IL NUOVO CIMENTO Vol. 37 C, N. 2 Marzo-Aprile 2014
DOI 10.1393/ncc/i2014-11746-x

Colloquia: LC13

QCD calculations for jet substructure

M. Dasgupta

School of Physics and Astronomy, The University of Manchester
Oxford Road, Manchester M13 9PL, UK

ricevuto il 22 Gennaio 2014

Summary. — We present results on novel analytic calculations to describe invariant mass distributions of QCD jets with three substructure algorithms: trimming, pruning and the mass-drop taggers. These results not only lead to considerable insight into the behaviour of these tools, but also show how they can be improved. As an example, we discuss the remarkable properties of the modified mass-drop tagger.

PACS 12.38.-t – Quantum chromodynamics.
PACS 12.38.Cy – Summation of perturbation theory.
PACS 12.38.Aw – General properties of QCD (dynamics, confinement, etc.).
PACS 12.38.Bx – Perturbative calculations.

1. – The boosted regime and substructure tools

The main aims of the research programme carried out at the Large Hadron Collider (LHC) at CERN are to understand the mechanism of electroweak symmetry breaking and to explore the TeV scale for signs of new physics beyond the Standard Model of particle physics. In order to achieve this, protons are collided at energies far above the electroweak scale, opening up the possibility of producing electroweak-scale particles with a large boost. In these situations, their hadronic decay products are collimated into a single jet. Consequently a vibrant research field has emerged in recent years, investigating how best to identify the characteristic substructure that appears inside "signal" jets in order to differentiate them from background (QCD) jets (for a review of the field see refs. [1,2]). Many "grooming" and "tagging" algorithms have been developed, successfully tested and are already being used experimental analyses.

Until very recently, nearly all the theoretical studies of substructure tools have been done using Monte Carlo parton showers. While these are powerful general purpose tools, their essentially numerical nature offers little insight into the results produced or their detailed and precise dependence on tagger parameters and the parameters of jet finding. Such a detailed level of understanding, which can be achieved for example via analytical formulae, is in fact crucial in order for substructure studies to realise their full potential. However it has been *far from obvious* that given their inherent complexity, substructure taggers can be understood to any extent analytically.

In two recent papers [3, 4] we have developed the first comprehensive theoretical understanding of three commonly used substructure tools: trimming [5], pruning [6, 7] and the mass drop tagger [8]. In these proceedings we review the main results of those papers, focussing on the perturbative properties of jet mass distributions of QCD jets with the application of substructure algorithms, and compare the results to the plain jet mass distribution.

2. – The perturbative structure of jet mass distributions

Jet mass distributions are affected by logarithmic corrections in the ratio of jet invariant mass (m) over its transverse momentum (p_t). When this ratio becomes small, as happens for highly boosted configurations, these logarithms are large and fixed-order perturbation theory is not a reliable way to organise the calculation. One then needs to resum these large corrections to all orders in perturbation theory. Resummed results can be then matched to fixed-order ones, typically obtained at next-to-leading order (NLO), in order to obtain a reliable estimate of jet masses over a wide range of m/p_t.

2˙1. *Plain jet mass.* – Resummed calculations are usually discussed in terms of the cumulative distributions, *i.e.* the integral of the jet mass distribution up to a fixed value:

$$\Sigma(\rho) = \frac{1}{\sigma}\int^{\rho} \frac{\mathrm{d}\sigma}{\mathrm{d}\rho'}\,\mathrm{d}\rho', \qquad \rho = \frac{m^2}{p_t^2 R^2}\,, \tag{1}$$

where R is the jet radius. In our discussion, we will work in the small jet radius limit $R \ll 1$. This considerably simplifies our expressions because we only have to consider the radiation from the parton that initiated the jet: large-angle radiation from other final-state partons and from the initial-state partons result in contributions which are power suppressed in R. For brevity, we also limit ourselves to the case of quark-initiated jets.

