_T$:

$$\widetilde{N}_{\rm jet}(p_{T\rm cut}, R) = \sum_{i \in \rm event} \frac{p_{Ti}}{p_{Ti,R}} \Theta(p_{Ti,R} - p_{T\rm cut}),\tag{2.9}$$

$$\widetilde{H}_T(p_{T\rm cut}, R) = \sum_{i \in \rm event} p_{Ti}\, \Theta(p_{Ti,R} - p_{T\rm cut}),\tag{2.10}$$

$$\widetilde{\not{p}}_T(p_{T\rm cut}, R) = \left| \sum_{i \in \rm event} \vec{p}_{Ti}\, \Theta(p_{Ti,R} - p_{T\rm cut}) \right|,\tag{2.11}$$

where $p_{Ti,R}$ is defined in eq. (2.7). For the sake of simplicity, in eq. (2.11) we approximated $\vec{p}_{Ti,R} \approx p_{Ti,R}\, \hat{p}_{Ti}$, which is strictly true only for infinitely narrow jets.[4] For events consisting of infinitely narrow jets separated by more than R, the event shapes $\widetilde{N}_{\rm jet}$, $\widetilde{H}_T$, and $\widetilde{\not{p}}_T$ yield identical values to their jet-based counterparts $N_{\rm jet}$, H_T, and $\not{p}_T$. We describe applications and generalizations of this procedure to other inclusive jet (and subjet) observables in section 5.

To get a sense for how these event shapes behave, it is useful to study how they correlate with their jet-based counterparts. For this study, we generate event samples for the $\sqrt{s} = 8\,{\rm TeV}$ LHC in MadGraph 5 [40], with showering and hadronization carried out in Pythia 8.157 [41].[5] For the standard jet-based observables, we use FastJet 3.0.2 with the anti-k_T jet algorithm [3] with a jet radius $R = 0.6$ and $p_{T\rm cut} = 25$ GeV. For the event shapes, we use the same value of R and $p_{T\rm cut}$. In order to (artificially) highlight the

---

[4]Alternatively, one could recover eq. (2.11) by noticing that if we assume $\mathcal{F}_{\rm jet} \equiv \vec{p}_{T\rm jet} \simeq \sum_{j \in \rm jet} \vec{p}_{Tj}$, then we can skip the first replacement in eq. (2.5), and directly convert the double sum into a sum over the event.

[5]Unless otherwise specified, this will be the standard setup for Monte Carlo studies throughout the paper.
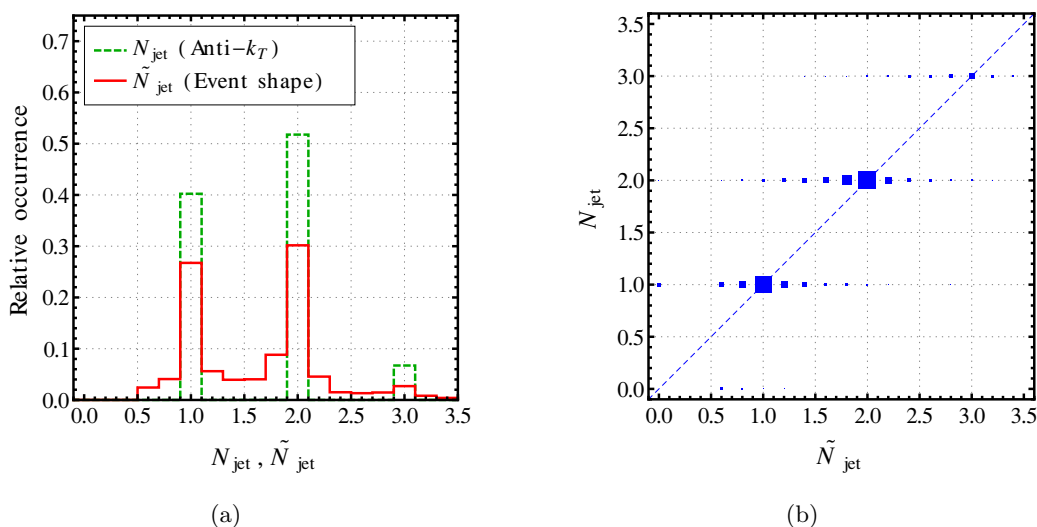
Figure 2: Jet multiplicity (i.e. N_{jet}) for QCD dijet events. figure 2a shows the distribution of the number of anti-k_T jets with $R = 0.6$ and $p_{T\text{cut}} = 25\,\text{GeV}$ (green dashed curve), and of the corresponding event shape with the same values of R and $p_{T\text{cut}}$ (red curve). Only events with $N_{\text{jet}} \geq 1$ or $\widetilde{N}_{\text{jet}} \geq 0.5$ are shown, and a parton level cut of $p_{T\text{cut}}^{\text{parton}} = 25$ GeV is employed to give a reasonable sample of both one jet and two jet events. Whereas N_{jet} takes on only integer values, the event shape $\widetilde{N}_{\text{jet}}$ is continuous, albeit with spikes near integer values. figure 2b shows the correlation between the two observables, where the area of the squares is proportional to the fraction of events in each bin. In the correlation plot, events that fail one of the jet cut criteria are assigned the corresponding value of zero.

behavior of our event shapes on both one jet and two jet events, we set the minimum p_T at the parton level in MADGRAPH to $p_{T\text{cut}}^{\text{parton}} = 25$ GeV.[6]

In figure 2, we compare N_{jet} versus $\widetilde{N}_{\text{jet}}$ for QCD dijet events. Whereas N_{jet} takes on discrete values, $\widetilde{N}_{\text{jet}}$ yields a continuous distribution, though the observables are correlated on an event-by-event basis. Here and in the following plots we only show events with $N_{\text{jet}} \geq 1$ and $\widetilde{N}_{\text{jet}} \geq 0.5$; the choice of the lower limit on $\widetilde{N}_{\text{jet}}$ will be justified in section 3.1. In figure 3, we compare H_T versus $\widetilde{H}_T$ again for QCD dijet events. Because of the $p_{T\text{cut}} = 25$ GeV cut, H_T exhibits two spikes that rise starting at 25 GeV (for one jet events) and 50 GeV (for two jet events), whereas $\widetilde{H}_T$ is smoother in this turn-on region.[7] In the tail region, the distributions of H_T and $\widetilde{H}_T$ are very similar. In figure 4, we compare $\not{p}_T$ versus $\widetilde{\not{p}}_T$ for Z plus jet events where the Z decays to neutrinos. Again we see a spike that rises starting at 25 GeV for $\not{p}_T$ which is milder in the event shape $\widetilde{\not{p}}_T$, though the distributions are quite similar throughout.

---

[6]Without a $p_{T\text{cut}}^{\text{parton}}$ cut, there would of course be more one jet than two jet events. We checked that the event shape distributions remain correlated with their jet-based counterparts as $p_{T\text{cut}}^{\text{parton}} \to 0$.

[7]With $p_{T\text{cut}}^{\text{parton}} \to 0$, the same features are visible, albeit with the one jet spike being much larger than the two jet spike.
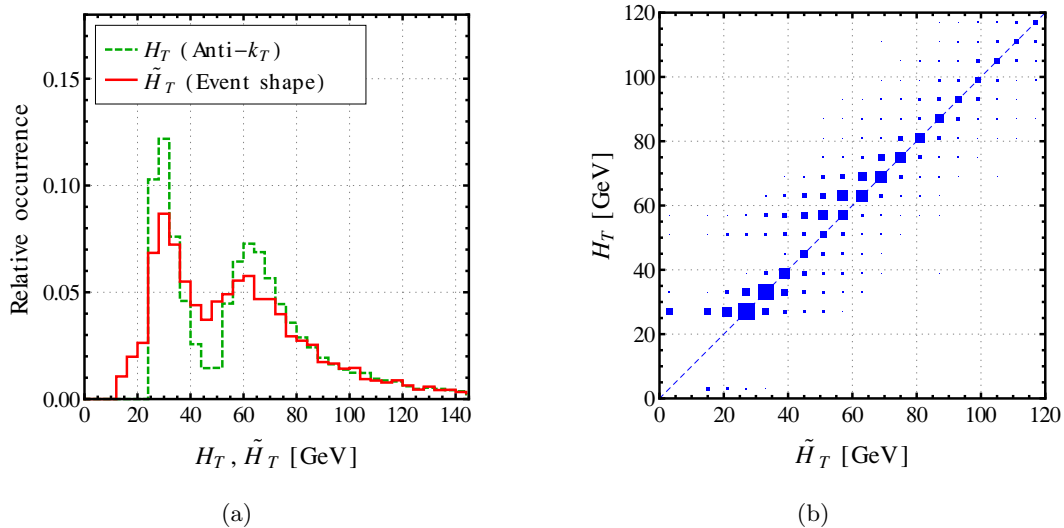
(a)

(b)

Figure 3: Summed scalar transverse momentum (i.e. H_T) for QCD dijet events. The jet parameters, formatting, and cuts are the same as for figure 2. Because of the smoother behavior of the event shape $\widetilde{H}_T$, the peaks rising at $p_{T\mathrm{cut}}$ and $2\,p_{T\mathrm{cut}}$ are less pronounced than for H_T.
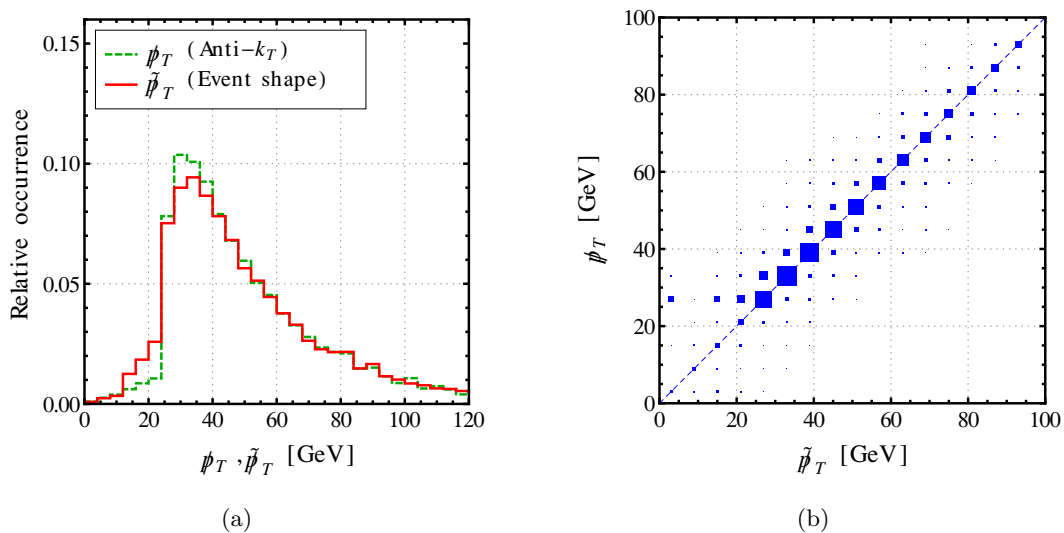

(a)

(b)

Figure 4: Missing transverse momentum (i.e. $\not{p}_T$) for $Z(\to \nu\bar{\nu}) + j$ events. The jet parameters, formatting, and cuts are the same as for figure 2. Again, we see a smoother turn on behavior for $\widetilde{\not{p}}_T$ compared to $\not{p}_T$.
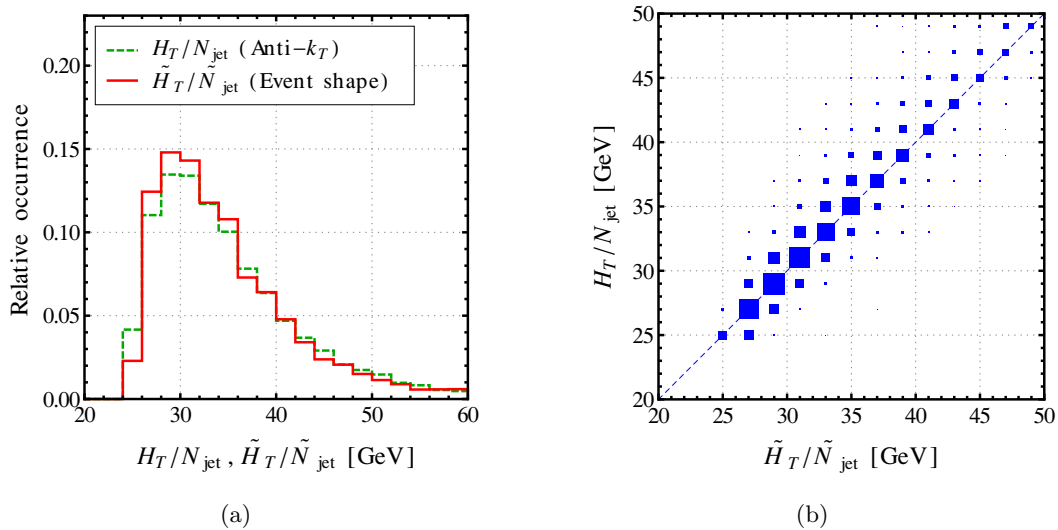
(a)

