

Methods for jet clustering algorithms

A SUMMER STUDENT PROJECT

Konstantin Shchablo*

PH-UCM Group, CERN

*Federal State Budgetary Educational Institution of Higher Education Dubna State University, 141982, Russia

ShchabloKV@gmail.com

Supervisions: Maurizio Pierini, Maxime Gouzevitch

(Dated: August 21, 2015)

Abstract

The project of a Future Circular Collider will demand new algorithms of measuring events which will have high power level. In the context of a CERN summer student project, one of the possible ways for the modernization of jet algorithms was studied. This method allows to simplify work with a radius choice and the method has filter pileup.

1. INTRODUCTION

The project of a Future Circular Collider [6] which will support proton collisions at energies up to 100 TeV. This new power level for collision will demand more precise algorithms in data processing as events will have more particles than compared to the Large Hadron Collider [1], in which is creating complex data for use of jet algorithms. High energy hadron collisions produce quarks and gluons, which cannot exist as free particles and therefore have to be hadronize. After a hadronization the received hadrons can break up about the subsequent and then recombine to form new hadrons. In this project will simulate "The Drell-Yan process" [5]. The Drell-Yan process occurs in high energy hadron-hadron scattering. It takes place when a quark of one hadron and an antiquark of another hadron annihilate, creating a virtual photon or Z boson which then decays into a pair of oppositely-charged leptons. This process leads to creation of complex cascade of particles. The analysis of this cascade includes creation of jets from particles and work with jets for the determination processes, which had started the cascade. This method works because conservation of momentum and energy dictate that particles from cascade contains information about the processes from which they had originated. Future data processing will be complexity, with the correct choice of algorithms of a clustering of particles in jets, because the density of the particles in the events will be high.

In this project there is one of possible method for the jets creation algorithm improvement was presented. These methods allow to filter pileup, and also to simplify work with a radius choice.

2. METHODS

There are many algorithms for jets clusterization like "The kt algorithm", "The Cambridge and Aachen algorithms", "The anti-kt algorithm," etc. [7].

For approbation of these methods which are offered in this paper will use "The anti-kt algorithm", because this algorithm is one of the few algorithms, which leads to jets with shapes that are not influenced by soft perpendicular and strong parallel radiation; as a result this algorithm has a stable radius.

2.1. The anti-kt algorithm

The anti-kt jet clustering algorithm has determined in the following way: [2]:

As usual, one introduces distances d_{ij} between entities (particles, jets) i and j and d_{iB} between entity i and the beam (B). The (inclusive) clustering proceeds by identifying the smallest of the distances and if it is a d_{ij} recombining entities i and j , while if it is d_{iB} calling i a jet and removing it from the list of entities. The distances are recalculated and the procedure repeated until no entities are left.

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2} \quad (1)$$

$$d_{iB} = k_{ti}^{2p} \quad (2)$$

where

$$\Delta_{ij}^2 = (\gamma_i - \gamma_j)^2 + (\phi_i - \phi_j)^2 \quad (3)$$

and k_{ti} , γ_i and ϕ_i are respectively the transverse momentum, rapidity and azimuth of particle i . In addition to the usual radius parameter R , we have added a parameter p to govern the relative power of the energy versus geometrical (Δ_{ij}^2) scales.

When $p = 1$, (1) is called as the "anti-kt" jet-clustering algorithm.

2.2. "Jet of Jet" method

The method is based on the double clusterization particles in jets.

The first step includes making jets with calibrating radius. In the second step, the method creates jets with the last radius (a radius which is needed for methodological research) from the jets which were created in the previous step.

Schematically, the algorithm can be represented as follows: In the first phase, jets are created with a small radius. If jets have an energy level less than the minimum energy level which is required, they will be filtered. In the second phase, the small jets with were not filtered will cluster with a larger radius.

If clustering radiuses from the first step have calibrated, then this approach will allow one to choose an optimum radius. This calibration allows conditions in which the task for calculating the last radius will be parameterized. The Filter Effect works because all particles which have clusterized in jets with energy less than the minimum energy level, have been filtered.

This method can be determined in the following way: In the first step, we use the formula (1) for calculating jets. In the second step we also use (1), but i and j in this step are jets.

2.3. "Wide Jet" method

This Method includes two steps for the clusterization of jets. At the first step, all particles make jets with the calibration of the minimum radius, afterwards it will find the two biggest jets. The biggest jets with the last radius (radius which is needed for methodological research) are collected with all the jets from the first step.

"Wide jet" allows one to collect jets around the two biggest jets which have the last radius. Also, like "Jet Of Jet" this method filters jets with small energy.

Also, we can determine the method in the following way. In the first step, we use the formula (1) for calculating the jets. In the second step, we shall collect the jets around the two biggest jets with filter rules.

