PROCEEDINGS OF SCIENCE

Subjet algorithm and N-subjettiness for identifying fat jets

Jihun Kim*
FPRD and Department of Physics, Seoul National University, Seoul, 151-747, Korea
E-mail: jihuni@phya.snu.ac.kr

The rest frame subjet algorithm is introduced to define the subjets for the SISCone jet; with this algorithm, an infrared and collinear safe jet shape observable N-subjettiness, τ_N^j, is defined to discriminate the fat jet, from a highly boosted color singlet particle decaying to N partons, from the QCD jet. Using rest frame subjets and τ_2^j on dijets from highly boosted $H/W/Z$ bosons through $pp \to HW$, HZ with $m_H = 120$GeV, we found that statistical significance of the signal, from the fully hadronic channels, is about 2σ for 14TeV collisions with $\mathscr{L} \sim 30\text{fb}^{-1}$.

The 2011 Europhysics Conference on High Energy Physics-HEP 2011,
July 21-27, 2011
Grenoble, Rhône-Alpes France

*Speaker.

PoS(EPS-HEP2011)231

1. Introduction

The electroweak precision fits prefer light standard model Higgs; and experimental con- straints suggest standard model Higgs mass of about 120GeV. However, the Higgs lighter than 135GeV dominantly decays to a b-quark pair. Thus, the signals involve huge QCD background; itâĂŹs why light standard model Higgs has been considered hard to discover.

Colored partons radiate to produce soft and collinear quarks and gluons which evolve to a numerous hadrons collimated along some directions, so called hadron jets. Jet algorithms are for grouping them. Among them, inclusive kT, Cambridge/Aachen, Anti-kT, and SISCone jet algorithm [1] are infrared and collinear safe and popular. All of them have a free parameter, so called a jet radius. Broadly speaking, it represent a distance measure, and a reach of jets; although its precise meanings depend on each jet algorithm. There is a tension to choose the jet radius. To capture the radiations of colored partons, larger jet radius is better. However,underlying event and pileup events at the LHC produce numerous soft particles. They ruin the signal such as invariant mass peak of the jets. To reduce the contaminations, smaller jet radius is better. Several studies on an optimal value for a jet radius.

2. Studies on jet substructure

So one need to compromise at some value. Two approaches are studied extensively. One is figure out optimal value for the jet radius from perturbative QCD. Other is set the jet radius value sufficiently large, and remove some of its constituent particle to reduce the contamination from underlying event and pileup. They are so called jet substructure algorithms; including mass drop-filtering[2], trimming, pruning[3, 4, 5, 6].

With these, one get jets which capture the radiations from the original parton while rejecting other radiations come from other source. Given the jet, next question would be how one can identify its origin. It may comes from quark, gluon, or W/Z bosons, Higgs or new particles. Usually, it is very hard job because backgrounds are much larger than those of lepton signals.

To be specific, consider Higgs production associated with W/Z bosons, $pp \to HW/Z \to b\bar{b} + leptons$. One of main backgrounds is $t\bar{t} \to b\bar{b}W^{+}W^{-}$. If some of decay products of W bosons are not detected by the detector, it gives same objects : b-jets and leptons / missing energy. Consider the Higgs highly boosted so that the b-quarks are collimated to be identified as a singlet jet. We call it a fat jet. In this case, leptons from the W/Z bosons are also highly boosted to opposite directions. For $t\bar{t}$, it is very hard to give such topologies without producing other hadronic activity. In other words, geometric shape of fat jet signals are constrained, and background is smaller than ordinary cases.

3. N-subjettiness

Given an event with fat jets, however, how one can know whether it is a signal or not? Although there are many studies on fat jet tagging, they do not tell us whether the jet comes from the signal or QCD background. It is what the N-subjettiness[7] is devised for. N-subjettiness is a jet shape observable to find highly boosted objects decaying to N partons; In other words, N-subjettiness is for selecting fat jets with N-subjets.