(b)

Figure 5: Average jet transverse momentum (i.e. H_T divided by N_{jet}) for QCD dijet events. The jet parameters, formatting, and cuts are the same as for figure 2.
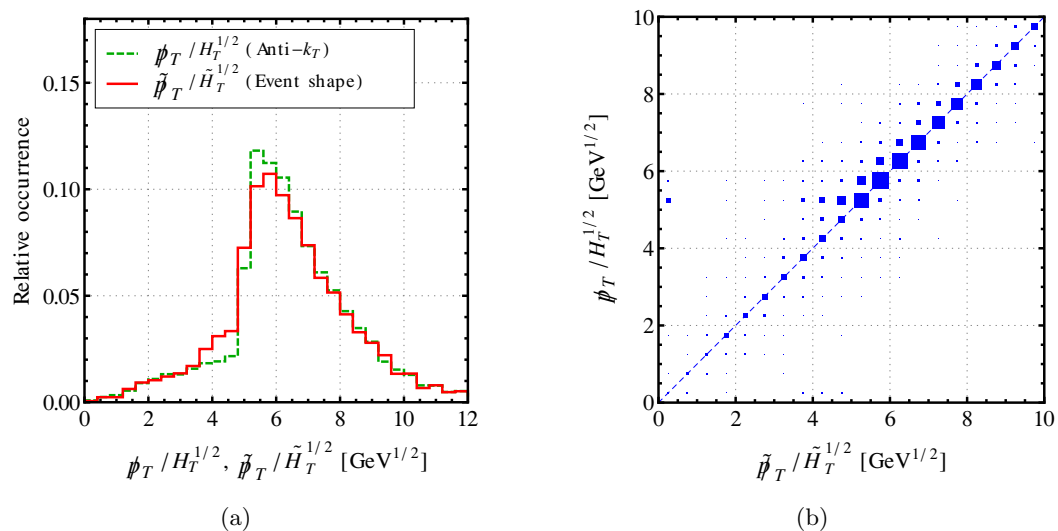

(a)

(b)

Figure 6: Missing transverse momentum significance (i.e. $\not{p}_T$ divided by $\sqrt{H_T}$) for $Z(\to \nu\bar{\nu}) + j$ events. The jet parameters, formatting, and cuts are the same as for figure 2.

Just as for ordinary jet-based observables, one can construct interesting composite functions with the event shapes. For example, one can consider the average p_T of the jets in an event, and we compare H_T/N_{jet} versus $\widetilde{H}_T/\widetilde{N}_{\text{jet}}$ in figure 5. Another useful composite variable is missing p_T significance [43, 44], and we compare $p\!\!\!/_T/\sqrt{H_T}$ versus $\widetilde{p\!\!\!/}_T/\sqrt{\widetilde{H}_T}$ in figure 6.

The differences between the jet-like event shapes and their jet-based counterparts reflects the intrinsic ambiguity in how to define a jet, seen most strikingly in the fact that $\widetilde{N}_{\text{jet}}$ does not take on integer values. For jet observables that are inclusive over all jets, $\widetilde{N}_{\text{jet}}$, $\widetilde{H}_T$, and $\widetilde{p\!\!\!/}_T$ characterize the global properties of the event without defining a clustering procedure, and appear to give very similar information to N_{jet}, H_T, and $p\!\!\!/_T$ for the same values of R and $p_{T\text{cut}}$. Of course, because there is no clustering, one cannot determine the kinematics of any individual jet with the event shape alone (see however section 3 below). In terms of computational costs, the bottleneck is calculating $p_{Ti,R}$ in eq. (2.7) for every particle i, which naively scales like N^2 for an event with N hadrons, though the computational costs are dramatically reduced if one has an efficient way to determine which particles are within a radius R of particle i.[8] In practice, calculating $\widetilde{N}_{\text{jet}}$ using our FASTJET 3 add-on with a standard laptop takes about as long as calculating N_{jet} with anti-k_T. Moreover, $\widetilde{N}_{\text{jet}}$ can be parallelized since it only depends on the contributions from particles within a radius R (i.e. it is defined "locally"). This feature makes it possible to implement $\widetilde{N}_{\text{jet}}$ in a low-level trigger for sufficiently small R. The key question at the trigger level is whether an event-shape-based trigger has better properties (e.g. turn-on, stability, calibration, etc.) than a jet-based trigger, but a detailed study of this issue is beyond the scope of this work.

3 Characterizing individual jets

While inclusive jet observables are useful for characterizing the gross properties of an event, one would still like to gain more exclusive information about the kinematics of individual jets. In general, our jet-like event shapes do not yield that kind of exclusive information, but we will demonstrate a novel way to extract the (approximate) transverse momentum of individual jets by using the full functional form of $\widetilde{N}_{\text{jet}}$. We will then define a hybrid event shape density that incorporates (local) jet clustering information in order to determine the constituents of individual jets.

3.1 Jet transverse momentum

Consider the jet multiplicity event shape $\widetilde{N}_{\text{jet}}(p_{T\text{cut}}, R)$. As shown in appendix A, there is a computationally efficient way to find the pseudo-inverse of this function with respect

---

[8]In our FASTJET add-on, we make a crude attempt in this direction by partitioning the event into overlapping blocks of size $2R \times 2R$ and by caching the results of repeated calculations. Our implementation could potentially be further optimized by using, for example, an alternative distance heuristic.
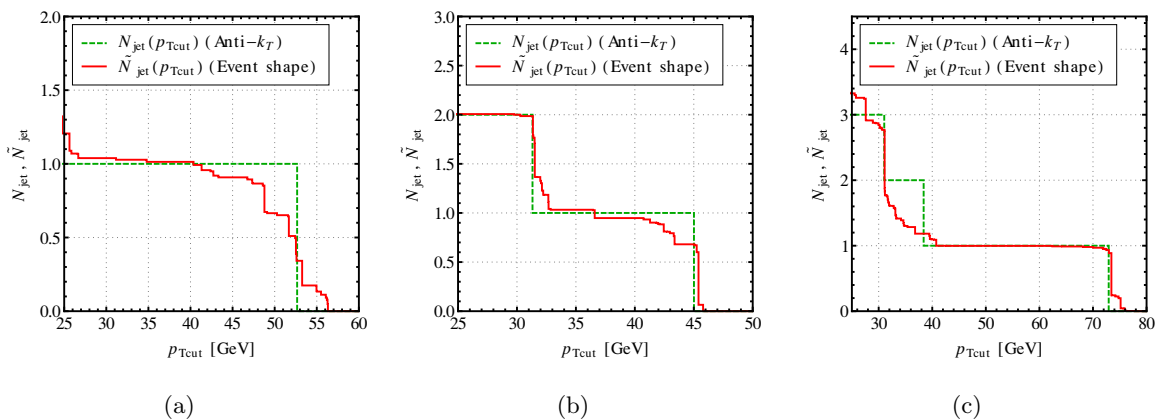
(a)   (b)   (c)

Figure 7: Number of jets N_{jet} as a function of $p_{T\text{cut}}$ for fixed $R = 0.6$, for three QCD dijet events. figures 7a, 7b, and 7c show example events with 1, 2, and 3 anti-k_T jets with $p_T > 25\,\text{GeV}$, respectively. The anti-k_T curve (green dashed line) takes integer steps at values of $p_{T\text{cut}}$ corresponding to the p_T of the jets. The event shape curve (red line) takes smaller steps, and it roughly intersects the anti-k_T curve at $\widetilde{N}_{\text{jet}} = \{0.5, 1.5, 2.5\}$.

to $p_{T\text{cut}}$, namely $p_{T\text{cut}}(\widetilde{N}_{\text{jet}}, R)$.[9] We will see in a moment that it is useful to introduce an offset n_{off}, so we define

$$\widetilde{p}_T(n, R) = p_{T\text{cut}}(n - n_{\text{off}}, R) \quad \text{with} \quad 0 \lesssim n_{\text{off}} \lesssim 1, \tag{3.1}$$

where the default value of n_{off} is 0.5. The corresponding function for ordinary jets is denoted $p_T(n, R)$.

The function $\widetilde{p}_T(n, R)$ effectively gives the p_T of the n-th hardest jet. That is, it gives the value of the p_T threshold needed to include the n-th jet's contribution to $\widetilde{N}_{\text{jet}}$. For infinitely narrow jets separated by more than R, $p_{T\text{cut}}(\widetilde{N}_{\text{jet}}, R)$ takes discrete jumps as $\widetilde{N}_{\text{jet}}$ increases by integer values. More generally, the offset n_{off} accounts for the fact that an event with n jets most likely returns a value of $\widetilde{N}_{\text{jet}}$ between $n - 1$ and n.

Using the same QCD dijet event samples as in section 2, we can see how well $\widetilde{p}_T(n, R)$ corresponds to $p_T(n, R)$. First in figure 7, we show the function $\widetilde{N}_{\text{jet}}(p_{T\text{cut}}, R)$ for individual events compared to $N_{\text{jet}}(p_{T\text{cut}}, R)$, fixing $R = 0.6$. Besides the obvious point that N_{jet} takes integer steps whereas $\widetilde{N}_{\text{jet}}$ takes smaller steps, we see that the curves roughly intersect at values of $\widetilde{N}_{\text{jet}} = 0.5, 1.5, 2.5$, justifying the default value $n_{\text{off}} = 0.5$. In figure 8, we compare $\widetilde{p}_T(n, R)$ versus $p_T(n, R)$ for $n = 1, 2, 3$, where we see that they are highly correlated, as expected from the correlations already seen in the inclusive observables in section 2.

Besides just measuring the p_T of the n-th hardest jet, $\widetilde{p}_T(n, R)$ can be used to mimic analyses that require a fixed number of jets. For example, one may wish to measure H_T on just the n hardest jets above a given $p_{T\text{cut}}$. To do that with the event shape, one has

---

[9]The reason this is a pseudo-inverse is that $\widetilde{N}_{\text{jet}}(p_{T\text{cut}}, R)$ is a monotonically decreasing step-wise function of $p_{T\text{cut}}$, so there is a range of values of $p_{T\text{cut}}$ with the same $\widetilde{N}_{\text{jet}}$. Once the values of $p_{Ti,R}$ are known, the algorithm in appendix A scales like $N \log N$ for N particles.
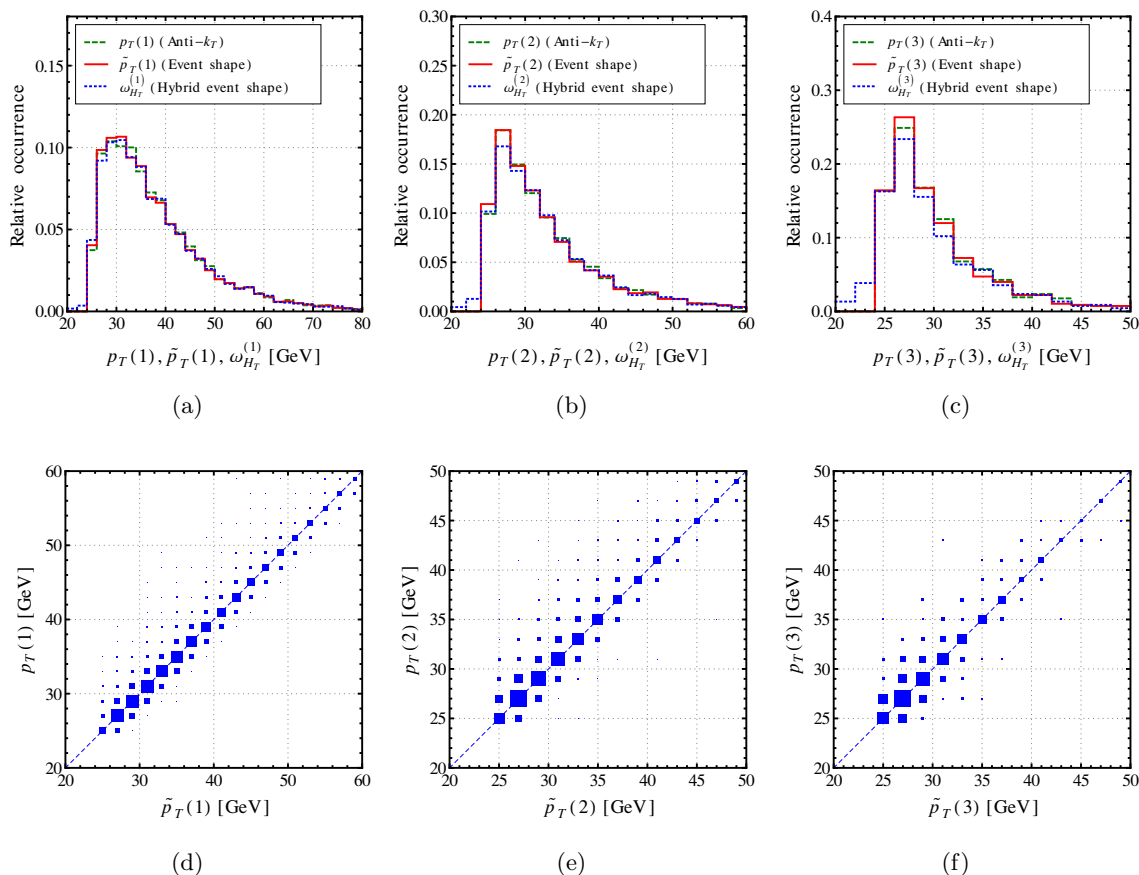
Figure 8: Transverse momentum of the three hardest jets (i.e. $p_T(1)$, $p_T(2)$, and $p_T(3)$ from left to right) for QCD dijet events. The top panels shows the transverse momentum distributions for anti-k_T jets with $R = 0.6$ and $p_{T\text{cut}} = 25\,\text{GeV}$ (green dashed curve), the corresponding event shape $\widetilde{p}_T(n)$ with the same R and $p_{T\text{cut}}$ (red curve), and the weights $\omega_{H_T}^{(n)}$ returned by the hybrid event shape with the same R but $p_{T\text{cut}} = 0$ (purple dotted curve, see section 3.2). The bottom panels shows the correlations between $p_T(n)$ and $\widetilde{p}_T(n)$, with the area of the squares proportional to the fraction of events in each bin. For plots of the $\{1\text{st}, 2\text{nd}, 3\text{rd}\}$-hardest jets, the corresponding selection criteria are $N_{\text{jet}} \geq 1, 2, 3$ (for anti-k_T) and $\widetilde{N}_{\text{jet}} \geq 0.5, 1.5, 2.5$ (for the event shape).