3. SIMULATION

By switching from the theory to the experimental realization we can see data simulation tools and an analysis scheme (see Fig. 1). This scheme describes links between tools for simulation and data analysis. Simulation data was used "BSMGen" and "Delphes 3" [4]. analysis data was based on "fastJet" [3] classes.

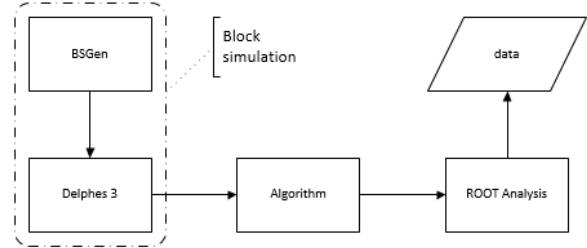


Figure 1: Simulation and data analysis scheme.

Where: BSMGen, Delphes 3 - programs for simulation data; Algorithm - block for create jets; ROOT Analysis - block for analysis jets.

In the first step data simulation in BSMGen program. In the second step, in Delphes block this data simulation with FCC detectors property. After we input data from Delphes processing in algorithm block. At last step result data for other algorithms compare between each other with ROOT CERN tools.

4. RESULTS, DISCUSSION

In the following section the resulting plots and tables are presented, i.e. the comparison of the data obtained from "The anti-kt clustering algorithm", "Jet of Jet method" and "Wide Jet method" for type collision quark quark with collision energies of 100 TeV and particle masses of 10 TeV, 30 TeV and 50 TeV for particles which

were collided. The data has been created with pileup and without pileup. We will show result with mass 10 TeV.

In the beginning, we find the radius for this collision energy and particles mass which gives a minimum resolution. We need to create jets with radius of 0.4, 0.45, ..., 1.6 (units) and find minimal diapason for dijet mass where we collect twenty-five percent of the events [7]. (see Fig. 2,3).

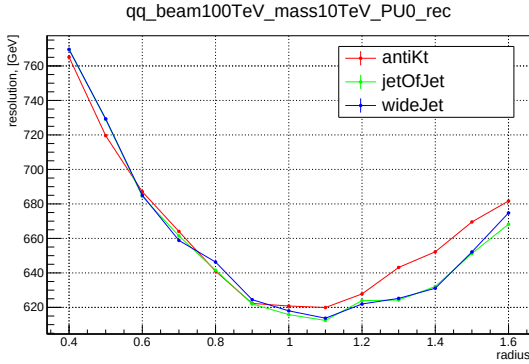


Figure 2: The graph shows resolution by radius for qq collision energy of 100 TeV and a mass of particles 10 TeV for the case with pileup.

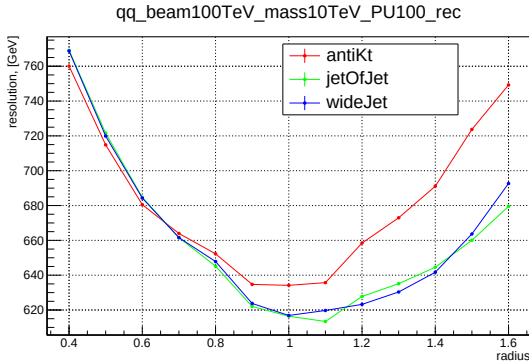


Figure 3: The graph shows resolution by radius for qq collision energy of 100 TeV and a mass of particles 10 TeV for the case with pileup.

We can see that graphs for resolution without pileup have the same shape and different shape between graphs for the situation with pileup. It happened because the filter effect for "Jet of Jet" and "wide jet" algorithms worked. From the calculation, we can use the radius with value equal to one, which is approximately the minimum resolution.

We compare data without pileup for "antiKt" with a radius equal to 0.6 a maximum calibrated radius, "Jet of Jet" = 1, "antiKt" = 1 and "Wide Jet" = 1 we can see, that for the radius with value equal to one, distributions do not overlap with each other; however, they have the same structure. Resolution for the radius = 1 is less than radius = 0.6 (see Fig. 4).

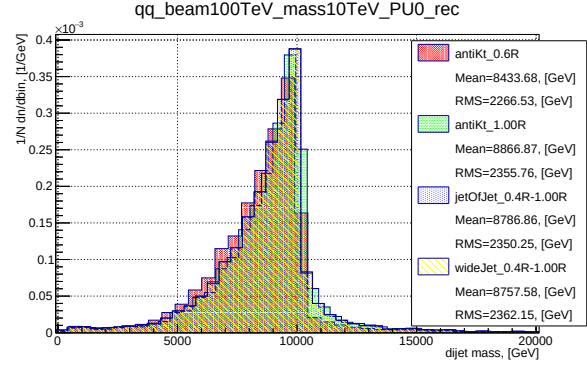


Figure 4: On graphic display there are plots for different algorithms for qq collision energy of 100 TeV and a mass of particles 10 TeV for the case without pileup

If we compare data without pileup for "antiKt" with radius equal to 0.6 a maximum calibrated radius) "Jet of Jet" = 1, "antiKt" = 1 and "Wide Jet" = 1 Also like in figure 4 for the radius with value equal to one, distributions do not overlap with each other, but "antiKt" = 0.6 with pileup has biggest tail than "antiKt" = 0.6 without pileup. (see Fig. 5).