To next-to-leading logarithmic (NLL) accuracy, *i.e.* control of terms $\alpha_s^n L^{n+1}$ and $\alpha_s^n L^n$ in $\ln \Sigma(\rho)$, where $L \equiv \ln \frac{1}{\rho}$, the cumulative distribution can be computed using an independent-emission approximation, ignoring subsequent splittings of those emissions, other than in the treatment of the running coupling. The NLL result, in the small-R limit, can be written as

$$\Sigma(\rho) = e^{-D(\rho)} \cdot \frac{e^{-\gamma_E D'(\rho)}}{\Gamma(1 + D'(\rho))} \cdot \mathcal{N}(\rho)\,. \tag{2}$$

The first factor, which is double logarithmic, accounts for the Sudakov suppression of emissions that would induce a (squared, normalised) jet mass greater than ρ. In a fixed coupling approximation the resummed exponent reduces to

$$D(\rho) \simeq \frac{\alpha_s C_F}{\pi}\left[\frac{1}{2}\ln^2\frac{1}{\rho} - \frac{3}{4}\ln\frac{1}{\rho} + \mathcal{O}(1)\right]. \tag{3}$$

The second factor in eq. (2), accounts for the fact that the effects of multiple emissions add together to give the jet's overall mass. The third factor, also single logarithmic, accounts for modifications of the radiation pattern in the jet (non-global logarithms and boundaries of the jet induced by soft radiation near the jet's edge.

2˙2. *Trimmed mass distribution.* – Trimming takes all the particles in a jet of radius R and reclusters them into subjets with a jet definition with radius $R_{\rm sub} < R$. All resulting subjets that satisfy the condition $p_t^{(\rm subjet)} > z_{\rm cut} p_t^{(\rm jet)}$ are kept and merged to form the trimmed jet. The other subjets are discarded.

We can get an idea of the trimmed jet mass behaviour by considering configurations in which the jet is made of a hard quark and a bunch of soft gluons. It is then clear that the algorithm will cut away soft radiation if emitted at angles larger than $R_{\rm sub}$, while arbitrarily soft gluons radiated at angles smaller than $R_{\rm sub}$ will contribute to the trimmed jet mass.

The full leading logarithmic (LL) calculation of the trimmed jet mass produces:

$$\Sigma^{(\rm trim)}(\rho) = \exp\left[- D(\max(z_{\rm cut},\rho)) - S(z_{\rm cut},\rho)\Theta(z_{\rm cut}-\rho) \right. \\ \left. -\Theta(z_{\rm cut} r^2 - \rho) \int_\rho^{z_{\rm cut} r^2} \frac{{\rm d}\rho'}{\rho'} \int_{\rho'/r^2}^{z_{\rm cut}} \frac{{\rm d}z}{z} \frac{C_F}{\pi}\, \alpha_s(\rho' z p_t^2 R^2) \right], \tag{4}$$

where $r = \frac{R_{\rm sub}}{R}$ and we have neglected finite $z_{\rm cut}$ corrections. The resummed exponent D is the same as in eq. (3), while the function S is single-logarithmic and in a fixed-coupling approximation is given by

$$S(a,b) \simeq \frac{\alpha_s C_F}{\pi} \left[\ln\frac{1}{z_{\rm cut}} - \frac{3}{4} + \mathcal{O}\left(z_{\rm cut}\right) \right] \ln\frac{a}{b}\,, \tag{5}$$

We can now discuss differences and similarities of the trimmed mass distribution in eq. (4) to the plain jet mass eq. (2). The main similarity from the point of view of resummed calculations is that in both cases the analysis of the one loop case essentially captures the LL behaviour to all orders (this is not the case for pruning or mass drop). However, the actual form of the one-gluon exponentiation in the case of trimming has a non-trivial dependence on the jet's kinematics. We can identify three distinct kinematic regions: for $\rho > z_{\rm cut}$ trimming is not active and the result is the same as plain jet mass. For $r^2 z_{\rm cut} < \rho < z_{\rm cut}$, the parameter $z_{\rm cut}$ provides a lower limit for the emissions' energy, resulting into a single-logarithmic distributions. The last region $\rho < r^2 z_{\rm cut}$ is again double logarithmic and it correspond to configurations in which soft gluons are emitted at angles smaller than $R_{\rm sub}$, as mentioned above.