to find the value of a new scale $p'_{T\text{cut}}$ such that $(n+1)$-th jet would not contribute to $\widetilde{H}_T$ but the n-th jet is largely unmodified. A convenient choice for that scale is

$$p'_{T\text{cut}} = \max\{p_{T\text{cut}}, \widetilde{p}_T(n+1)\}, \tag{3.2}$$

and we will use $p'_{T\text{cut}}$ in some of the studies in section 5.

By using an algorithm similar to the one described in appendix A, one could also try to invert the number of jets $\widetilde{N}_{\text{jet}}(R)$ as a function of R, for fixed $p_{T\text{cut}}$. Strictly speaking this inverse is not possible, since $\widetilde{N}_{\text{jet}}(R)$ is not guaranteed to be a monotonic function of R. Still, we expect that the R dependence of the event shapes could be exploited much in
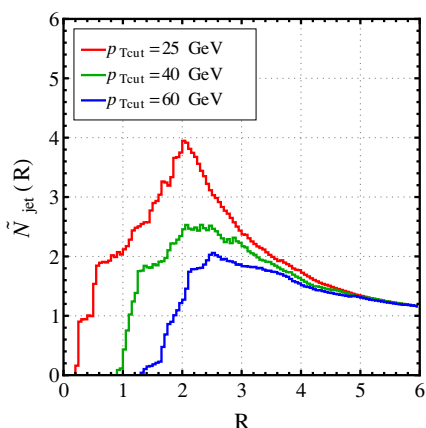
Figure 9: Number of jets $\widetilde{N}_{\mathrm{jet}}$ as a function of R for a single QCD dijet event. Shown are three values of $p_{T\mathrm{cut}} = \{25\ \mathrm{GeV}, 40\ \mathrm{GeV}, 60\ \mathrm{GeV}\}$.

the same way as for telescoping jets [45]. For example, one could measure the volatility of an event shape (à la Q-jets [34, 35]) as R is varied. A detailed study of R dependence is beyond the scope of this work, but in figure 9 we show an example of $\widetilde{N}_{\mathrm{jet}}(R)$ for a QCD dijet event, which suggests that there is interesting information to be gained by looking at multiple R values.

3.2 Jet axes and constituents

By themselves, the event shapes do not have a clustering interpretation, so in order to (uniquely) assign particles to jets we will build a hybrid event shape that incorporates some kind of clustering procedure. Before doing that, though, it is helpful to introduce the concept of an "event shape density".

Consider the following probability density for a jet axis to lie in a given direction $\hat{n}$, as determined by a standard jet clustering algorithm:

$$\rho_{N_{\mathrm{jet}}}(\hat{n}) = \sum_{\mathrm{jets}} \delta(\hat{n} - \hat{n}_{\mathrm{jet}}^r)\,\Theta(p_{T\mathrm{jet}} - p_{T\mathrm{cut}}), \tag{3.3}$$

where the superscript r reminds us that we must choose a recombination scheme for defining the jet axis $\hat{n}_{\mathrm{jet}}^r$ in terms of the constituents of that jet. For example, in the standard E-scheme, the jet axis lies in the direction of the summed constituent four-momenta. The reason $\rho_{N_{\mathrm{jet}}}$ is a density is that if we integrate over all directions $\hat{n}$ then $\int d^2\hat{n}\,\rho_{N_{\mathrm{jet}}} = N_{\mathrm{jet}}$, but $\rho_{N_{\mathrm{jet}}}$ itself has delta function spikes at the jet locations $\hat{n}_{\mathrm{jet}}^r$ identified by the jet algorithm. Similarly, we can define a transverse momentum density,

$$\rho_{H_T}(\hat{n}) = \sum_{\mathrm{jets}} p_{T\mathrm{jet}}\,\delta(\hat{n} - \hat{n}_{\mathrm{jet}})\,\Theta(p_{T\mathrm{jet}} - p_{T\mathrm{cut}}), \tag{3.4}$$

where $\int d^2\hat{n}\,\rho_{H_T} = H_T$ and the height of the delta functions correspond to the p_T of the corresponding jets.

Following the general strategy outlined in section 2, we can define corresponding event shape densities:

$$\widetilde{\rho}_{N_{\text{jet}}}(\hat{n}) = \sum_{i \in \text{event}} \frac{p_{Ti}}{p_{Ti,R}} \delta(\hat{n} - \hat{n}_{i,R}^r) \, \Theta(p_{Ti,R} - p_{T\text{cut}}), \tag{3.5}$$

$$\widetilde{\rho}_{H_T}(\hat{n}) = \sum_{i \in \text{event}} p_{Ti} \, \delta(\hat{n} - \hat{n}_{i,R}^r) \, \Theta(p_{Ti,R} - p_{T\text{cut}}), \tag{3.6}$$

where $\int d^2\hat{n} \, \widetilde{\rho}_{N_{\text{jet}}} = \widetilde{N}_{\text{jet}}$ and $\int d^2\hat{n} \, \widetilde{\rho}_{H_T} = \widetilde{H}_T$. Here, $\hat{n}_{i,R}^r$ is the direction of the recombined momenta in a cone of radius R around particle i, which of course depends on the recombination scheme r. If we choose to do recombination via the E-scheme, then $\widetilde{\rho}_{N_{\text{jet}}}$ and $\widetilde{\rho}_{H_T}$ can still be considered event shapes, since $\hat{n}_{i,R}^r$ can be written in closed form (i.e. in terms of the four-vector sum of constituents). For more general recombination schemes, though, $\widetilde{\rho}_{N_{\text{jet}}}$ and $\widetilde{\rho}_{H_T}$ are hybrid event shapes, since the specific direction of $\hat{n}_{i,R}^r$ depends on the recombination algorithm (which in general cannot be written in closed form). In contrast to standard jet clustering algorithms, finding $\hat{n}_{i,R}^r$ is a "local" procedure since it only requires knowledge about particles within a radius R of particle i.

Whereas the jet-based densities have n delta function spikes for an n-jet event, the event shape densities typically exhibit a more continuous distribution. In particular, the distribution will still show peaks corresponding to jet directions, although smeared because nearby particles will typically have (slightly) different values of $\hat{n}_{i,R}^r$. In this way, the event shape densities are similar in spirit to the jet energy flow project [26], since they effectively give a probability distribution for the jet axis locations.

Concretely, if we let $\{\hat{m}_j^r\}$ be the set of distinct directions in $\{\hat{n}_{i,R}^r\}$, we can rewrite the distributions in eqs. (3.5) and (3.6) as

$$\widetilde{\rho}_X(\hat{n}) = \sum_j \omega_{Xj} \, \delta(\hat{n} - \hat{m}_j^r), \quad X = N_{\text{jet}}, H_T. \tag{3.7}$$

The coefficients ω_{Xj} can be thought as weights corresponding to each candidate jet axis $\hat{m}_j^r$ and are given by:

$$\begin{aligned}
\omega_{N_{\text{jet}}j} &= \sum_{i \in \text{event}} \frac{p_{Ti}}{p_{Ti,R}} \Theta(p_{Ti,R} - p_{T\text{cut}}) \, \delta_{\{\hat{n}_{i,R}^r; m_j^r\}}, \\
\omega_{H_Tj} &= \sum_{i \in \text{event}} p_{Ti} \, \Theta(p_{Ti,R} - p_{T\text{cut}}) \, \delta_{\{\hat{n}_{i,R}^r; m_j^r\}},
\end{aligned} \tag{3.8}$$

where $\delta_{\{\hat{n}_{i,R}^r; m_j^r\}}$ is a Kronecker delta over the discrete sets of directions $\{\hat{n}_{i,R}^r\}$ and $\{m_j^r\}$. The weights $\omega_{N_{\text{jet}}j}$ indicate the (fractional) number of jets that should be associated with a given axis, while ω_{H_Tj} indicate the associated transverse momentum. For an isolated narrow jet, a typical recombination scheme will yield a single axis $\hat{m}^r$ with $\omega_{N_{\text{jet}}} = 1$ and $\omega_{H_T} = p_{T\text{jet}}$.

We emphasize that in this hybrid approach, a separate clustering algorithm is applied to each particle i, using just the particles within its neighborhood of radius R. For an event with N final state hadrons, one has to run N clustering algorithms, yielding N values of

$\hat{n}^r_{i,R}$, though not all of them will be distinct. In practice, it is inconvenient to have $\mathcal{O}(N)$ candidate jet axis locations, so ideally we want a recombination scheme that returns $\mathcal{O}(n)$ unique axes $\hat{m}^r_j$ for an n-jet event.

For this purpose, we will use a "winner-take-all" recombination scheme when performing the local clustering around each particle.[10] This scheme guarantees that the recombined direction will always coincides with one of the input particles, dramatically decreasing the number of unique $\hat{m}^r_j$ values. In the context of a pairwise clustering algorithm like anti-k_T, the recombination scheme determines how two pseudo-jets p_1 and p_2 will be merged to form a combined pseudo-jet p_r. In the winner-take-all scheme, the transverse momentum of p_r is given by the sum of the two pseudo-jets, but the direction of p_r is given by the hardest pseudo-jet:

$$p_{Tr} = p_{T1} + p_{T2}, \qquad \hat{n}_r = \begin{cases} \hat{n}_1 & \text{if } p_{T1} > p_{T2}, \\ \hat{n}_2 & \text{if } p_{T2} > p_{T1}. \end{cases} \tag{3.9}$$

For simplicity, we take p_r to be a massless four-vector. When used with an infrared/collinear safe clustering measure (anti-k_T in the later plots), the winner-take-all scheme is also infrared/collinear safe. Because the winner-take-all scheme always returns a jet direction aligned along one of the input particles (often the hardest particle), the set of recombined jet directions $\{\hat{m}^r_j\}$ is much smaller than the number of hadrons in the final state.[11] Of course, for later analysis, one probably wants to use the summed four-vector of the jet constituents instead of the jet axis.[12]

Another practical consideration concerns the value of $p_{T\text{cut}}$. As stated above, one can think of $\omega_{H_T j}$ in eq. (3.8) as the transverse momentum associated with jet j, so that a way to find the n hardest jets is by taking the n highest values of ω_{H_T}. However, although the sum of the ω_{H_T} returns $\widetilde{H}_T$, $p_{T\text{cut}}$ would distort the jet p_T spectrum. The reason is that the $p_{T\text{cut}}$ requirement in eqs. (3.5) and (3.6) vetoes particles near the periphery of jets which would be captured using standard clustering procedures. Note that this effect is relevant only for jets close to the $p_{T\text{cut}}$ threshold. This effect was not seen in figure 8 for $\widetilde{p}_T(n)$ because there we could compensate for the loss of peripheral particles by using $n_{\text{off}} = 0.5$ in eq. (3.1). This effect is visible, however, in figure 3a for H_T where the peaks in the event shape $\widetilde{H}_T$ (corresponding to jets at threshold) are below the peaks for the jet-based H_T because of leakage towards smaller values of $\widetilde{H}_T$. The most convenient way to restore the vetoed particles is to simply take $p_{T\text{cut}} = 0$ in eq. (3.8), in which case the

---

[10]We thank Andrew Larkoski, Duff Neill, and Gavin Salam for discussions on this point. The winner-take-all scheme is also discussed in ref. [46] in the context of recoil-free observables.

