MIT–CTP xyz

Systematic Quark/Gluon Identification with Ratios of Likelihoods

Eric M. Metodiev,[a] **Ian Moult,**[b,c] **Benjamin Nachman,**[d] **and Stefan Prestel**[e]

[a] *Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

[b] *Berkeley Center for Theoretical Physics, University of California, Berkeley, CA 94720, USA*

[c] *Theoretical Physics Group, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*

[d] *Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*

[e] *Department of Astronomy and Theoretical Physics, Lund University, S-223 62 Lund, Sweden*

E-mail: metodiev@mit.edu, ianmoult@lbl.gov, bpnachman@lbl.gov, stefan.prestel@thep.lu.se

Abstract: The radiation pattern within high energy jets contains a wealth of information about the originating particle. Numerous techniques have been developed based on qualitative features of jet substructure, with the resulting observables calculated to high perturbative accuracy in many cases. At the same time, there have been many attempts to fully exploit the jet radiation pattern using tools from statistics and machine learning. We propose a new approach that combines a deep analytic understanding of jet substructure with the optimality promised by machine learning and statistics. After specifying an approximation to the full emission phase space, the optimal observable is calculated exactly. This procedure is demonstrated for the case of quark and gluons jets in the leading logarithmic and modified leading logarithmic approximation, where we prove that linear combinations of weighted multiplicity is the optimal observable. In addition to providing a new and powerful framework for systematically improving jet substructure observables, the specific quark versus gluon jet tagging observable is demonstrated to improve upon the performance of existing substructure observable approaches.

Contents

1 Introduction

Collimated sprays of particles initiated by high energy quarks and gluons are a ubiquitous feature of collisions at the Large Hadron Collider (LHC). The radiation pattern within these *jets* contains a wealth of information about the originating particle. Since the start of the LHC, there has been a revolution in jet substructure that has resulted in many tools developed to tag the origin of a jet [1–5]. These tools have significantly extended the reach of searches for phenomena beyond the Standard Model (SM) as well precision measurements in the SM.

The most successful and widely used techniques apply to situations where the radiation pattern within the jet exhibits hierarchical structures. This occurs in many cases of relevance, such as n-prong tagging for identifying hadronically decaying bosons or top quarks. In such cases, power counting [6] can be used to identify optimal observables as an expansion in the hierarchical parameter. For instance, the ATLAS collaboration two-prong tagger default D_2 [6], and the CMS collaboration two-prong tagger N_2 [7] were constructed by power counting. This approach has the advantage that it makes clear the physics identified by the observable. Once motivated and defined, many of the state-of-the-art observables have been calculated with high formal precision [8]. However, this approach fails when there does not exist a clear expansion parameter. One such case is quark/gluon jet tagging, which, despite having significant potential applications at the LHC, has achieved limited theory success due to the lack of a systematic approach to the construction of observables.

Modern machine learning has revolutionized the way classification techniques are optimized in many fields, including high energy physics [9–11]. Deep neural networks trained on a low-level representation of the full radiation pattern inside jets have been found to outperform observables simple physics motivated features. For a review of machine learning techniques in jet physics, see Ref. [12]. A key challenge with these machine learning tools is that they are intractable to understand physically. This has motivated a series of proposals to use physically-motivated bases of high-level features [13–16]. These proposals guarantee important properties for classification, such as completeness or linearity, but nonetheless still may not result in sufficient analytic control over the final classifier to be understood simply. While a first-principles understanding of a classifier is not strictly required for its performance, it can add robustness when the technique is applied beyond its training domain and furthermore can lead to a deeper understanding of the underlying physics.

In this paper, we introduce a new approach that combines the benefits of both first-principles techniques and statistical machine learning algorithms. First, an approximation of the full emission phase space is constructed. A well-known theorem due to Neyman and Pearson [17] states that the optimal classifier for distinguishing two classes is a threshold on the ratio of the class probability densities (likelihood ratio). With a description of the full emission phase space for two classes, one can take the full ratio and reorganize terms to arrive at an analytic prediction for the optimal classifier. In this sense it is similar to shower deconstruction [18]. This procedure is systematically improvable as the approximation to the full phase space is improved. Unlike the power counting approach, this approach does not rely on the existence of a small parameter. Compared to standard machine learning techniques, this approach makes clear what physics is being exploited and what additional information must be included to achieve increased performance, since it is based on understanding the analytic approximation used to derive the likelihood.

We will apply this approach to the case of quark/gluon discrimination, which has so far resisted organizing principles. We will prove that in the approximation of eikonal splitting functions, the optimal observable is multiplicity. We then extend this to modified leading logarithmic accuracy, and prove that the optimal observable is a linear combination of weighted multiplicities. Using this, we introduce a new quark/gluon discriminant, which we find in parton shower simulations outperforms standard observables.

This paper is organized as follows. In Sec. 2 we provide a review of the structure of quark and gluon jets. Section 3 describes an explicit example of an optimal classifier for quark versus gluon jet tagging with leading logarithmic (LL) and then at modified leading logarithmic (MLL) precision. The performance of the resulting tagger is shown in Sec. 4 using parton shower (PS) Monte Carlo (MC) simulations. Finally, in Sec. 5 we discuss how this approach could be extended to other tagging problems before concluding.