Equation (4) does not capture full NLL accuracy *i.e.* all terms $\alpha_s^n L^n$ in $\ln\Sigma(\rho)$. The missing terms include non-global logarithms, related clustering logarithms, and multiple-emission effects on the observable. They should all be relatively straightforward to include, if desired, since they follow the structure of corresponding terms for the plain jet-mass distribution.

In order to test that the approximations made in order to obtain the resummed result in eq. (4) capture the relevant physical effects, we compare our result to the one obtained with a Monte Carlo parton shower. This comparison is shown in fig. 1. Our calculations indeed reproduce the shape of the distribution in all three distinct regions, as well as the position of the transition points between these regions (indicated by vertical arrows), which confirms that we have analytically captured the essence of trimming.

2˙3. *Pruned mass distribution.* – Pruning takes an initial jet, and from its mass deduces a pruning radius $R_{\rm prune} = R_{\rm fact} \cdot \frac{2m}{p_t}$, with $R_{\rm fact}$ of order 1 (here we choose $R_{\rm fact} = 0.5$

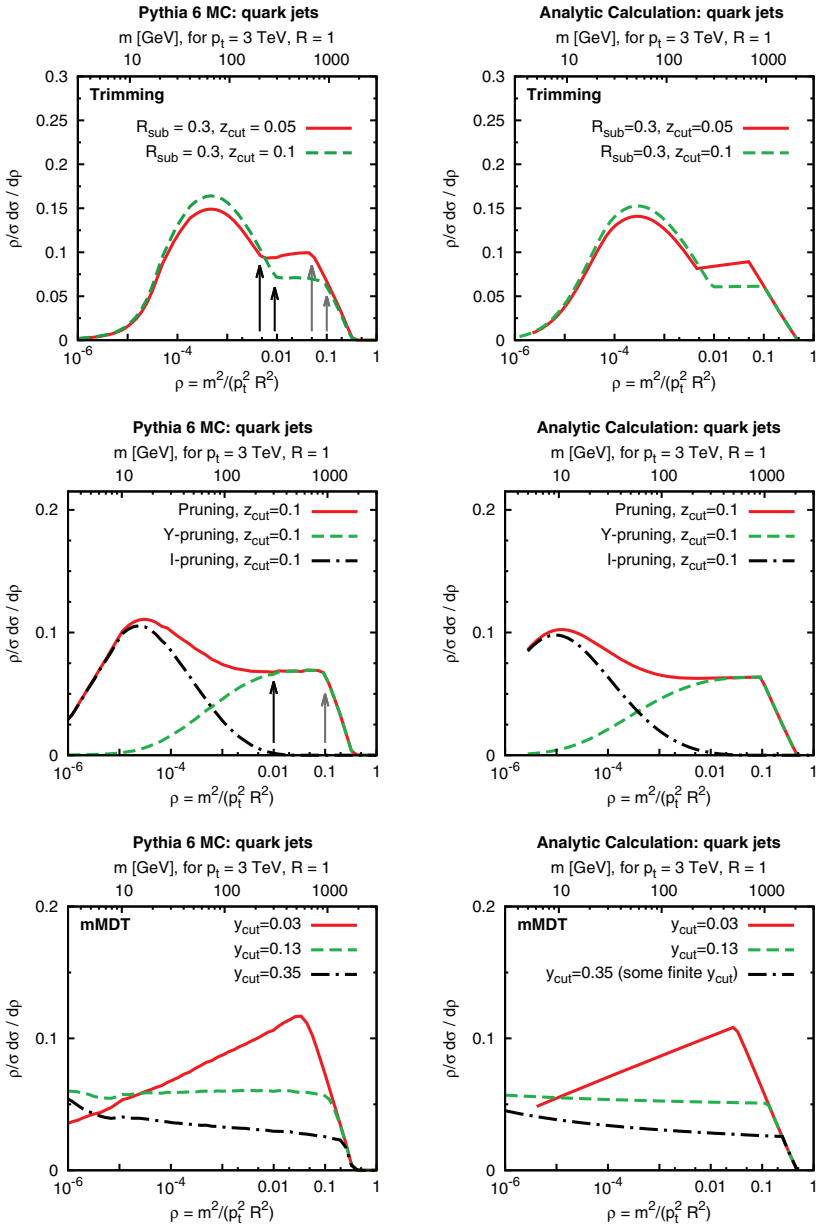

Fig. 1. – Comparisons of the resummed calculations for the mass distributions (right-hand panels) to a standard parton shower (left-hand plots). The arrows indicate the analytic prediction for the position of the transition points. The results on the left-hand panels have been obtained from Monte Carlo simulation with Pythia 6.425 in the DW tune (virtuality-ordered shower), with a minimum p_t cut in the generation of 3 TeV, for 14 TeV pp collisions, at parton level, including initial and final-state showering, but without the underlying event (multiple interactions).

but our main conclusions do not depend on this choice). It then reclusters the jet and for every clustering step, involving objects a and b, it checks whether $\Delta_{ab} > R_{\rm prune}$ and $\min(p_{ta}, p_{tb}) < z_{\rm cut} p_{t,(a+b)}$, where $z_{\rm cut}$ is a second parameter of the tagger. If so, then the

softer of the a and b is discarded. Otherwise a and b are recombined as usual. Clustering then proceeds with the remaining objects, applying the pruning check at each stage.

We can start our analysis by considering the $\mathcal{O}(\alpha_s)$ contribution. At this perturbative order the jet is made of two partons, a and b with $m = m_{ab}$ and $\Delta_{ab} > R_{\text{prune}} = \frac{m_{ab}}{p_t}$. Thus, the two partons are kept only if they pass the energy condition, irrespectively of their angular distance. This has a remarkable consequence: the LO pruned mass distribution receives no contributions from the soft region and it has only a single logarithmic, which is of pure collinear origin.