[11]To further reduce the number of jet directions, we could further insist that the winner-take-all axes are globally consistent. That is, if particle a has winner-take-all axis aligned with particle b, but particle b has winner-take-all axis aligned with particle c, then we could assign particle a the axis aligned with c (recursing further if necessary). This consistency criteria would ensure that the final set of jet directions $\{\hat{m}^r_j\}$ are their own winner-take-all axes. It would also imply that the jet regions can expand beyond a cone of radius R from the jet axes. This option is available in the FASTJET add-on, but not used in the following plots.

[12]Unlike in the E-scheme, the jet axis and the jet four-momentum (i.e. the summed four-momenta of the jet constituents) will not typically be aligned in the winner-take-all scheme.
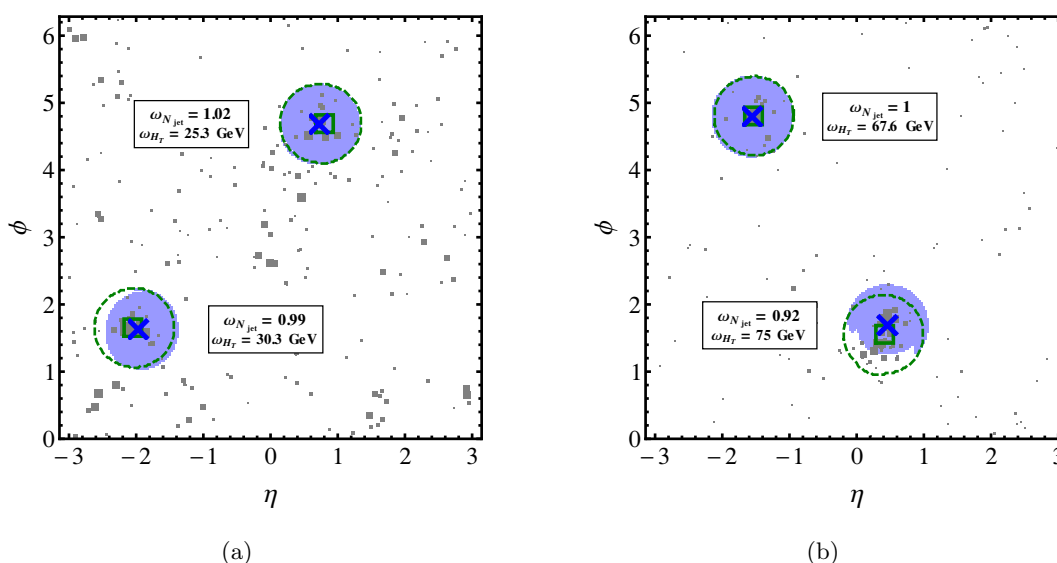
Figure 10: Two QCD dijet event displays with $R = 0.6$. The anti-k_T jet axes (green) are compared to the ones obtained using the hybrid event shape approach with $p_{T\text{cut}} = 0$ (blue). The standard E-scheme is used for the anti-k_T jets, whereas the hybrid event shape uses the winner-take-all recombination scheme, as explained in section 3.2. The light blue shaded region corresponds to (passive) ghost particles which are clustered to the given axis, and the dashed green curve gives the anti-k_T boundary. The weights $\omega_{N_{\text{jet}}} \simeq 1$ and $\omega_{H_T} \simeq p_{T\text{jet}}$ associated with the event shape axes are also shown.

sum of the ω_{H_T} yields the total sum of scalar p_T in the event (though the sum of the $\omega_{N_{\text{jet}}}$ is no longer infrared safe).

We now compare standard jet clustering to the hybrid event shape approach. For anti-k_T jets, we use the standard E-scheme recombination, whereas for the hybrid event shape, we use the anti-k_T clustering measure with winner-take-all recombination for the local clustering around each particle. In figure 10 we show two QCD dijet events comparing the two hardest jets from anti-k_T with the jets defined by the two highest weights ω_{H_T} (with $p_{T\text{cut}} = 0$). We also show the corresponding values of $\omega_{N_{\text{jet}}}$ and ω_{H_T}. The displayed jet regions are determined by adding (passive) ghost particles [21]. There are differences between the jet axes caused by the different recombination schemes, and differences in the jet regions from the different effective jet splitting criteria. But overall, there is a good correlation between the two methods, and the fact that $\omega_{N_{\text{jet}}} \simeq 1$ is a nice cross check.

Turning to the QCD dijet event sample, back in figure 8 we showed distributions for the three highest weights ω_{H_T}, which correlate strongly with the three hardest jets from anti-k_T (and with the inverse multiplicity $\widetilde{p}_T(n)$). In figure 11, we compare the direction of the axis of the hardest jet found with both methods, again seeing good agreement, apart from a small set of events where the azimuth differs by π because the choice of hardest jet is ambiguous. In the three panels of figure 12, we show various effects on the hardest jet of having $p_{T\text{cut}} = 0$ versus non-zero $p_{T\text{cut}}$. The (passive) jet areas are shown in figure 12a,

where the jet area distribution is peaked around πR^2 for $p_{T\text{cut}} = 0$ (similar to anti-k_T) whereas the area is smaller for non-zero $p_{T\text{cut}}$ because of peripheral vetoes. The same effect is seen in figure 12b, where a non-zero $p_{T\text{cut}}$ decreases the ω_{H_T} value. The effect is less visible for $\omega_{N_{\text{jet}}}$ in the figure 12c, since most events peak at 1, but there is a shift to lower $\omega_{N_{\text{jet}}}$ as $p_{T\text{cut}}$ is increased. We thus conclude that $p_{T\text{cut}} = 0$ gives results that are closer to the expectation from standard jet clustering.

In terms of computational cost, the hybrid event shape approach is significantly more costly than anti-k_T, since one has to effectively run a separate jet clustering procedure for each particle i to determine the direction $\hat{n}^r_{i,R}$. On a standard laptop, it is roughly a factor of four slower on dijet events. Despite the speed issue, this approach to identifying candidate jet regions might still be appropriate for trigger-level analyses because of the parallelizable and local nature of the hybrid event shapes. The winner-take-all recombination is crucial for this approach to work, since it ensures that only a small number of candidate jet axes are identified. It also has the nice feature that a given jet axis is guaranteed to align along one of the input particle directions.

4 Shape trimming

Thus far, we have only discussed event shapes for observables built as a sum over all jets in an event. As discussed further in section 5, the same basic strategy can be applied to observables which are a sum over all subjets in all jets in an event. A simple application of this is to implement jet trimming [19] via an event shape. We refer to traditional jet trimming as "tree trimming" and the corresponding event shape version as "shape trimming".

In tree trimming, one first clusters particles into jets of radius R and $p_{T\text{jet}} > p_{T\text{cut}}$, typically via the anti-k_T algorithm. For each jet, one reclusters its constituents into subjets with characteristic radius $R_{\text{sub}} < R$, typically via the CA algorithm [47–49] or k_T algorithm [50, 51]. Subjets whose transverse momentum fraction $p_{T\text{sub}}/p_{T\text{jet}}$ are above a certain threshold f_{cut} are kept, while the remaining subjets are removed. The four-momentum of a trimmed jet can be written as

$$t^\mu_{\text{jet}} = \sum_{\text{subjets}} p^\mu_{\text{sub}} \Theta \left(\frac{p_{T\text{sub}}}{p_{T\text{jet}}} - f_{\text{cut}} \right), \tag{4.1}$$

where p^μ_{sub} is the four-momentum of the subjet, $p_{T\text{sub}}$ is the corresponding transverse momentum, and $p_{T\text{jet}}$ is the transverse momentum of the un-trimmed jet. The trimmed four-momentum of the entire event is

$$t^\mu_{\text{event}} = \sum_{\text{jets}} t^\mu_{\text{jet}} \Theta(p_{T\text{jet}} - p_{T\text{cut}}) = \sum_{\text{jets}} \sum_{\text{subjets}} p^\mu_{\text{sub}} \Theta \left(\frac{p_{T\text{sub}}}{p_{T\text{jet}}} - f_{\text{cut}} \right) \Theta(p_{T\text{jet}} - p_{T\text{cut}}). \tag{4.2}$$

Along with the clustering algorithms used, the trimming procedure is specified by the jet parameters $\{p_{T\text{cut}}, R\}$ and the subjet parameters $\{f_{\text{cut}}, R_{\text{sub}}\}$.

To recast trimming as an event shape, we can follow the strategy outlined in section 2, but adding an extra step to deal with the presence of subjets. Since $p^\mu_{\text{sub}} \simeq \sum_{i \in \text{subjet}} p^\mu_i$
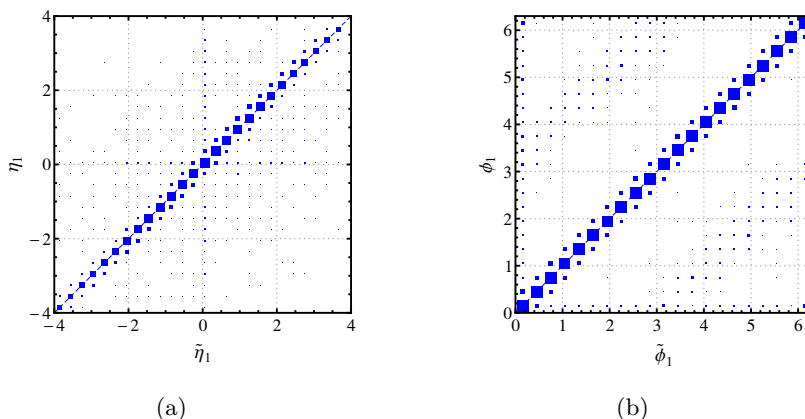
(a)   (b)

Figure 11: Position of the hardest jet axis for QCD dijet events, using the same jet clustering as figure 10. The (η_1, ϕ_1) coordinates correspond to the jet axis identified with anti-k_T, and $(\widetilde{\eta}_1, \widetilde{\phi}_1)$ are the coordinates found with the hybrid event shape. The area of the squares is proportional to the fraction of events in each bin. There is a slight difference in the jet direction due to the different recombination schemes (E-scheme for anti-k_T, winner-take-all for the hybrid event shape). Note the (small) accumulation of events at $|\phi_1 - \widetilde{\phi}_1| = \pi$, which occur when the two algorithms disagree about which of the dijets is the hardest.