2 The Structure of Quark and Gluon Jets

We begin by analyzing the perturbative structure of a jet sourced by a hard parton radiating massless gauge bosons (the radiation of fermions will be addressed later). Here, we will restrict ourselves to LL or MLL, where we can use the splitting functions. While this is very well understood, here we present what is perhaps a different perspective that proves particularly useful for understanding quark vs. gluon discrimination.[1]

Although we will be focused on the physically realized case of QCD, it is interesting to view the problem more generally. We consider the splitting function for the radiation of a massless gauge boson

$$= 4C_i \frac{\alpha}{4\pi}\left(\frac{x}{1-x} + (1-x)g_i(x)\right)\,, \qquad \begin{cases} g_\phi(x) = 0 \\ g_\lambda(x) = \frac{1}{2} \\ g_V(x) = x + \frac{1}{x} \end{cases}\,. \tag{2.1}$$

(write it in the symmetrized form for the gluon –ijm) (discuss relation to spin ijm
and classical limit and mass/ EECs –ijm) (view it like lbk for the discriminator ijm
on the emission –ijm) (also motivates measuring EEC etc in this region of spin ijm
to tune quark vs gluon. –ijm) (clarifies the physics that is probed in quark vs ijm
gluon discrimination or is most important –ijm) (advertise these parton shower ijm
modfications for tests of ML and what they are learning. would be itneresting ijm
to train on with the color equal, to just focus on spin, etc –ijm) (allows to focus ijm
more directly on physical effects that are being probed. –ijm) The function $g_i(x)$ satisfies the condition that it has a regular Taylor series expansion about $x = 1$. Here we see that the splitting probability depends on the parton species through two factors: the Casimir C_i which describes the color charge of the particle, and the function $g_i(x)$.

In the $x \to 1$ limit, in which the emitted radiation is soft, and which is enhanced by the soft singularity, the splitting function exhibits a universal form, depending only on the color factor, C_i, which to be explicit, are

$$C_F = \frac{N^2-1}{2N} \to \frac{4}{3}\,, \qquad C_A = N \to 3\,. \tag{2.2}$$

The coefficient appearing in the splitting function is simply the one-loop cusp anomalous dimension. The result in this limit is of course simply the classical result for radiation in the eikonal limit [22].[2] The classical result is (as it must be) independent of the spin of the emitting parton. In this limit, there is no information contained in the structure of the radiation within the jet. The only information is contained in the quantity of radiation (multiplicity). We will prove this rigorously in Sec. 3. The second factor $g_i(x)$ in the splitting function depends on the nature of the emitting parton. Its contribution to the splitting function is suppressed by two powers of $(1-x)$. This is guaranteed by the Low-Burnett-Kroll (LBK) theorem [23–25] describing the soft limits of gauge theories.[3] As expected, $g_i(x)$ vanishes for a scalar, which has no structure beyond its color.

Having written the splitting functions in the form above in Eq. (2.1) makes abundantly clear the difficulty for quark gluon jet discrimination: the structure of a jet is dominated by

[1]This perspective has been beautifully advocated for the simplification of the higher order splitting functions in Refs. [19–21]. **(phrasing issue here –emm)**

[2]We wish to emphasize that the leading double logarithmic Sudakov is classical.

[3]In this case, the LBK theorem is applied after taking the collinear limit.

its classical result, and therefore depends only on the color charge. Futhermore, we note an amusing difference. Expanding the function g_i in $(1-x)$, we have

$$g_\phi = 0\,, \qquad g_\lambda = \frac{1}{2}\,, \qquad g_V = \frac{1}{2} + \frac{1}{4}(1-x)^2 + \cdots\,. \tag{2.3}$$

So the difference between a quark and a gluon is in fact more suppressed than it needs to be.

We note that in QCD, there is an additional distinction between quarks and gluons, namely that gluons can split into a $q\bar{q}$ pair,

$$P_{g\to q\bar{q}}(z) = \frac{n_f T_R}{2C_A}[z^2 + (1-z)^2]. \tag{2.4}$$

This splitting does not obey the LBK theorem and is down only by a single power of z. See Refs. [26–30] for a discussion. **ijm**

(can one argue something about the behavior of the fragmentation function –ijm)

(may want a subsection here on the qg "alphabet" and other q/g approaches (or lack thereof). –emm) **emm**

3 Quark and Gluon Jet Likelihoods

In this section, we introduce an approach which allows us to prove the optimality of quark/gluon discriminants. This approach consists of analyzing the likelihood function of the jet radiation at a given accuracy. The Neyman-Pearson Lemma [17] then states that the optimal classifier for distinguishing two classes is a threshold on the ratio of the class probability densities (likelihood ratio). We believe, however, that this approach is more general, and it would be interesting to apply it to other cases in jet substructure. **ijm**

(make some blab about power counting and having a small parameter to expand in –ijm)[6][7][31]

This approach allows one to work in some approximation for the likelihood distribution of quarks and gluons (e.g. eikonal approximation, independent emission, etc.) and then analyze this likelihood using machine learning or information theoretic techniques to prove results about the optimal observables. These optimality results will then apply to any observables that can be computed at a given accuracy from this likelihood. Since we will always work in some approximation for the likelihood, one can always come up with observables that can evade the proven results, if they cannot be correctly computed from this approximation. However, we view this as a virtue, as it shows how the proven theorems can be evaded.