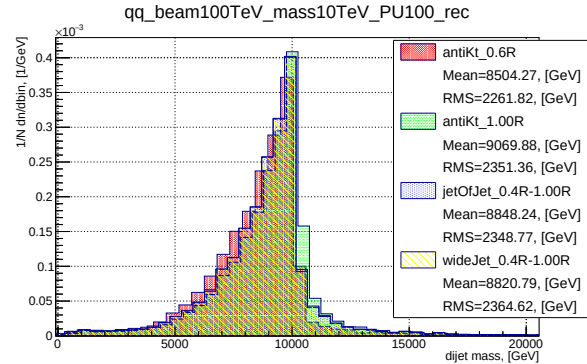


Figure 5: On graphic display there are plots for different algorithms for qq collision energy of 100 TeV and a mass of particles 10 TeV for the case with pileup

We can see that the pileup filtering will affect the stability of the mean value of the plots, which

shows that the differences are due to the effect the filter. (see Fig. 6) The mean has determined like weight function for three maximum bins. We show mean for algorithms with optimal radius for differences pileup.

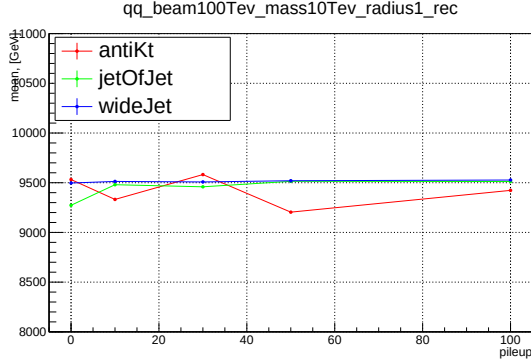


Figure 6: The graph shows mean by pileup for qq collision energy of 100 TeV and a mass of particles 10 TeV.

The Method "Jet Of Jet" and "wide jet" have more stability of mean and resolution than just a clustering algorithm. It means that "Jet Of Jet" and "wide jet" have filter effects.

Also need show that for quark quark collision with energy (30 TeV, 50 TeV) radius from the minimum resolution will different.

5. CONCLUSION

In this project, we have been simulating data for 100 TeV collision with different mass for particles (10 TeV, 30 TeV, 50 TeV) as well as with and without pileup. We have tested methods "Jet Of Jet" and "Wide Jet" to create jets and showed that applying those methods allow use optimum radius for methodological research and the filter effect make algorithms stability.

ACKNOWLEDGMENTS

I would like to extend my gratitude towards my supervisors Maurizio Pierini and Maxime

Gouzevitch for their great support to my project and all matters related with my work. Also, I would like to say thanks to organizers CERN summer students programme 2015 and PH-UCM group for a great opportunity to work with researchers and other summer students from all over the world and give me an unforgettable experience.

REFERENCES

- [1] Oliver Sim BrÄijning, Paul Collier, P Lebrun, Stephen Myers, Ranko Ostojic, John Poole, and Paul Proudlock. *LHC Design Report*. CERN, Geneva, 2004.
- [2] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. *The anti- k_t jet clustering algorithm*, volume 04. Feb 2008.
- [3] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. *FastJet user manual*, volume 72. Nov 2011. Comments: 69 pages. FastJet 3 is available from <http://fastjet.fr/>.
- [4] J de Favereau, C Delaere, P Demin, A Giammanco, V LemaÄotre, A Mertens, and M Selvaggi. *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*. Number arXiv:1307.6346. Jul 2013. Comments: JHEP 1402 (2014).
- [5] Tung-Mow A Drell, Sidney D.; Yan, V LemaÄotre, A Mertens, and M Selvaggi. *Massive Lepton-Pair Production in Hadron-Hadron Collisions at High Energies*. 08 1970.
- [6] Roman Martin, Rogelio Tomas Garcia, and L Medina. *Status of the FCC-ee Interaction Region Design*. Feb 2015.
- [7] Gavin P Salam. *Towards Jetography*. Number arXiv:0906.1833. Jun 2009. Comments: 88 pages, 27 figures, an extended version of lectures given at the CTEQ/MCNET school, Debrecen, Hungary, August 2008.