(a)   (b)   (c)
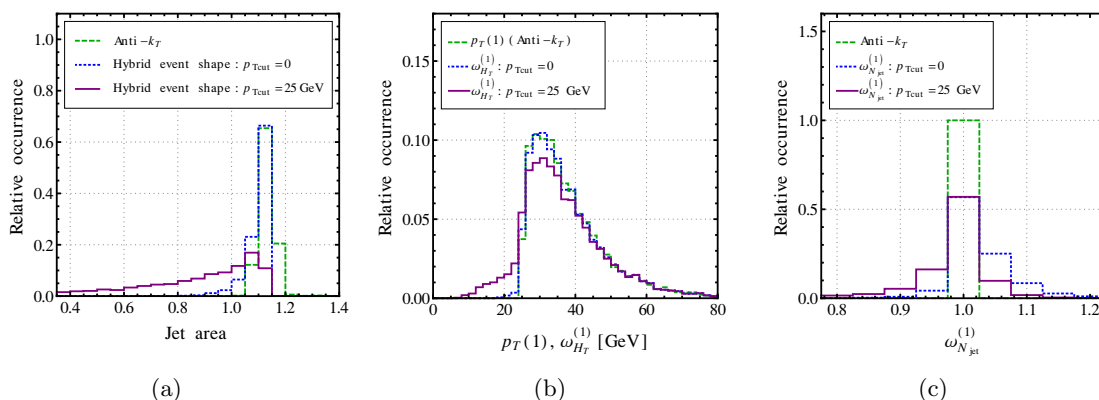
Figure 12: Comparison of the hardest jet found with anti-k_T (green dashed curve), the hybrid event shape result with $p_{T\text{cut}} = 0$ (blue dotted curve), and the hybrid event shape with $p_{T\text{cut}} = 25\,\text{GeV}$ (purple curve), all for QCD dijet events with $R = 0.6$. Left: Passive jet area, where the first two methods peak at πR^2. Center: jet p_T (or ω_{H_T}). Right: fractional jet weight $\omega_{N_{\text{jet}}}$, where all methods peak at 1. In all cases, the $p_{T\text{cut}} = 0$ event shape is closer to the anti-k_T result, since it restores peripheral particles that are vetoed with non-zero $p_{T\text{cut}}$.

can be written as a sum over subjet's constituents, we can skip the first replacement in eq. (2.5), and directly make the replacement

$$\sum_{\text{jets}} \sum_{\text{subjets}} p_{\text{sub}}^{\mu} \to \sum_{i \in \text{event}} p_i^{\mu}. \tag{4.3}$$

Moreover,

$$p_{T\text{jet}} \to p_{Ti,R}, \qquad p_{T\text{sub}} \to p_{Ti,R_{\text{sub}}}, \tag{4.4}$$

where $p_{Ti,R_{\text{sub}}}$ is analogous to $p_{Ti,R}$ in eq. (2.7), except it only includes particles contained in a cone around particle i of radius R_{sub}. The trimmed event shape corresponding to the overall four-momentum is therefore

$$\widetilde{t}_{\text{event}}^{\mu} = \sum_{i \in \text{event}} p_i^{\mu} \Theta \left(\frac{p_{Ti,R_{\text{sub}}}}{p_{Ti,R}} - f_{\text{cut}} \right) \Theta(p_{Ti,R} - p_{T\text{cut}}). \tag{4.5}$$

For defining more general event shapes (or for use in other jet-based analyses), we can interpret $\widetilde{t}_{\text{event}}^{\mu}$ as defining a weight for each individual particle:

$$w_i = \Theta \left(\frac{p_{Ti,R_{\text{sub}}}}{p_{Ti,R}} - f_{\text{cut}} \right) \Theta(p_{Ti,R} - p_{T\text{cut}}). \tag{4.6}$$

Here w_i is either 0 or 1, but one could generalize w_i to take on continuous values by smoothing out the theta functions. In practice, we implement eq. (4.6) as a `Selector` in our FASTJET add-on, which takes a collection of particles and only returns those particles with $w_i = 1$. Instead of applying trimming event wide ("event shape trimming"), one could first find jets with an ordinary jet algorithm and then apply eq. (4.6) with $p_{Ti,R}$ replaced by $p_{T\text{jet}}$; we have implemented this "jet shape trimming" option as a `Transformer` in FASTJET.

One could also use the weights directly in the event shapes. For example, we could define the trimmed jet multiplicity as

$$\widetilde{N}_{\text{jet}}^{\text{trim}}(p_{T\text{cut}}, R; f_{\text{cut}}, R_{\text{sub}}) = \sum_{i \in \text{event}} \frac{p_{Ti}}{p_{Ti,R}} \Theta \left(\frac{p_{Ti,R_{\text{sub}}}}{p_{Ti,R}} - f_{\text{cut}} \right) \Theta(p_{Ti,R} - p_{T\text{cut}}), \tag{4.7}$$

and one could define the trimmed inverse $\widetilde{p}_T^{\text{trim}}(n, R; f_{\text{cut}}, R_{\text{sub}})$ accordingly. Note that applying the weights in eq. (4.6) first and then calculating $\widetilde{N}_{\text{jet}}$ is not the same as calculating $\widetilde{N}_{\text{jet}}^{\text{trim}}$ directly, since in the former case, the value of $p_{Ti,R}$ is affected by the weights. In most cases, one gets better performance by using the weights first, especially if the jet observable $\mathcal{F}_{\text{jet}}$ is non-linear in the inputs (as is the case for jet mass studied in section 5.1). For $\widetilde{N}_{\text{jet}}^{\text{trim}}$ there is only a mild difference, so we use $\widetilde{N}_{\text{jet}}^{\text{trim}}$ for simplicity in some of the case studies in section 5.

To compare the behavior of ordinary tree trimming and shape trimming, we use event samples from the BOOST 2010 report [14]. In particular, we analyze a boosted top signal and the corresponding QCD background in the p_T bin 500 GeV $< p_T <$ 600 GeV.[13] In

---

[13]Event samples from BOOST 2010 and details about events generation can be found at http://www.lpthe.jussieu.fr/~salam/projects/boost2010-events/herwig65 and http://tev4.phys.washington.edu/TeraScale/boost2010/herwig65. These events are for the 7 TeV LHC generated with HERWIG 6.510 [52], with underlying event given by JIMMY [53] with an ATLAS tune [54].
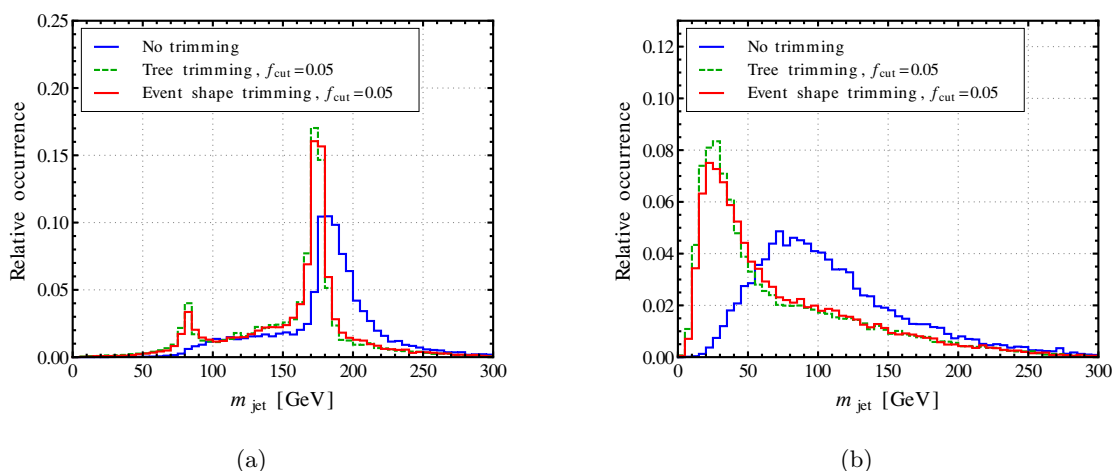
Figure 13: Boosted top sample (left) and corresponding QCD background (right) from the BOOST 2010 event samples [14]. For ordinary tree trimming, we identify jets anti-k_T jets with $R = 1.0$ and $p_{T\text{cut}} = 200$ GeV, and then applying trimming with $R_{\text{sub}} = 0.3$ and $f_{\text{cut}} = 0.05$. For shape trimming, we apply event-wide trimming using the same R_{sub} and f_{cut} parameter before clustering with anti-k_T. In both cases, we plot the masses of the two hardest jets per event.

figure 13, we show the effect of trimming on the jet mass spectrum for the boosted top signal and the corresponding QCD background. For tree trimming, we build anti-k_T jets with $R = 1.0$ and $p_{T\text{cut}} = 200$ GeV and trim with $R_{\text{sub}} = 0.3$ and $f_{\text{cut}} = 0.05$. For shape trimming, we use the same set of parameters to trim the entire event according to weights from eq. (4.6), and then build anti-k_T jets with $R = 1.0$ and $p_{T\text{cut}} = 200$ GeV. We see that the behavior of both trimming methods is very similar, and that both methods emphasize the boosted top mass peak while suppressing the high-mass QCD background. Jet shape trimming is not shown in figure 13 as it performs very similarly to event shape trimming.

One important application for trimming is pileup mitigation [55, 56]. To study its effectiveness, we take our sample of $Z(\to \nu\nu) + j$ events from section 2 and overlay N_{PV} soft QCD events generated with PYTHIA 8.157 [41].[14] We consider three options: ordinary tree trimming, shape trimming applied to the individual jets (jet shape trimming), and shape trimming applied to the entire event (event shape trimming). figure 14 shows the average of the hardest jet mass as a function of N_{PV}, where the jets are built using anti-k_T with $R = 1.0$ and $p_{T\text{cut}} = 500$ GeV. Taking $R_{\text{sub}} = 0.3$ in all cases, we find a comparable degree of stability against pileup for tree trimming with $f_{\text{cut}} = 0.05$ (as was done in ref. [56]), jet shape trimming with $f_{\text{cut}} = 0.07$, and event shape trimming with $f_{\text{cut}} = 0.07$. Note that event shape trimming has the largest variation with N_{PV}, as expected since $p_{Ti,R}$ is typically lower than $p_{T\text{jet}}$, and therefore does not groom as aggressively. Part of the reason we need a different f_{cut} value for tree trimming versus shape trimming is that the effective subjet areas of the two methods are different.

---

[14]Here, the minimum p_T for the hard process at generator level has been reset to $p_{T\text{cut}}^{\text{parton}} = 350$ GeV.

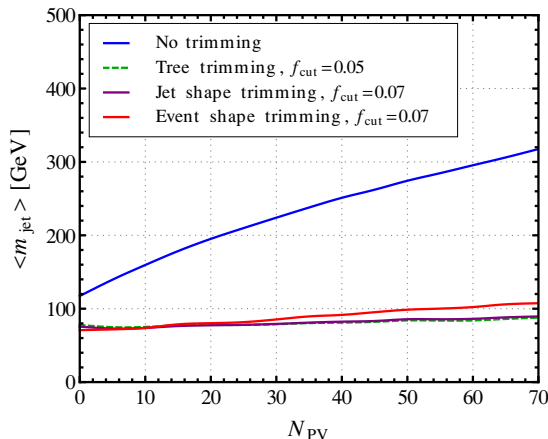**Figure 14**: Pileup mitigation for $Z(\to \nu\nu) + j$ events. Shown is the mean of the hardest jet mass distribution as a function of the number of primary vertices $N_{\rm PV}$. In all cases we use $R_{\rm sub} = 0.3$, with $f_{\rm cut} = 0.05$ for tree-trimming and $f_{\rm cut} = 0.07$ for the two shape trimming options.

A complementary way to do pileup mitigation is via area subtraction [20–22]. It is straightforward to correct the trimming weights in eq. (4.6) using area subtraction, because $p_{Ti,R_{\rm sub}}$ and $p_{Ti,R}$ are defined in terms of fixed-radius regions, and therefore have fixed areas $\pi R_{\rm sub}^2$ and πR^2 respectively. At present, our FASTJET implementation of shape trimming does not include area subtraction, but we anticipate including that functionality in a future version.

5 Generalizations

5.1 Other jet-like event shapes

The general procedure to build event shapes $\widetilde{\mathcal{F}}$ from single jet observables $\mathcal{F}_{\rm jet}$ was given in section 2. Here we give a few more examples beyond $\widetilde{N}_{\rm jet}$, $\widetilde{H}_T$, and $\widetilde{\not{p}}_T$.

As a simple generalization of $N_{\rm jet}$ and H_T, consider the jet-based observable

$$H_T^n(p_{T\rm cut}, R) = \sum_{\rm jets} p_{T\rm jet}^n \, \Theta(p_{T\rm jet} - p_{T\rm cut}),\tag{5.1}$$

where $n = 0$ ($n = 1$) corresponds to $N_{\rm jet}$ (H_T). Using the method in section 2, the corresponding event shape is

$$\widetilde{H}_T^n(p_{T\rm cut}, R) = \sum_{i \in \rm event} \frac{p_{T,i}}{p_{Ti,R}} (p_{Ti,R})^n \, \Theta(p_{Ti,R} - p_{T\rm cut}).\tag{5.2}$$

In figure 15, we compare $H_T^n(p_{T\rm cut}, R)$ to $\widetilde{H}_T^n(p_{T\rm cut}, R)$ for $n = -1$ in QCD dijet events, using the same event generation scheme as section 2.