3.1 Analysis in the Eikonal Limit

We begin by considering the optimal quark/gluon discriminant in the eikonal limit. We define the eikonal limit as independent emissions with the universal part of the splitting functions

$$dP_{i\to ig}(z,\theta) = \frac{2\alpha_s C_i}{\pi}\frac{dz}{z}\frac{d\theta}{\theta}, \tag{3.1}$$

where $i \in \{q/g\}$.

Before proceeding, we would like to clarify a number of issues related to our referring to this as the eikonal, or classical eikonal limit. Another name which could be used is the LL limit. We wish to avoid this language, since it is best defined when their is an observable of interest. While it is of course true that for Sudakov-type observables, the above will generate the leading logarithms, it is straightforward to identify IRC safe observables (even those that probe only two particle correlations) for which the leading logarithm will not be produced by the above approximated splitting function.

Since we have shown that splitting function is identical in this limit for all partons up to the color factor in Sec. 2, we should be able to prove that multiplicity is the optimal observable. To prove this, we consider the quark/gluon likelihood ratio $L_{q/g}$ for radiating a collection of gluons[4] with kinematics $\{(z_n, \theta_n)\}_{n=1}^{M}$

$$\ln L_{q/g}^{\mathrm{LL}} = \sum_{n=1}^{M} \ln \left(\frac{dP_{q\to qg}(z_n, \theta_n)}{dP_{g\to gg}(z_n, \theta_n)} \right) = \ln \frac{C_F}{C_A} \sum_{n=1}^{M} 1 = M \ln \frac{C_F}{C_A}. \tag{3.2}$$

Here M is simply the multiplicity, and so by the Neyman-Pearson Lemma [17], the optimal observable at this level of accuracy is the multiplicity. Experimentally the multiplicity is indeed found to perform well as a discriminant, however, we are now able to prove this. **(worth remarking on hadron (expt.) vs. parton (here) multiplicity? –emm)** **emm**

In hindsight, this is completely obvious, and follows simply from the universal nature of soft emissions in gauge theory, or in other words, the classical structure of the jet. This behavior is true for a very wide class of observables, including standard Sudakov observables, Sudakov observables that have been soft dropped [33], groomed multiplicity [34]. We also note that the above result holds regardless of whether the coupling is taken to be running or not, since it cancels out of the ratio in Eq. (3.2).

We therefore state our first result:

Theorem I: For an observable whose LL result can be computed using the eikonal splitting functions of Eq. (3.1) in the independent emission approximation (with or without running coupling), this LL result can not achieve better quark/gluon discrimination than multiplicity.

In addition to showing that multiplicity is optimal, we can in fact derive its probability distribution. Since the sum of Poissonians is Poissonian, so we can immediately read off

$$p_i^{\mathrm{LL}}(M) = \mathrm{Pois}\left[\int \frac{dz}{z} \int \frac{d\theta}{\theta} \frac{2\alpha_s C_i}{\pi} \Theta(z, \theta) \right] . \tag{3.3}$$

[4]Here we use the well known fact that there is a probabilistic interpretation for the twist two splitting functions. This, however, should not be taken for granted, and indeed such a probabilistic interpretation fails at higher twist [32].

Poisson observables were emphasized in [34]

There is a simple way of visualizing this proof in the Lund plane, shown in Fig. 1. In the eikonal limit of the splitting functions, emissions are uniformly distributed in the Lund plane. We can therefore tesselate some perturbative region of the Lund plane using triangles. The probability for n emissions in a given triangle with area Δ is then Poisson distributed according to

$$\Pr(n_i = n) = \frac{\lambda^n e^{-\lambda}}{n!}\,, \qquad \lambda = \frac{2\alpha_s C_i \Delta}{\pi}\,. \tag{3.4}$$

Due to the assumption of independent emissions, we then have that the radiation counts throughout the emission plane are distributed according to

$$\Pr_i(n_1, n_2, \cdots, n_N) = \prod_{j=1}^{N} \frac{\lambda_i^{n_j} e^{-\lambda_i}}{n_j!}\,. \tag{3.5}$$

The optimal quark/gluon discriminant is then

$$\ln \frac{\Pr_q(n_1, \cdots, n_N)}{\Pr_g(n_1, \cdots, n_N)} = \ln \frac{C_F}{C_A} \sum_{j=1}^{N} n_j + \text{const.} \tag{3.6}$$

We can take the $\Delta \to 0$ limit to find again that the optimal quark gluon discriminant is simply the multiplicity.