PoS(EPS-HEP2011)231

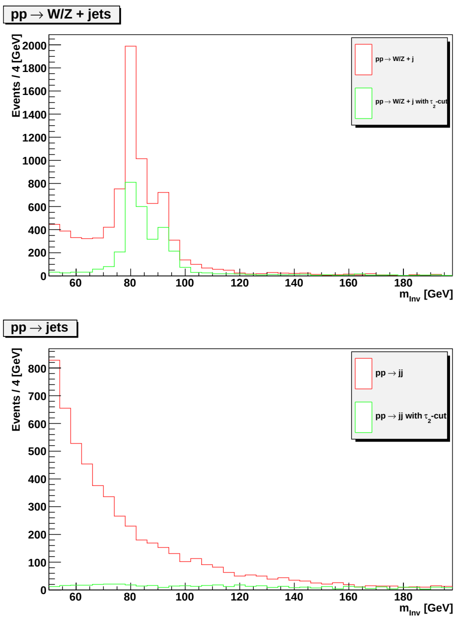

Figure 1: before(red curve) and after(green curve) applying two-subjettiness to the leading jets from $pp \to W/Z + jets$ (above figure) and its QCD background $pp \to jets$ (bottom figure).

It is a variation of the global event shape observable N-jettiness[8], suggested by Ian Stewart, et al. Here is a definition of N-subjettiness. It is sum of distances between its constituents particles and subjets. We use Lorentz product divided by jet mass for the distance. (Of course, one can use other measure; and J. Thaler suggested different distance measures.)

To see why it works, consider how the $b\bar{b}$ quarks evolves. They radiate quarks and gluons and finally hadronize. However, they do not exchange four momentum with outer system since they form a color singlet. In contrast, $b\bar{b}$ quarks from gluon form a color octet. They evolve according to the splitting kernels of the DGLAP equation. τ_2^j distributions of jets with $110\text{GeV} \leq m_{inv}^{jet} \leq 130\text{GeV}$, $p_T > 200\text{GeV}$, $|y| < 2.5$ are shown at Fig. 2.

How underlying event, pileup, and imperfection of jet reconstruction affect the rest frame subjet and N-subjettiness? Let's consider two extreme cases. The first case: soft particles are nearly parallel to the jet axis. For the first case, the particle's energy is even suppressed by the boost into the jet rest frame. Moreover, the particle doesn't change the jet mass much. So its effect on the N-subjettiness is negligible. The second case : soft particles that reside around the edge of the cone. In this case, the particles energy can be increased about factor 3 in case of a jet with radius 0.7 and Lorentz factor of a mother particle is about 10. But they are collimated to the boost axis, say around $\cos\theta_c < 0.97$. So it doesn't affect the leading energetic rest frame subjets; it only affects jet mass. If jet mass is changed about 10%, τ_N^{rest} can be changed about 20%. In any cases, N-subjettiness is not changed much by inclusion of a ultra soft particle. Thus, N-subjettiness is infra-red safe observable. By definition, N-subjettiness is collinear safe, too.

Fig. 1 is before and after applying two-subjettiness to the leading jets from $pp \to W/Z + jets$ and its QCD background $pp \to jets$. And this is a $S/\sqrt{B}$ improvement graph as a function of τ_2^{rest}-cut value. For the signal process, only well reconstructed fat jets pass the τ_2^{rest}-cut. We applied

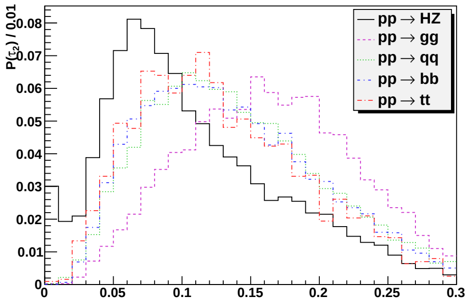

Figure 2: τ_2^j distributions of jets with $110\text{GeV} \leq m_{inv}^{jet} \leq 130\text{GeV}$, $p_T > 200\text{GeV}$, $|y| < 2.5$.