A more complicated example is the sum of jet masses in an event,

$$M_J(p_{T\rm cut}, R) = \sum_{\rm jets} m_{\rm jet} \, \Theta(p_{T\rm jet} - p_{T\rm cut}).\tag{5.3}$$
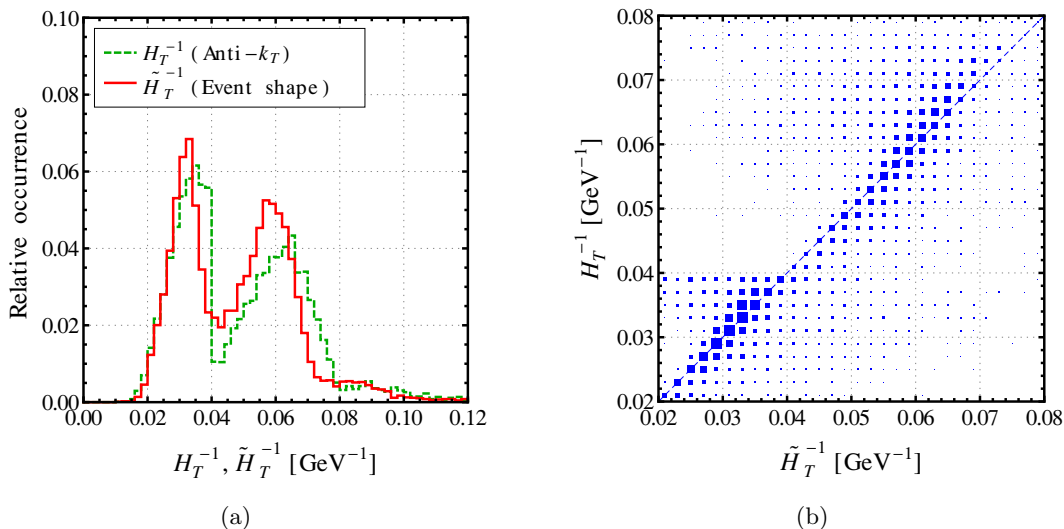
– 20 –

Figure 15: Summed transverse momentum inverse (i.e. H_T^n with $n = -1$) for QCD dijet events. The jet parameters, formatting, and cuts are the same as for figure 2.
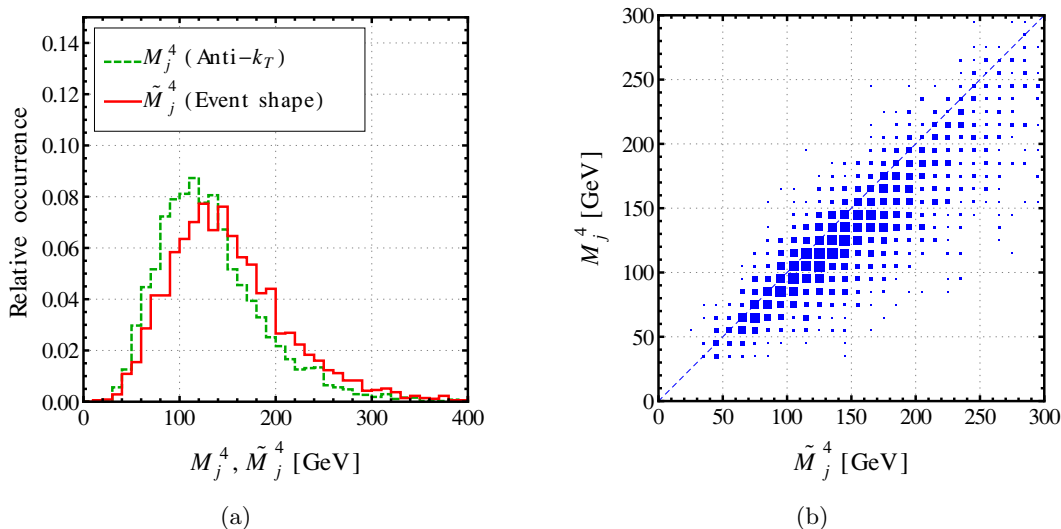

Figure 16: Summed jet mass analysis for M_J^4 that mimics ref. [39]. Shown is a QCD four-jet sample with the (trimmed) summed jet mass of the four hardest jets. For the anti-k_T version, the trimmed jets have $R = 1.2$, $p_{T\text{cut}} = 50$ GeV, $R_{\text{sub}} = 0.3$, and $f_{\text{cut}} = 0.05$, requiring at least four such jets and the hardest jet above 100 GeV. For the event shape version, the event selection criteria is $\widetilde{N}_{\text{jet}}^{\text{trim}} > 3.5$ and $\widetilde{p}_T^{\text{trim}}(1) > 100$ GeV with the same jet and trimming parameters above, and the observable is $\widetilde{M}_J^4$ defined in eq. (5.5), calculated after the trimming weights in eq. (4.6) are applied.

The corresponding event shape is given by

$$\widetilde{M}_J(p_{T\text{cut}}, R) = \sum_{i \in \text{event}} \frac{p_{T,i}}{p_{Ti,R}} m_{i,R} \, \Theta(p_{Ti,R} - p_{T\text{cut}}), \tag{5.4}$$

where $m_{i,R} = \sqrt{|p^\mu_{i,R}|^2}$. One could of course raise $m_{i,R}$ to a power in analogy with $\widetilde{H}^n_T$.

Summed jet mass is a potentially powerful variable to study high jet multiplicity events at the LHC [37], and can be combined with other substructure observables to control QCD multijet backgrounds to new physics searches [38, 39]. As an example, it is instructive to see how to mimic aspects of such an analysis using event shapes. In ref. [39], events were clustered into fat jets with $R = 1.2$, the fat jets were trimmed ($R_{\text{sub}} = 0.3$, $f_{\text{cut}} = 0.05$), and events were retained if they had at least four fat jets above $p_{T\text{cut}} = 50$ GeV and the hardest jet above 100 GeV. Then the (trimmed) summed jet mass was taken for just the four hardest jets. To mimic the selection procedure, one would take events with $\widetilde{N}^{\text{trim}}_{\text{jet}}(p_{T\text{cut}}, R; f_{\text{cut}}, R_{\text{sub}}) > 3.5$ (see eq. (4.7)) and $\widetilde{p}^{\text{trim}}_T(1) > 100$ GeV. To mimic the observable, one would first apply the shape trimming weights from eq. (4.6), and then define

$$\widetilde{M}^4_J(p_{T\text{cut}}, R) \equiv M_J(p'_{T\text{cut}}, R), \qquad p'_{T\text{cut}} = \max\{p_{T\text{cut}}, \widetilde{p}^{\text{trim}}_T(5)\}, \tag{5.5}$$

where $p'_{T\text{cut}}$ effectively picks out the four hardest jets (see eq. (3.2)). In figure 16, we compare the distributions of the (trimmed) summed mass calculated using the two different methods on a QCD four-jet sample. Despite the somewhat complicated form of the event shape version, there are clear correlations between the methods. We will discuss the subjet counting aspect of ref. [39] in section 5.3.[15]

5.2 Subjet-like jet shapes

Thus far, we have focused on jet-like event shapes, but it is clear that the same technique can be applied to subjet-like jet shapes. These jet shapes would probe the substructure of a given jet, and can be defined according to eqs. (2.4) and (2.8) with "jet" replaced by "subjet" and "event" replaced by "jet". Concretely, given a jet found using an ordinary jet algorithm, consider a subjet-based observable built from subjets of radius R_{sub} above $p_{T\text{subcut}}$:

$$\mathcal{G}(p_{T\text{subcut}}, R_{\text{sub}}) = \sum_{\text{subjets}} \mathcal{G}_{\text{subjet}} \, \Theta(p_{T\text{sub}} - p_{T\text{subcut}}), \tag{5.6}$$

where $\mathcal{G}_{\text{subjet}} \equiv g(\{p^\mu_j\}_{j \in \text{subjet}})$ depends on the kinematics of the individual subjet constituents. The corresponding jet shape would be

$$\widetilde{\mathcal{G}}(p_{T\text{cut}}, R_{\text{sub}}) = \sum_{i \in \text{jet}} \frac{p_{Ti}}{p_{Ti,R_{\text{sub}}}} \mathcal{G}_{i,R_{\text{sub}}} \, \Theta(p_{Ti,R_{\text{sub}}} - p_{T\text{subcut}}), \tag{5.7}$$

where $\mathcal{G}_{i,R_{\text{sub}}} \equiv g(\{p^\mu_j \, \Theta(R_{\text{sub}} - \Delta R_{ij})\}_{j \in \text{jet}})$.

---

[15]The event-subjettiness variable of ref. [38] is defined as a geometric mean of N-subjettiness ratios [25, 57] measured on individual jets. To convert that to an event shape, we would first take the logarithm, since that would correspond to a sum over the logs of individual jet observables, and is therefore in the form needed in eq. (2.4).
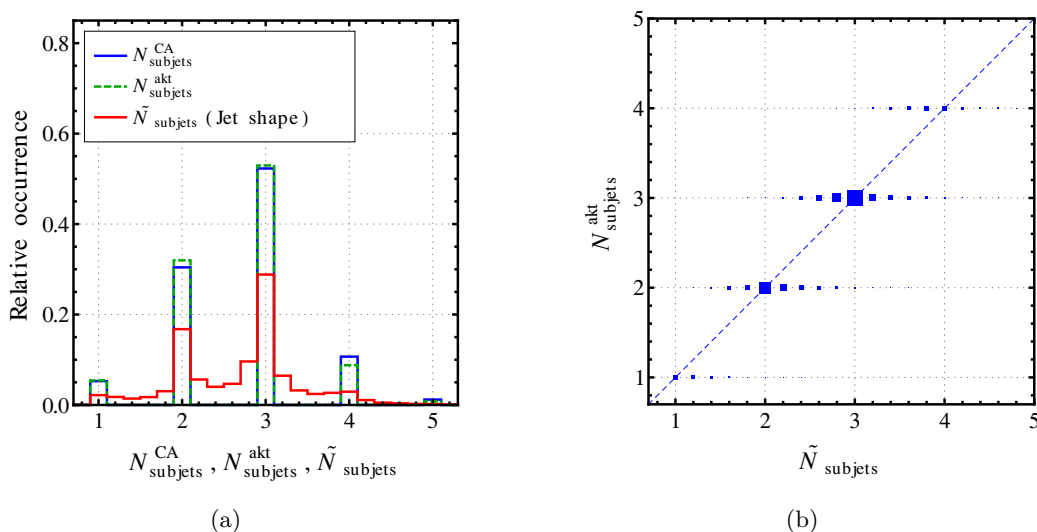
Figure 17: Subjet multiplicity (i.e. N_{subjet}) on the boosted top sample from BOOST 2010. The jet selection is the same as in figure 13, and we count subjets either with Cambridge-Aachen clustering, anti-k_T clustering, or $\widetilde{N}_{\text{subjet}}$. In all cases, we take $R_{\text{sub}} = 0.3$ and $p_{T\text{subcut}} = 0.05\, p_{T\text{jet}}$. In the case of Cambridge-Aachen clustering, this is equivalent to counting the subjets left from (CA) trimming with $R_{\text{sub}} = 0.3$ and $f_{\text{cut}} = 0.05$.

As an example, a jet shape that counts the subjet multiplicity is

$$\widetilde{N}_{\text{subjet}}(p_{T\text{subcut}}, R_{\text{sub}}) = \sum_{i \in \text{jet}} \frac{p_{Ti}}{p_{Ti,R_{\text{sub}}}} \Theta(p_{Ti,R_{\text{sub}}} - p_{T\text{subcut}}). \tag{5.8}$$

In figure 17 we study subjet multiplicity for the same boosted top sample analyzed in section 4. Starting from anti-k_T jets with $R = 1.0$ and $p_{T\text{cut}} = 200$ GeV, we count the number of subjets in three different ways. First, we count the number of Cambridge-Aachen subjets left after trimming is applied with $R_{\text{sub}} = 0.3$ and $f_{\text{cut}} = 0.05$. Second, we re-run anti-k_T clustering on the jet with $R_{\text{sub}} = 0.3$ and $p_{T\text{subcut}} = f_{\text{cut}}\, p_{T\text{jet}}$. Third, we use the jet shape $\widetilde{N}_{\text{subjet}}$ with the same value of R_{sub} and $p_{T\text{subcut}}$. The first two methods necessarily yield integer values, whereas $\widetilde{N}_{\text{subjet}}$ is continuous. All three methods peak at $N_{\text{subjet}} = 3$, as expected since this is a boosted top quark sample.