This version of the proof makes clear that we can apply a cutoff and only consider the likelihood function in the perturbative regime. **ijm**

(discussion about NP and soft drop mimics LL structure of the shower. –ijm) **ijm**

(This also makes clear that one can apply an NP cut, and it still holds –ijm)

3.2 Beyond Eikonal

The radiation phase space from the previous section can be systematically improved. Observables computed with modified leading logarithmic accuracy should include the effects of the running coupling as well as subleading terms in the splitting function. In particular, the radiation phase space is given by

$$dP_{i\to ij}(z, \theta) = \frac{2\alpha_s(z\theta p_T)\, C_i}{\pi} p_{i\to ij}(z) dz\, \frac{d\theta}{\theta}, \tag{3.7}$$

where p_{ij} are the QCD splitting functions. In general, at this beyond-Eikonal (BE) order, there are can be flavor changes between quarks and gluons, since $\int_{1/2}^{1} dz p_{q\to qg}(z) > 0$ and a finite part of the gluon splitting function contains a contribution from $g \to q\bar{q}$. Ignoring flavor changing, the formalism from the previous section still holds, only now the optimal observable is more complicated than simply multiplicity. In particular:

$$\text{JL}^{\text{BE'}} \equiv \ln L_{q/g}^{\text{BE'}} = \sum_{n=1}^{M} \ln \frac{dP_{q\to qg}(z_n, \theta_n)}{dP_{g\to gg}(z_n, \theta_n)} = \sum_{n=1}^{M} \left[\ln \frac{C_F}{C_A} + \ln \frac{p_{q\to qg}(z_n)}{p_{g\to gg}(z_n)} \right]. \tag{3.8}$$

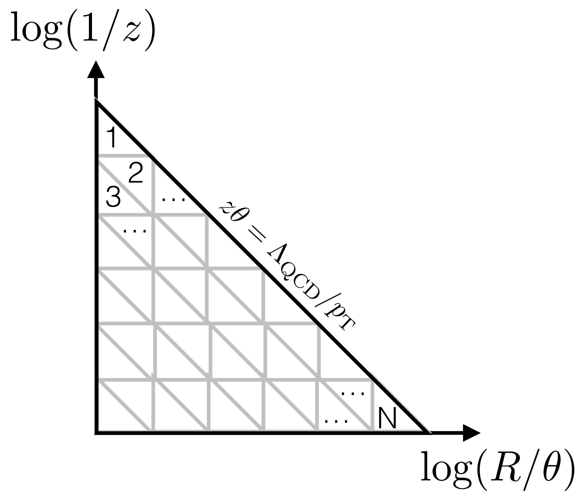

Figure 1. A tiled version of the Lund plane where each of the N triangles has the same area. The non-perturbative regime is defined by $z\theta > \Lambda_{\mathrm{QCD}}/p_{\mathrm{T}}$.

where BE' denotes the beyond Eikonal approximation for the radiation phase space, but ignoring flavor changing. By inserting the corresponding the following splitting functions[5]

$$p_{q\to qg}(z) = \frac{1+(1-z)^2}{2z}, \tag{3.9}$$

$$p_{g\to gg}(z) = \frac{1-z}{z} + \frac{1}{2}z(1-z), \tag{3.10}$$

the optimal observable is

$$\mathrm{JL}^{\mathrm{BE'}} = \sum_{n=1}^{M}\left[\ln\frac{C_F}{C_A} + \ln\frac{1+(1-z_n)^2}{(1-z_n)(2+z_n^2)}\right] \approx M\ln\frac{C_F}{C_A} + \frac{1}{2}\sum_{n=1}^{M} z_n^3 + \mathcal{O}(z_n^4). \tag{3.11}$$

Equation 3.11 shows that the multiplicity is still nearly optimal when including the non-flavor changing corrections to the radiation phase space. A full BE calculation requires a proper accounting of flavor changing. This is accomplished in two steps. As a first step, we consider a new basis of particles in which there no flavor changes. For a quark or gluon at a given step in the jet formation, the flavor transition matrix is given by

$$\begin{pmatrix} q \\ g \end{pmatrix} \mapsto \begin{pmatrix} p(q|q) & p(g|q) \\ p(q|g) & p(g|g) \end{pmatrix} \begin{pmatrix} q \\ g \end{pmatrix}, \tag{3.12}$$

where

[5]The $p_{q\to qg}$ still includes the flavor change, but this fact is ignored in the BE' approximation.

$$
\begin{aligned}
p(q|q) &= \int_0^{1/2} p_{q\to qg}(z)dz = \frac{1}{2}\int_0^{1/2} dz \left(\frac{1+(1-z)^2}{(z)_+}\right) = \frac{1}{2}\int_0^{1/2} dz \frac{(1-z)^2}{z} \\
p(g|q) &= \int_{1/2}^1 p_{q\to qg}(z)dz = \int_{1/2}^1 dz \frac{1+(1-z)^2}{2z} = \\
p(g|g) &= \int_0^1 p_{g\to gg}(z)dz = \int_0^1 dz \frac{1-z}{z} + \frac{1}{2}z(1-z) = \\
p(q|g) &= \int_0^1 p_{g\to qq}(z)dz = \int_0^1 \frac{n_f T_R}{2C_A}[z^2+(1-z)^2] =
\end{aligned}
\tag{3.13}
$$

$$\mathrm{JL}^{\mathrm{MLL}} \equiv \ln L_{q/g}^{\mathrm{MLL}} = \sum_{n=1}^{M} \ln \frac{dP_{q\to qg}(z_n,\theta_n)}{dP_{g\to gg}(z_n,\theta_n)} = \sum_{n=1}^{M} \left[\ln \frac{C_F}{C_A} + \ln \frac{P_q(z_n)}{P_g(z_n)}\right]. \tag{3.14}$$

Can see at this point that the optimal observable is a linear combination of the weighted multiplicities considered in

The quark and gluon splitting functions, when summed over final states, become:

$$P_q(z) = \frac{1+(1-z)^2}{2z}, \tag{3.15}$$

$$P_g(z) = \frac{1-z}{z} + \frac{1}{2}z(1-z) + \frac{n_f T_R}{2C_A}[z^2+(1-z)^2]. \tag{3.16}$$

where the $z \leftrightarrow 1-z$ symmetry is used put the gluon singularity entirely at $z \to 0$.