This behavior certainly appears desirable from the viewpoint of taming the background jet mass distribution as it rids us of double logarithms. Thus we may wonder if the above feature holds to all orders. However an analysis of the NLO contributions reveal that this is not the case and the pruned mass distribution receives contributions from soft emissions beyond LO and consequent double logarithmic enhancements appear.

In particular, we consider NLO configurations (which involve three partons), where there is a soft parton (p_3) that dominates the total jet mass thus setting the pruning radius, but it is soft enough that it fails the z_{cut} threshold and therefore it does not contribute to the pruned mass; meanwhile there is another parton (p_2), within the pruning radius, that contributes to the pruned jet mass independently of how soft it is. We call this "I-pruning", because at the angular scale R_{prune} (set by the soft parton p_3), the final pruned jet consists of a single hard prong. On the other hand, we call "Y-pruning" those configurations that contributed to the leading order result for which at an angular scale R_{prune}, the pruned jet always consisted of two prongs.

The above analysis can be generalised to all orders and a resummed result for pruning and its Y-pruning and I-pruning components can be found but for brevity here we avoid writing down the formulae and refer the reader to out original articles. Several comments can be made about the perturbative structure of the results.

I-pruning has a double logarithmic behaviour $\alpha_s^n L^{2n}$ like the plain jet mass. On the other hand, Y-pruning is essentially a Sudakov suppression of the leading order result and, therefore, $\Sigma^{(\text{Y-prune})}(\rho)$ is as singular as $\alpha_s^n L^{2n-1}$. Interestingly, when considering full pruning, *i.e.* the sum of the two components, a cancellation occurs in the $z_{\text{cut}}^2 < \rho < z_{\text{cut}}$ region and one obtains a distribution which is only single-logarithmic.

As for trimming, to reach full NLL accuracy would require the treatment of several additional effects: non-global logarithms and related clustering logarithms and multiple-emission effects on the observable.

The comparison between the analytic calculation and the Pythia shower is shown in fig. 1 in the middle panel. There one observes excellent agreement between the shapes of the analytical and MC distributions, indiacting once again a successful analytical description of pruning.