5.3 Subjet-like event shapes

Our final generalization is to observables that are inclusive over the subjets in an entire event. That is, we want to start from an observable defined in terms of the constituents in a subjet, summed over all subjets in each jet, and then further summed over all jets in the event. Consider an observable built from jets of radius R above $p_{T\text{cut}}$ with subjets of radius R_{sub} above $p_{T\text{subcut}}$:

$$\mathcal{H}(p_{T\text{cut}}, R_{\text{sub}}; p_{T\text{subcut}}, R_{\text{sub}}) = \sum_{\text{jets}} \sum_{\text{subjets}} \mathcal{H}_{\text{subjet}}\Theta(p_{T\text{sub}} - p_{T\text{subcut}})\Theta(p_{T\text{jet}} - p_{T\text{cut}}), \tag{5.9}$$
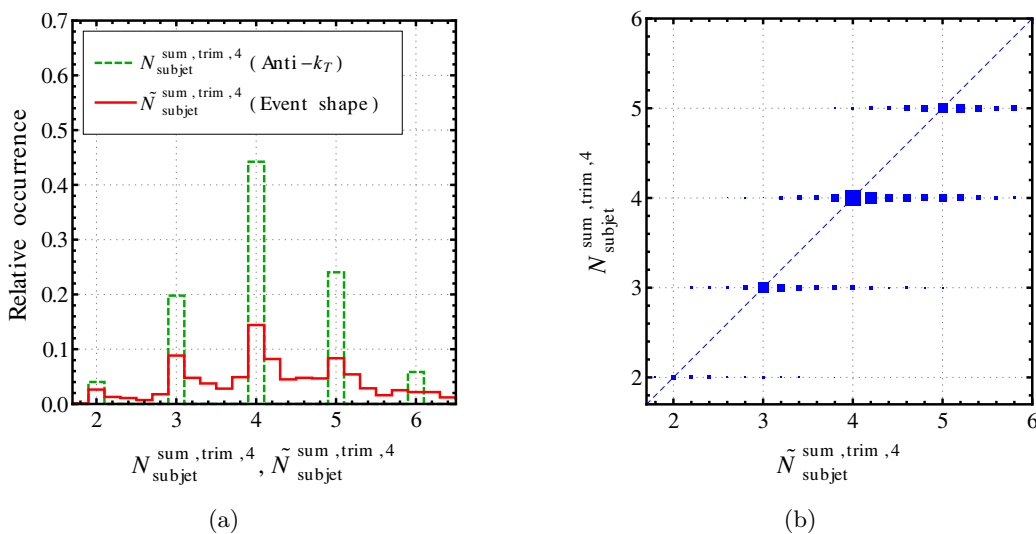
Figure 18: Summed subjet multiplicity (i.e. $N_{\text{subjet}}^{\text{sum}}$) on a QCD four-jet sample, in the spirit of ref. [39]. The jet selection is the same as figure 16, and we use the event shape $\widetilde{N}_{\text{subjet}}^{\text{sum,trim,4}}$ from eq. (5.14).

where $\mathcal{H}_{\text{subjet}} \equiv h(\{p_j^\mu\}_{j\in\text{subjet}})$ depends on the kinematics of the subjet constituents. The corresponding event shape is

$$\widetilde{\mathcal{H}}(p_{T\text{cut}}, R; p_{T\text{subcut}}, R_{\text{sub}}) = \sum_{i\in\text{event}} \frac{p_{Ti}}{p_{Ti,R_{\text{sub}}}} \mathcal{H}_{i,R_{\text{sub}}} \, \Theta(p_{Ti,R_{\text{sub}}} - p_{T\text{subcut}})\Theta(p_{Ti,R} - p_{T\text{cut}}), \tag{5.10}$$

where $\mathcal{H}_{i,R_{\text{sub}}} \equiv h(\{p_j^\mu \Theta(R_{\text{sub}} - \Delta R_{ij})\}_{j\in\text{event}})$. Note that the weight factor depends on $p_{Ti,R_{\text{sub}}}$, and $p_{Ti,R}$ only appears for testing $p_{T\text{cut}}$.

For measurement functions $\mathcal{H}_{\text{subjet}}$ that are expressible as a sums over the subjet constituents,

$$\mathcal{H}_{\text{subjet}} = \sum_{j\in\text{subjet}} \tilde{h}(p_j^\mu), \tag{5.11}$$

where $\tilde{h}$ is a single particle measurement function, we can elide the $p_{Ti}/p_{Ti,R_{\text{sub}}}$ weighting factor and directly write down the event shape

$$\widetilde{\mathcal{H}}(p_{T\text{cut}}, R; p_{T\text{subcut}}, R_{\text{sub}}) = \sum_{i\in\text{event}} \tilde{h}(p_i^\mu) \, \Theta(p_{Ti,R_{\text{sub}}} - p_{T\text{subcut}})\Theta(p_{Ti,R} - p_{T\text{cut}}). \tag{5.12}$$

The shape trimming technique from section 4 can be expressed as such an event shape, with $p_{T\text{subcut}} = f_{\text{cut}}\, p_{Ti,R}$ and $\tilde{h}(p_j^\mu) = p_j^\mu$ (see eq. (4.5)).

Following the example of subjet multiplicity $\widetilde{N}_{\text{subjet}}$ in eq. (5.8), we can define the (trimmed) summed subjet multiplicity:

$$\widetilde{N}_{\text{subjet}}^{\text{sum,trim}}(p_{T\text{cut}}, R; f_{\text{cut}}, p_{T\text{subcut}}, R_{\text{sub}}) \tag{5.13}$$
$$= \sum_{i\in\text{event}} \frac{p_{Ti}}{p_{Ti,R_{\text{sub}}}} \Theta(p_{Ti,R_{\text{sub}}} - p_{T\text{subcut}})\Theta\left(\frac{p_{Ti,R_{\text{sub}}}}{p_{Ti,R}} - f_{\text{cut}}\right)\Theta(p_{Ti,R} - p_{T\text{cut}}),$$

where the trimming criteria on the subjets is only imposed if it is stricter than the $p_{T\text{subcut}}$ requirement. A similar variable was used in ref. [39] to isolate high jet multiplicity events at the LHC, in concert with the summed jet mass already mentioned in section 5.1. Here, however, we are restricted to defining subjets with a fixed radius R_{sub}, as opposed to the more dynamical subjet finding procedures advocated in ref. [39].[16] In figure 18, we compare subjet counting using anti-k_T for both fat jets and subjets to the comparable procedure with $\widetilde{N}_{\text{subjet}}^{\text{sum}}$ on the QCD four-jet sample. We use the same event selection as in section 5.1, and define

$$\widetilde{N}_{\text{subjet}}^{\text{sum,trim,4}}(p_{T\text{cut}}, R; f_{\text{cut}}, p_{T\text{subcut}}, R_{\text{sub}}) = \widetilde{N}_{\text{subjet}}^{\text{sum,trim}}(p'_{T\text{cut}}, R; f_{\text{cut}}, p_{T\text{subcut}}, R_{\text{sub}}) \quad (5.14)$$

with $p'_{T\text{cut}} = \max\{p_{T\text{cut}}, \widetilde{p}_T^{\text{trim}}(5)\}$ to effectively isolate the four hardest jets. Apart from the non-integer nature of $\widetilde{N}_{\text{subjet}}^{\text{sum,trim,4}}$, there is a clear correlation between the methods.

6 Conclusions

In this paper, we have shown how inclusive jet observables can be recast as jet-like event shapes. By replacing an inclusive sum over jets in an event with an inclusive sum over particles in an event, we have removed the dependence on the jet clustering procedure, while still maintaining the jet-like radius R and jet-like momentum cut $p_{T\text{cut}}$ expected in jet-based analyses. While our original method can only be applied to inclusive jet observables, we have shown one example where more exclusive information about single jets was obtained by inverting the jet multiplicity event shape $\widetilde{N}_{\text{jet}}$ to determine the p_T of the n-th hardest jet. Our focus was on event shapes in this paper, though we have shown that there is a straightforward generalization to jet shapes, which may find use in jet substructure studies.

A promising possible application of these event shapes is for event selection at the trigger level, especially given their local computational structure. To the best of our knowledge, all jet triggers presently in use on the ATLAS and CMS experiments can be mimicked by appropriate combinations of $\widetilde{N}_{\text{jet}}$, $\widetilde{H}_T$, and $\widetilde{\slashed{p}}_T$ cuts (choosing different values of R and $p_{T\text{cut}}$ as needed). It may even be possible to do preliminary jet identification at the trigger level using the hybrid event shapes with winner-take-all recombination; the local nature of the clustering means that the approach can be parallelized across the detector without double-counting. Of course, more detailed feasibility studies are needed to see whether these event shapes can be incorporated into the trigger upgrades planned for high-luminosity LHC running.

For analysis-level jet studies, the event shapes provide a complementary characterization of the gross jet-like nature of the event. From the correlations seen in section 2, one should expect $\mathcal{F}$ and $\widetilde{\mathcal{F}}$ to have similar performance in an experimental context. There can be important differences, however, in regions of phase space where jets are overlapping or otherwise ambiguous. Thus, a comparison between, say, a selection criteria based on N_{jet} and one based on $\widetilde{N}_{\text{jet}}$ would offer a useful test for the robustness of an analysis.

---

[16]In principle, one could choose the subjet radius R_{sub} to be a (local) function of the particles within a radius R of particle i.

A novel application of our method is for jet grooming via shape trimming. This worked because ordinary tree trimming [19] can be written as a double sum over subjets and jets in an event, allowing an application of the general techniques in section 5.3. Shape trimming can be applied to event shapes themselves, or it can be interpreted as simply assigning a weight to each particle in an event, after which one can perform a traditional jet-based analysis. Shape trimming has similar pileup mitigation performance to tree trimming, but can be more easily applied event-wide since it does not require the explicit identification of jets or subjets.

Other grooming techniques beyond trimming deserve future study, though we do not know (yet) how to cast them as event shapes. For example, filtering [16] is based on keeping a fixed number of subjets, which we do not know how to implement as an inclusive sum over all particles in an event. Similarly, pruning [17, 18] and (modified) mass drop [16, 58, 59] are based on recursively applying a selection criteria, which have no obvious event shape counterpart. The modified mass drop procedure is particularly interesting because it removes Sudakov double logarithms [58, 59], and a non-recursive event shape version of this procedure would help for understanding this unique behavior.

Finally, these event shapes are particularly interesting for future analytic studies in perturbative QCD. Formally, an inclusive jet observable $\mathcal{F}$ and its event shape counterpart $\widetilde{\mathcal{F}}$ are exactly equivalent for infinitely narrow jets separated by more than R, such that they share the same soft-collinear structure. Therefore, up to non-singular and power-suppressed terms, we expect $\mathcal{F}$ and $\widetilde{\mathcal{F}}$ to have similar (if not identical) factorization and resummation properties. That said, there is clearly a difference between the integer-valued jet multiplicity N_{jet} and the continuous event shape $\widetilde{N}_{\mathrm{jet}}$, though the difference does not show up until $\mathcal{O}(\alpha_s)$ (for jets separated by more than R but less than $2R$) or $\mathcal{O}(\alpha_s^2)$ (for jets separated by more than $2R$). We expect that understanding the origin of non-integer $\widetilde{N}_{\mathrm{jet}}$ values is likely to shed considerable light on the jet-like nature of QCD.

Acknowledgments

A Inverting jet multiplicity

In section 3, we want to find the pseudo-inverse of $\widetilde{N}_{\text{jet}}(p_{T\text{cut}}, R)$ as a function of $p_{T\text{cut}}$ to get the function $\widetilde{p}_T(n, R)$. Here, we provide a computationally efficient way to perform this inverse. Consider a general function of the form

$$f(c) = \sum_{i=1}^{N} f_i \, \Theta(c_i - c), \tag{A.1}$$

where f_i and c_i are properties of the i-th particle, and c is some value of a cut. We wish to calculate the pseudo-inverse $c(f)$, which exists because $f(c)$ is a monotonically decreasing function of c. There is an ambiguity in the inverse because $f(c)$ is a step-wise function (with N steps), so there exists a range of values for c with the same value of f.

First, construct a list of length N with all of the values of c_i, keeping track of the corresponding value of f_i for each entry:

$$\{\{c_1, f_1\}, \{c_2, f_2\}, \ldots, \{c_N, f_N\}\}. \tag{A.2}$$

This list can be sorted from highest value of c_i to lowest value of c_i with computational scaling $N \log N$. Let i_j be the particle number i for the j-th highest element in the sorted list:

$$\{\{c_{i_1}, f_{i_1}\}, \{c_{i_2}, f_{i_2}\}, \ldots, \{c_{i_N}, f_{i_N}\}\}, \tag{A.3}$$

with $c_{i_1} > c_{i_2} > \ldots > c_{i_N}$. (If there are two value of c_i that are truly identical, one can add a small offset to arbitrarily break the degeneracy.)