The optimal observable at MLL from Eq. (3.14) is independent of the running of the strong coupling constant. The splitting functions influence the optimal observable via the log likelihood ratio $\ln P_q(z)/P_g(z)$ Restore perturbative multiplicity as the optimal observable in the soft $z_n \to 0$ limit as quarks and gluons have the same soft singularity so the log splitting function ratio approaches zero.

$$\mathrm{JL}^{\mathrm{MLL}} = \sum_{n=1}^{M} \left[\ln \frac{C_F}{C_A} + \ln \frac{1+(1-z_n)^2}{(1-z_n)(2+z_n^2) + \frac{n_f T_R}{C_A} z_n (z_n^2 + (1-z_n)^2)}\right]. \tag{3.17}$$

$$\mathrm{JL}^{\mathrm{MLL}} = \sum_{n=1}^{M} \left(\ln \frac{C_F}{C_A} - \frac{n_f T_R}{2C_A} z_n + \frac{n_f T_R (4C_A + n_F T_R)}{8C_A^2} z_n^2 + \cdots\right) \tag{3.18}$$

$$= \ln \frac{C_F}{C_A} n^{(\kappa=0)} - \frac{n_f T_R}{2C_A} n^{(\kappa=1)} + \frac{n_f T_R (4C_A + n_F T_R)}{8C_A^2} n^{(\kappa=2)} + \cdots \tag{3.19}$$

emm

(Should redo without $g \to q\bar{q}$, but then leading term is $\mathcal{O}(z_n^3)$. Coefficients should become much simpler. –emm)

We can therefore extend to our second result:

Theorem II: For an observable whose LL or MLL result can be computed using splitting functions in the independent emission approximation, this result can not achieve better quark/gluon discrimination than a linear combination of weighted multiplicities.

3.3 Perturbative multiplicity and parton showers

Parton showers are numerical tools that generate multi-parton phase-space points on which observables can be measured. The shower proceeds iteratively, producing its own initial conditions for subsequent evolution. This procedure requires physically sensible intermediate states, i.e. on-shell phase-space points during its ordered evolution. Parton showers recover leading-logarithmic results for observables that are sufficiently similar to their ordering variable, and for which a strong ordering of emissions is guaranteed. The latter aspect and the desire for physically sensible intermediate states can sometimes lead to conflicting needs, especially for obserables depending/relying on states containing several partons that cannot be interpreted as "ordered", and that are sufficiently different from the ordering variable. In such observables, kinematic recoil effects can no longer be neglected, since different schemes to ensure momentum conservation become distinguishable. Perturbative multiplicity is such an observable.

One particularly common scheme to enforce momentum conservation arises from the leading-color approximation of QCD, in which color-dipoles radiate soft gluons coherently. This dipole picture suggests to produce physical $(n+1)$-parton states from n-parton states by distributing the momenta of two originator partons over three new partons, thus allowing momentum-conservation and on-shell conditions throughout. The originator pair is, in most cases, determined by color connections. Quarks and antiquarks will thus contribute to the evolution of a single dipole, while gluons will participate in two dipoles.

In parton showers, quarks and gluons yield different radiation patterns (and thus perturbative multiplicities) because of differing color factors, and because of different phase-space dependence of hard-collinear contributions. Moreover, the effect of kinematic recoil is handled differently for both. This is easily seen by comparing the evolution of $Z \to q\bar{q}$ to $h \to g_1\bar{g}_2$. In the former, the original color connection between q and $\bar{q}$ is broken after the first gluon emission, while for the latter, only one of the two original connections is severed by emitting the first gluon. If momentum is re-distributed locally within color-connected dipoles (as is commonly the case in parton showers, including the Dire shower employed below), no subsequent gluon emission from the q will influence the $\bar{q}$ momenta, while emissions from g_1 can continue to influence the g_2 momentum, until the second original connection is severed. This simple choice of recoil mechanism has the advantage that it makes the action of the parton shower trivially invertible, and allows stopping and restarting the shower at will without consequences. These factors are crucial for matching and merging methods to improve the overall event generator fidelity. However, such a recoil scheme clearly complicates the subleading-color bahavior of the parton shower, and can lead to unphysical artifacts in quark-vs-gluon discrimination.

To assess the impact of recoil on perturbative multiplicities, we introduce a simple extension of the local recoil strategy of Dire, with the aim to obtain an identical recoil handling for quarks and gluons. For this, we introduce a "backbone" dipole for gluon evolution, which is never allowed to split. The backbone dipole is defined by the left-over original color-connection after the first gluon emission in the shower. If the original state consisted of a $q\bar{q}$ pair, no backbone is present. The emission rate from the backbone dipole is forced to vanish. If a parton contributes both to the backbone and to another dipole evolution (i.e. is a gluon), then the rate of emissions from the parton in the latter dipole is increased to ensure that the overall $g \to gg$ rate is recovered. This guarantees that recoil effects in $Z \to q\bar{q}$ and $h \to g_1\bar{g}_2$ are treated completely identically, by effectively treating the recoil of emissions from gluons connected to the backbone identical to emission from (anti)quarks.