2˙4. *MDT and mMDT mass distribution.* – The mass-drop tagger (MDT) [8] is a declustering algorithm to be used with Cambridge/Aachen jets. In its original incarnation, the algorithm starts from a jet j, then undoes the last step of the clustering finding two subjets j_1 and j_2, with $m_{j_1} > m_{j_2}$ If there was a significant mass drop, $m_{j_1} < \mu m_j$, and the splitting is not too asymmetric, $y = \min(p_{tj_1}^2, p_{tj_2}^2)\Delta R_{j_1 j_2}^2/m_j^2 > y_{\text{cut}}$, then the jet j is tagged. Otherwise j is redefined to be equal to j_1 and the algorithm iterates (unless j consists of just a single particle, in which case the original jet is deemed untagged).

At $\mathcal{O}(\alpha_s)$ the mass-drop condition is always satisfied, so we only need to check for the y_{cut} condition, which is essentially a cut on the energy sharing between the two prongs. This situation is completely analogous to what we have encountered for pruning at LO

and the resulting mass distribution has only a single logarithm. However, starting from NLO the behaviour of MDT is far from straightforward. Complications arise because MDT recurses on the more massive branch, which in principle can be the softer of a given subjet pair. This was not what was intended in the original design, intended to tag hard substructure, and is to be considered a flaw.

The modified mass drop tagger (mMDT) is instead defined in such a way that it recurses on the subjet with the largest $m^2 + p_t^2$. Not only does the mMDT eliminate the wrong-branch issue, but it also turns out to greatly facilitate the resummation of the tagged mass distribution. We find that the all-order mMDT mass distribution is simply given by the exponential of the one-loop result:

$$\Sigma^{(\mathrm{mMDT})}(\rho) = \exp\left[-D(\max(y_{\mathrm{cut}}, \rho)) - S(y_{\mathrm{cut}}, \rho)\Theta(y_{\mathrm{cut}} - \rho)\right]. \tag{6}$$

The mass distribution above has remarkable properties: it only contains single-logarithmic ($\alpha_s^n L^n$) contributions. All contributions from soft emissions have been successfully removed. It is to our knowledge the first time that a jet-mass type observable is found with this property. We will analyse the salient properties of mMDT in more detail in the next section.

The comparison between our analytic calculation and the Pythia shower is shown in fig. 1 in the bottom panel and yet again we note that our resummation perfectly captures the behaviour of the mMDT.

3. – Conclusions and prospects

We have succeeded in providing a first systematic understanding of commonly used jet substructure tools. Prior to our work it was not possible to comapare in detail the performance of the various tools across a wide range of their parameters and a wide kinematic range. This is because one would have to provide detailed Monte Carlo analyis separately for all the regions and then hope that significant features of the taggers would be immediately apparent. With analytical formulae however one is powerfully placed to understand in detail the nature of the tools for arbitrary parameter values and kinematic regions. Such formulae can then provide the guiding principles on which more robust and superior tools can be engineered. We believe our work will be most relevant to such developments and look forward to further studies.

REFERENCES

[1] Abdesselam A., Kuutmann E. B., Bitenc U., Brooijmans G., Butterworth J., Bruckman de Renstrom P., Buarque Franzosi D. and Buckingham R. *et al.*, *Eur. Phys. J. C*, **71** (2011) 1661.
[2] Altheimer A., Arora S., Asquith L., Brooijmans G., Butterworth J., Campanelli M., Chapleau B. and Cholakian A. E. *et al.*, *J. Phys. G*, **39** (2012) 063001.
[3] Dasgupta M., Fregoso A., Marzani S. and Salam G. P., *JHEP*, **1309** (2013) 029.
[4] Dasgupta M., Fregoso A., Marzani S. and Powling A., *Eur. Phys. J. C*, **73** (2013) 2623.
[5] Krohn D., Thaler J. and Wang L.-T., *JHEP*, **1002** (2010) 084.
[6] Ellis S. D., Vermilion C. K. and Walsh J. R., *Phys. Rev. D*, **81** (2010) 094023.
[7] Ellis S. D., Vermilion C. K. and Walsh J. R., *Phys. Rev. D*, **80** (2009) 051501.
[8] Butterworth J. M., Davison A. R., Rubin M. and Salam G. P., *Phys. Rev. Lett.*, **100** (2008) 242001.