From the sorted list, one then calculates the cumulative totals for the corresponding f_{i_j}:

$$\{\{c_{i_1}, f_{i_1}\}, \{c_{i_2}, f_{i_1} + f_{i_2}\}, \ldots, \{c_{i_N}, \sum_{j=1}^{N} f_{i_j}\}\}. \tag{A.4}$$

eq. (A.4) gives the function $f(c)$. To find the pseudo-inverse $c(f)$, one finds the value of J such that

$$\sum_{j=1}^{J} f_{i_j} < f < \sum_{j=1}^{J+1} f_{i_j}. \tag{A.5}$$

(For a sorted list, the computational cost of searching scales like $\log N$.) The pseudo-inverse $c(f)$ can then take on any value between c_{i_J} and $c_{i_{J+1}}$. For concreteness, we use

$$c(f) = c_{i_{J+1}}, \tag{A.6}$$

which makes sure that calculating $\widetilde{p}_T(n, R)$ with $0 < n_{\text{off}} < 1$ on infinitely narrow jets gives back the p_T of the n-th jet.

References

[1] S.D. Ellis, J. Huston, K. Hatakeyama, P. Loch and M. Tonnesmann, *Jets in hadron-hadron collisions*, *Prog. Part. Nucl. Phys.* **60** (2008) 484 [arXiv:0712.2447] [INSPIRE].

[2] G.P. Salam, *Towards Jetography*, *Eur. Phys. J.* **C 67** (2010) 637 [arXiv:0906.1833] [INSPIRE].

[3] M. Cacciari, G.P. Salam and G. Soyez, *The Anti-k(t) jet clustering algorithm*, *JHEP* **04** (2008) 063 [arXiv:0802.1189] [INSPIRE].

[4] M. Dasgupta and G.P. Salam, *Event shapes in* e^+e^- *annihilation and deep inelastic scattering*, *J. Phys.* **G 30** (2004) R143 [hep-ph/0312283] [INSPIRE].

[5] ALEPH collaboration, A. Heister et al., *Studies of QCD at* e^+e^- *centre-of-mass energies between 91-GeV and 209-GeV*, *Eur. Phys. J.* **C 35** (2004) 457 [INSPIRE].

[6] DELPHI collaboration, J. Abdallah et al., *A Study of the energy evolution of event shape distributions and their means with the DELPHI detector at LEP*, *Eur. Phys. J.* **C 29** (2003) 285 [hep-ex/0307048] [INSPIRE].

[7] L3 collaboration, P. Achard et al., *Studies of hadronic event structure in* e^+e^- *annihilation from 30-GeV to 209-GeV with the L3 detector*, *Phys. Rept.* **399** (2004) 71 [hep-ex/0406049] [INSPIRE].

[8] OPAL collaboration, G. Abbiendi et al., *Measurement of event shape distributions and moments in* $e^+e^- \to$ *hadrons at 91-GeV - 209-GeV and a determination of* α_s, *Eur. Phys. J.* **C 40** (2005) 287 [hep-ex/0503051] [INSPIRE].

[9] A. Banfi, G.P. Salam and G. Zanderighi, *Resummed event shapes at hadron - hadron colliders*, *JHEP* **08** (2004) 062 [hep-ph/0407287] [INSPIRE].

[10] A. Banfi, G.P. Salam and G. Zanderighi, *Phenomenology of event shapes at hadron colliders*, *JHEP* **06** (2010) 038 [arXiv:1001.4082] [INSPIRE].

[11] CDF collaboration, T. Aaltonen et al., *Measurement of Event Shapes in Proton-Antiproton Collisions at Center-of-Mass Energy 1.96 TeV*, *Phys. Rev.* **D 83** (2011) 112007 [arXiv:1103.5143] [INSPIRE].

[12] CMS collaboration, *First Measurement of Hadronic Event Shapes in pp Collisions at* $\sqrt{s} = 7$ *TeV*, *Phys. Lett.* **B 699** (2011) 48 [arXiv:1102.0068] [INSPIRE].

[13] ATLAS collaboration, *Measurement of event shapes at large momentum transfer with the ATLAS detector in pp collisions at* $\sqrt{s} = 7$ *TeV*, *Eur. Phys. J.* **C 72** (2012) 2211 [arXiv:1206.2135] [INSPIRE].

[14] A. Abdesselam et al., *Boosted objects: A Probe of beyond the Standard Model physics*, *Eur. Phys. J.* **C 71** (2011) 1661 [arXiv:1012.5412] [INSPIRE].

[15] A. Altheimer et al., *Jet Substructure at the Tevatron and LHC: New results, new tools, new benchmarks*, *J. Phys.* **G 39** (2012) 063001 [arXiv:1201.0008] [INSPIRE].

[16] J.M. Butterworth, A.R. Davison, M. Rubin and G.P. Salam, *Jet substructure as a new Higgs search channel at the LHC*, *Phys. Rev. Lett.* **100** (2008) 242001 [arXiv:0802.2470] [INSPIRE].

[17] S.D. Ellis, C.K. Vermilion and J.R. Walsh, *Techniques for improved heavy particle searches with jet substructure*, *Phys. Rev.* **D 80** (2009) 051501 [arXiv:0903.5081] [INSPIRE].

[18] S.D. Ellis, C.K. Vermilion and J.R. Walsh, *Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches*, *Phys. Rev.* **D 81** (2010) 094023 [arXiv:0912.0033] [INSPIRE].

[19] D. Krohn, J. Thaler and L.-T. Wang, *Jet Trimming*, *JHEP* **02** (2010) 084 [arXiv:0912.1342] [INSPIRE].

[20] M. Cacciari and G.P. Salam, *Pileup subtraction using jet areas*, *Phys. Lett.* **B 659** (2008) 119 [arXiv:0707.1378] [INSPIRE].

[21] M. Cacciari, G.P. Salam and G. Soyez, *The Catchment Area of Jets*, *JHEP* **04** (2008) 005 [arXiv:0802.1188] [INSPIRE].

[22] G. Soyez, G.P. Salam, J. Kim, S. Dutta and M. Cacciari, *Pileup subtraction for jet shapes*, *Phys. Rev. Lett.* **110** (2013) 162001 [arXiv:1211.2811] [INSPIRE].

[23] S.D. Ellis, J. Huston and M. Tonnesmann, *On building better cone jet algorithms*, *eConf* **C 010630** (2001) P513 [hep-ph/0111434] [INSPIRE].

[24] 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].

[25] J. Thaler and K. Van Tilburg, *Maximizing Boosted Top Identification by Minimizing N-subjettiness*, *JHEP* **02** (2012) 093 [arXiv:1108.2701] [INSPIRE].

[26] C.F. Berger et al., *Snowmass 2001: Jet energy flow project*, *eConf* **C 010630** (2001) P512 [hep-ph/0202207] [INSPIRE].

[27] L. Angelini et al., *Jet analysis by deterministic annealing*, *Phys. Lett.* **B 545** (2002) 315 [hep-ph/0207032] [INSPIRE].

[28] L. Angelini et al., *Deterministic annealing as a jet clustering algorithm in hadronic collisions*, *Phys. Lett.* **B 601** (2004) 56 [hep-ph/0407214] [INSPIRE].

[29] D.Y. Grigoriev, E. Jankowski and F.V. Tkachov, *Towards a standard jet definition*, *Phys. Rev. Lett.* **91** (2003) 061801 [hep-ph/0301185] [INSPIRE].

[30] D.Y. Grigoriev, E. Jankowski and F.V. Tkachov, *Optimal jet finder*, *Comput. Phys. Commun.* **155** (2003) 42 [hep-ph/0301226] [INSPIRE].

[31] S. Chekanov, *A New jet algorithm based on the k-means clustering for the reconstruction of heavy states from jets*, *Eur. Phys. J.* **C 47** (2006) 611 [hep-ph/0512027] [INSPIRE].

[32] Y.-S. Lai and B.A. Cole, *Jet reconstruction in hadronic collisions by Gaussian filtering*, arXiv:0806.1499 [INSPIRE].

[33] I. Volobouev, *FFTJet: A Package for Multiresolution Particle Jet Reconstruction in the Fourier Domain*, arXiv:0907.0270 [INSPIRE].

[34] S.D. Ellis, A. Hornig, T.S. Roy, D. Krohn and M.D. Schwartz, *Qjets: A Non-Deterministic Approach to Tree-Based Jet Substructure*, *Phys. Rev. Lett.* **108** (2012) 182003 [arXiv:1201.1914] [INSPIRE].

[35] D. Kahawala, D. Krohn and M.D. Schwartz, *Jet Sampling: Improving Event Reconstruction through Multiple Interpretations*, *JHEP* **06** (2013) 006 [arXiv:1304.2394] [INSPIRE].

[36] A.J. Larkoski, G.P. Salam and J. Thaler, *Energy Correlation Functions for Jet Substructure*, *JHEP* **06** (2013) 108 [arXiv:1305.0007] [INSPIRE].

[37] A. Hook, E. Izaguirre, M. Lisanti and J.G. Wacker, *High Multiplicity Searches at the LHC Using Jet Masses*, *Phys. Rev.* **D 85** (2012) 055029 [arXiv:1202.0558] [INSPIRE].

[38] T. Cohen, E. Izaguirre, M. Lisanti and H.K. Lou, *Jet Substructure by Accident*, *JHEP* **03** (2013) 161 [arXiv:1212.1456] [INSPIRE].

[39] S. El Hedri, A. Hook, M. Jankowiak and J.G. Wacker, *Learning How to Count: A High Multiplicity Search for the LHC*, *JHEP* **08** (2013) 136 [arXiv:1302.1870] [INSPIRE].

[40] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer and T. Stelzer, *MadGraph 5: Going Beyond*, *JHEP* **06** (2011) 128 [arXiv:1106.0522] [INSPIRE].

[41] T. Sjöstrand, S. Mrenna and P.Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852 [arXiv:0710.3820] [INSPIRE].

[42] M. Cacciari, G.P. Salam and G. Soyez, *FastJet User Manual*, *Eur. Phys. J.* **C 72** (2012) 1896 [arXiv:1111.6097] [INSPIRE].

[43] B. Knuteson, M. Strovink, B. Olivier, U. Bassler, F. Fleuret, and G. Bernardi, *The missing transverse energy resolution of an event*, *DØ Note* **3629** (1999).

[44] B. Nachman and C.G. Lester, *Significance Variables*, *Phys. Rev.* **D 88** (2013) 075013 [arXiv:1303.7009] [INSPIRE].

[45] Y.-T. Chien, *Telescoping Jets: Multiple Event Interpretations with Multiple R's*, arXiv:1304.5240 [INSPIRE].

[46] A.J. Larkoski, D. Neill and J. Thaler, *Jet Shapes with the Broadening Axis*, arXiv:1401.2158 [INSPIRE].

[47] Y.L. Dokshitzer, G.D. Leder, S. Moretti and B.R. Webber, *Better jet clustering algorithms*, *JHEP* **08** (1997) 001 [hep-ph/9707323] [INSPIRE].

[48] M. Wobisch and T. Wengler, *Hadronization corrections to jet cross-sections in deep inelastic scattering*, hep-ph/9907280 [INSPIRE].

[49] M. Wobisch, *Measurement and QCD analysis of jet cross-sections in deep inelastic positron proton collisions at* $\sqrt{s} = 300\,GeV$, DESY-THESIS-2000-049 (2000).

[50] 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].

[51] S.D. Ellis and D.E. Soper, *Successive combination jet algorithm for hadron collisions*, *Phys. Rev.* **D 48** (1993) 3160 [hep-ph/9305266] [INSPIRE].

[52] G. Corcella et al., *HERWIG 6.5 release note*, hep-ph/0210213 [INSPIRE].

[53] J.M. Butterworth, J.R. Forshaw and M.H. Seymour, *Multiparton interactions in photoproduction at HERA*, *Z. Phys.* **C 72** (1996) 637 [hep-ph/9601371] [INSPIRE].

[54] ATLAS collaboration, *ATLAS Monte Carlo tunes for MC09*, ATL-PHYS-PUB-2010-002 (2010).

[55] CMS collaboration, *Studies of jet mass in dijet and W/Z + jet events*, *JHEP* **05** (2013) 090 [arXiv:1303.4811] [INSPIRE].

[56] ATLAS collaboration, *Performance of jet substructure techniques for large-R jets in proton-proton collisions at* $\sqrt{s} = 7\,TeV$ *using the ATLAS detector*, *JHEP* **09** (2013) 076 [arXiv:1306.4945] [INSPIRE].

[57] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N-subjettiness*, *JHEP* **03** (2011) 015 [`arXiv:1011.2268`] [ɪɴSPIRE].

[58] M. Dasgupta, A. Fregoso, S. Marzani and G.P. Salam, *Towards an understanding of jet substructure*, *JHEP* **09** (2013) 029 [`arXiv:1307.0007`] [ɪɴSPIRE].

[59] M. Dasgupta, A. Fregoso, S. Marzani and A. Powling, *Jet substructure with analytical methods*, *Eur. Phys. J.* **C 73** (2013) 2623 [`arXiv:1307.0013`] [ɪɴSPIRE].