3.4 Bla: A new observable for Quark vs. Gluon Tagging

This extends to a continuous variable the counting variable of [34].

3.5 Discussion and Extensions

can discuss how for these observables, higher orders in α_s the same will mostly apply in soft limit. Casimir scaling holds to at least three loops. Violated at four loops[35][36].[37] [38] these should be small higher order corrections.

Suggests a different direction one should look directly for observables that cannot be computed in these approximations.

study of observables with more interesting logarithmic structures.

(need to mention more about shower deconstruction and Lund images –emm) emm

(need to be explicit about neglecting e.g. flavor changing, $g \to q\bar{q}$, momentum conservation, ... –emm) emm

4 Monte Carlo Study

Using our Pythia + DIRE setup, we can do some tests:

1. Turn off $g \to qq$ and turn off hadronization (will revise these later).
   (a) LL and set $C_A = C_F$. Verify that we can't learn anything useful for q/g.
   (b) LL but real C_A and C_F.
   (c) Same, but MLL (LO splitting functions). For later - this is the 'quantum' component.
   (d) Now, tweak the finite parts of LO splitting functions and show that we can predict the optimal observable (even more fun to make the second term dominate over the first so that the best observable is not just multiplicity + ϵ; in particular, can set $C_A = C_F$ but have different subleading terms).

   emm

   (for (d): $P_g(z) = 1/z$ and $P_q(z) = c_0/z \times \exp(c_1 z + c_2 z^2)$. This choice is motivated by looking ahead to the optimal observable, which would be:

$$\sum_i \ln \frac{P_q}{P_g} = \sum_i \left(\ln \frac{c_0 C_F}{C_A} + c_1 z_i + c_2 z_i^2 \right) = n^{(0)} \ln \frac{c_0 C_F}{C_A} + c_1 n^{(1)} + c_2 n^{(2)} \qquad (4.1)$$

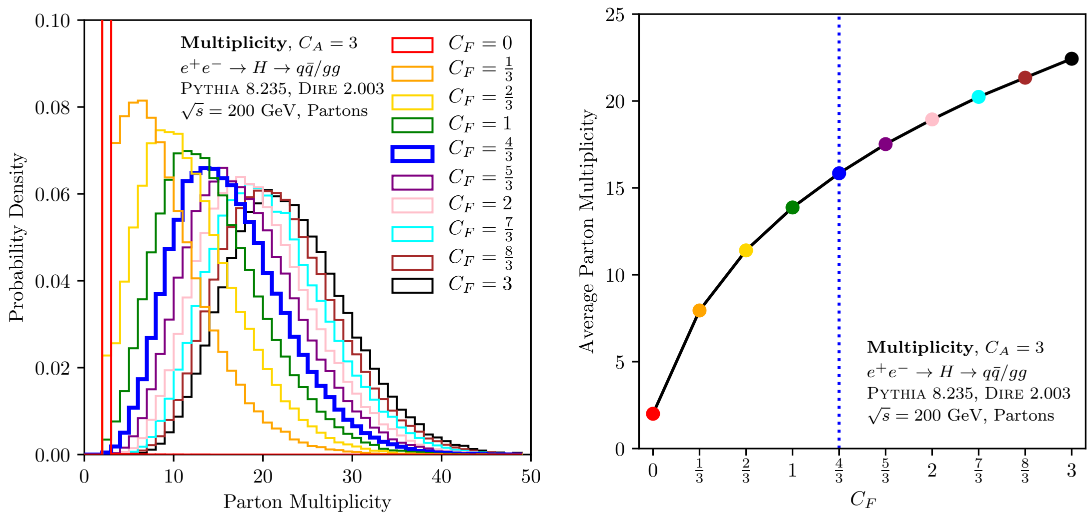

Figure 2. The (left) distribution of parton multiplicity and (right) average parton multiplicity in jets with $C_A = 3$ as the value of C_F is swept from $C_F = 0$ to $C_F = 3$. No additional partons are emitted from the originating two for $C_F = 0$. Here, physical quarks correspond to the case of $C_F = 4/3$ and physical gluons correspond to the case of $C_F = 3$.

where $n^{(\kappa)} = \sum_i z_i^\kappa$ is the energy-weighted multiplicity. So ideally we would have three tunable parameters: c_0, c_1, and c_2. –emm)

2. For one configuration above, check impact of hadronization and $g \to qq$.

Some notes: maybe we do this in e^+e^- so we don't have to worry about q/g definitions, etc. or pileup or UE.

5 Conclusions

In this paper we have presented a new approach to understanding jet substructure observables, analyzing their likelihood function at a fixed accuracy. This approach combines the standard approach from ML, with the use of analytic insights. We have applied this to the case of quark/gluon discrimination. Using an understanding of the universal nature of radiation in gauge theories, we have been able to prove a number of interesting results, which greatly clarify quark/gluon discrimination. In particular, we have proven that for observables that can be computed using eikonal splitting functions, their LL result cannot beat multiplicity. We extend this result to MLL accuracy, and showed that in this case on finds that the optimal observable is a linear combination of weighted multiplicities.