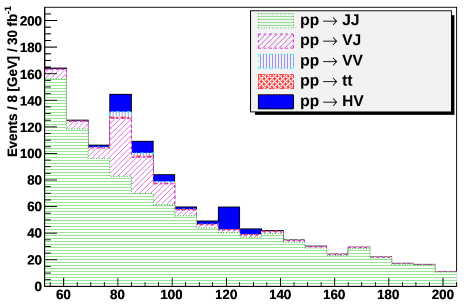

Figure 3: Expected m_{jet} distributions of the signal and background events that passes all cuts with $\mathscr{L} \sim 30\text{fb}^{-1}$. Pythia with the ATLAS MC09 parameter tune and SISCone jet with $R = 0.8$ is used. The JJ sample includes bb. The b-tag efficiency, c-jet misidentification and light-jet misidentification probabilities are assumed to be 70%, 10%, 1%.

two-subjettiness to the Higgs searches through fully hadronic decay channels of $pp \to HW/Z$. Because it involve two fat jets, the improvement factor is squared.

Here is the event selection scheme for the Higgs searches. Here is some usual selection cut : Both of two hardest jet have $p_T > 200\text{GeV}$, absolute values of pseudo-rapidity is smaller than 2.5, a pseudo-rapidity difference of the two jets is smaller than 2.0, third hardest jet must be not hard : $p_T^{j_3} < 30\text{GeV}$. And following is our additional selection cut : both of two hardest jet have $\tau_2^{rest} < 0.08$, $\cos\theta_s < 0.8$, and, for the Higgs candidate jets, two leading rest frame subjet is required to be b-tagged.

Fig. 3 is the result. Most large background comes from QCD dijet events. The statistical significance of the signal is about 2 σ. The result may depends on the gluon splitting modeling, or MC generator's parameter tune. We use PYTHIA 6.4.23 with the ATLAS MC09 parameter tune, and the modified leading-order MRST2007 parton distribution functions.

4. Conclusion

In conclusion, for the boosted color singlet particle searches, we have introduced the jet rest frame, the rest frame subjet, and the N-subjettiness. Using SISCone jet with τ_2^j, we see that, for the fully hadronic $pp \to HV$ channels of the SM Higgs with $m_H = 120$GeV, the statistical significance of the signals is about 2σ for 14TeV collisions with $\mathscr{L} \sim 30\text{fb}^{-1}$. Although it will be complementary to the known Higgs search channels, the scheme suggested in this letter is rather a proof of concept; the scheme will be improved further to increase the signal to background ratio, and to make full use of the jet rest frame. It involves comprehensive studies on theoretical uncertainties of the scheme, and we left them for the future study. The rest frame subjet can be defined by any jet algorithms, although the effects of underlying event, and pileup on the scheme depend on the jet algorithm. We also expect the scheme can also be employed for highly boosted colored particles.

References

[1] G. P. Salam and G. Soyez, JHEP **0705**, 086 (2007) [arXiv:0704.0292 [hep-ph]].

[2] J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, Phys. Rev. Lett. **100**, 242001 (2008) [arXiv:0802.2470 [hep-ph]].

[3] T. Plehn, G. P. Salam and M. Spannowsky, Phys. Rev. Lett. **104**, 111801 (2010) [arXiv:0910.5472 [hep-ph]].

[4] S. D. Ellis, C. K. Vermilion and J. R. Walsh, Phys. Rev. D **81**, 094023 (2010) [arXiv:0912.0033 [hep-ph]].

[5] D. E. Soper and M. Spannowsky, JHEP **1008**, 029 (2010) [arXiv:1005.0417 [hep-ph]].

[6] D. Krohn, J. Thaler and L. T. Wang, JHEP **1002**, 084 (2010) [arXiv:0912.1342 [hep-ph]].

[7] J. -H. Kim, Phys. Rev. D **83** (2011) 011502 [arXiv:1011.1493 [hep-ph]].

[8] I. W. Stewart, F. J. Tackmann and W. J. Waalewijn, Phys. Rev. Lett. **105**, 092002 (2010) [arXiv:1004.2489 [hep-ph]].

PoS(EPS-HEP2011)231