Based on this result, we introduced a new quark/gluon discriminant. This observable is a continuous observable which outperforms the soft drop multiplicity. We studied its performance in Monte Carlo simulation...

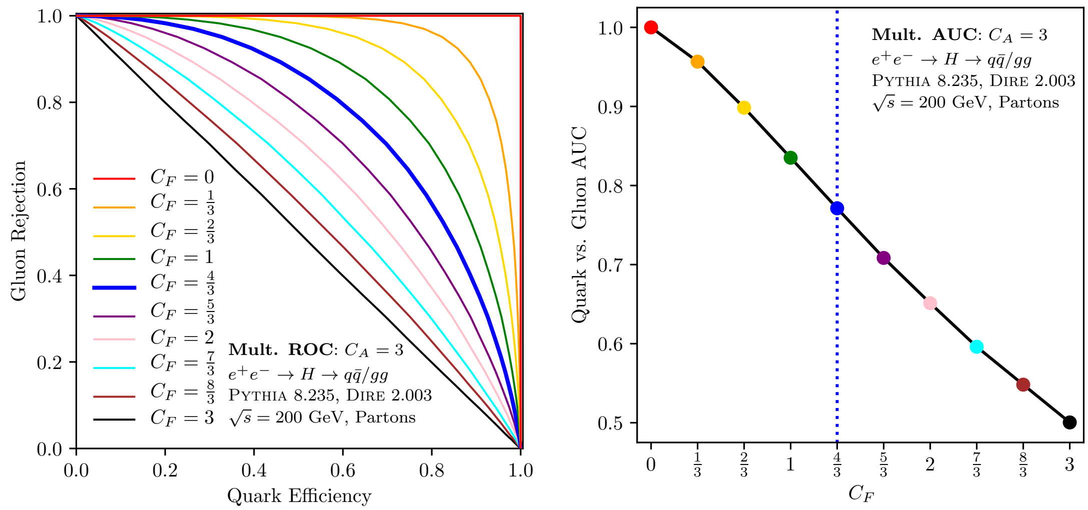

Figure 3. The quark versus gluon parton multiplicity (left) ROC curves and (right) AUC values with $C_A = 3$ as the value of C_F is swept from $C_F = 0$ to $C_F = 3$, with the physical value of $C_F = 4/3$ highlighted. Quarks and gluons are perfectly distinguishable when $C_F = 0$ and are indistinguishable with $C_F = C_A$, with a nearly linear behavior of the AUC at intermediate values for this energy scale.

A key feature of our approach is that the since the derivation of the optimal observable is based on an analytic likelihood function, it is clear what approximations have been made in the derivation. This suggests how observables can be designed that evade our theorems. It is also important to emphasize that although we have proven that a sum over weighted multiplicities is the optimal observable to modified LL accuracy, this observable is unique. It is an interesting question to find observables with desirable analytic properties which are good quark/gluon discriminants.

Our approach greatly clarifies

Acknowledgments

The authors are grateful to Lance Dixon, Gregory Soyez, Jesse Thaler for helpful discussions and suggestions. The work of E.M.M. is supported by the Office of Nuclear Physics of the U.S. Department of Energy (DOE) under grant DE-SC0011090 and the DOE Office of High Energy Physics under grant DE-SC-00012567. The work of B.N. is supported by the DOE Office of Science under contract DE-AC02-05CH11231. Cloud computing resources were provided through a Microsoft Azure for Research award.

References

[1] 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].

[2] A. Abdesselam et al., *Boosted objects: A Probe of beyond the Standard Model physics*, *Eur. Phys. J.* **C71** (2011) 1661, [arXiv:1012.5412].

[3] A. Altheimer et al., *Jet Substructure at the Tevatron and LHC: New results, new tools, new benchmarks*, *J. Phys.* **G39** (2012) 063001, [arXiv:1201.0008].

[4] A. Altheimer et al., *Boosted objects and jet substructure at the LHC. Report of BOOST2012, held at IFIC Valencia, 23rd-27th of July 2012*, *Eur. Phys. J.* **C74** (2014), no. 3 2792, [arXiv:1311.2708].

[5] D. Adams et al., *Towards an Understanding of the Correlations in Jet Substructure*, *Eur. Phys. J.* **C75** (2015), no. 9 409, [arXiv:1504.00679].

[6] A. J. Larkoski, I. Moult, and D. Neill, *Power Counting to Better Jet Observables*, *JHEP* **12** (2014) 009, [arXiv:1409.6298].

[7] I. Moult, L. Necib, and J. Thaler, *New Angles on Energy Correlation Functions*, arXiv:1609.07483.

[8] A. J. Larkoski, I. Moult, and D. Neill, *Analytic Boosted Boson Discrimination at the Large Hadron Collider*, arXiv:1708.06760.

[9] D. Guest, K. Cranmer, and D. Whiteson, *Deep Learning and its Application to LHC Physics*, arXiv:1806.11484.

[10] A. Radovic, M. Williams, D. Rousseau, M. Kagan, D. Bonacorsi, A. Himmel, A. Aurisano, K. Terao, and T. Wongjirad, *Machine learning at the energy and intensity frontiers of particle physics*, *Nature* **560** (2018), no. 7716 41–48.

[11] K. Albertsson et al., *Machine Learning in High Energy Physics Community White Paper*, arXiv:1807.02876.

[12] A. J. Larkoski, I. Moult, and B. Nachman, *Jet Substructure at the Large Hadron Collider: A Review of Recent Advances in Theory and Machine Learning*, arXiv:1709.04464.

[13] K. Datta and A. Larkoski, *How Much Information is in a Jet?*, *JHEP* **06** (2017) 073, [arXiv:1704.08249].

[14] K. Datta and A. J. Larkoski, *Novel Jet Observables from Machine Learning*, arXiv:1710.01305.

[15] H. Luo, M.-x. Luo, K. Wang, T. Xu, and G. Zhu, *Quark jet versus gluon jet: deep neural networks with high-level features*, arXiv:1712.03634.

[16] P. T. Komiske, E. M. Metodiev, and J. Thaler, *Energy flow polynomials: A complete linear basis for jet substructure*, arXiv:1712.07124.

[17] J. Neyman and E. S. Pearson, *Ix. on the problem of the most efficient tests of statistical hypotheses*, *Phil. Trans. R. Soc. Lond. A* **231** (1933), no. 694-706 289–337.

[18] D. E. Soper and M. Spannowsky, *Finding top quarks with shower deconstruction*, *Phys. Rev.* **D87** (2013) 054012, [arXiv:1211.3140].

[19] Yu. L. Dokshitzer, G. Marchesini, and G. P. Salam, *Revisiting parton evolution and the large-x limit*, *Phys. Lett.* **B634** (2006) 504–507, [hep-ph/0511302].

[20] Yu. L. Dokshitzer and G. Marchesini, *N=4 SUSY Yang-Mills: three loops made simple(r)*, *Phys. Lett.* **B646** (2007) 189–201, [hep-th/0612248].

[21] M. Beccaria, Yu. L. Dokshitzer, and G. Marchesini, *Twist 3 of the sl(2) sector of N=4 SYM and reciprocity respecting evolution*, *Phys. Lett.* **B652** (2007) 194–202, [arXiv:0705.2639].

[22] M. Levy and J. Sucher, *Eikonal approximation in quantum field theory*, *Phys. Rev.* **186** (1969) 1656–1670.

[23] F. E. Low, *Bremsstrahlung of very low-energy quanta in elementary particle collisions*, *Phys. Rev.* **110** (1958) 974–977.

[24] T. H. Burnett and N. M. Kroll, *Extension of the low soft photon theorem*, *Phys. Rev. Lett.* **20** (1968) 86.

[25] V. Del Duca, *High-energy Bremsstrahlung Theorems for Soft Photons*, *Nucl. Phys.* **B345** (1990) 369–388.

[26] I. Moult, L. Rothen, I. W. Stewart, F. J. Tackmann, and H. X. Zhu, *Subleading Power Corrections for N-Jettiness Subtractions*, arXiv:1612.00450.

[27] R. Boughezal, X. Liu, and F. Petriello, *Power Corrections in the N-jettiness Subtraction Scheme*, arXiv:1612.02911.

[28] I. Moult, L. Rothen, I. W. Stewart, F. J. Tackmann, and H. X. Zhu, *N-Jettiness Subtractions for $gg \to H$ at Subleading Power*, arXiv:1710.03227.

[29] R. Boughezal, A. Isgrò, and F. Petriello, *Next-to-leading-logarithmic power corrections for N-jettiness subtraction in color-singlet production*, arXiv:1802.00456.

[30] I. Moult, I. W. Stewart, G. Vita, and H. X. Zhu, *First Subleading Power Resummation for Event Shapes*, arXiv:1804.04665.

[31] A. J. Larkoski, I. Moult, and D. Neill, *Building a Better Boosted Top Tagger*, *Phys.Rev.* **D91** (2015), no. 3 034035, [arXiv:1411.0665].

[32] R. L. Jaffe and M. Soldate, *Twist Four in Electroproduction: Canonical Operators and Coefficient Functions*, *Phys. Rev.* **D26** (1982) 49–68.

[33] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, *Soft Drop*, *JHEP* **05** (2014) 146, [arXiv:1402.2657].

[34] C. Frye, A. J. Larkoski, J. Thaler, and K. Zhou, *Casimir Meets Poisson: Improved Quark/Gluon Discrimination with Counting Observables*, *JHEP* **09** (2017) 083, [arXiv:1704.06266].

[35] A. Armoni, *Anomalous Dimensions from a Spinning D5-Brane*, *JHEP* **11** (2006) 009, [hep-th/0608026].

[36] R. H. Boels, T. Huber, and G. Yang, *Four-Loop Nonplanar Cusp Anomalous Dimension in N=4 Supersymmetric Yang-Mills Theory*, *Phys. Rev. Lett.* **119** (2017), no. 20 201601, [arXiv:1705.03444].

[37] A. Grozin, J. Henn, and M. Stahlhofen, *On the Casimir scaling violation in the cusp anomalous dimension at small angle*, *JHEP* **10** (2017) 052, [arXiv:1708.01221].

[38] S. Moch, B. Ruijl, T. Ueda, J. A. M. Vermaseren, and A. Vogt, *Four-Loop Non-Singlet Splitting Functions in the Planar Limit and Beyond*, *JHEP* **10** (2017) 041, [arXiv:1707.08315].