



BNL-81350-2008-CP

The SISCone and anti- k_t jet algorithms

Grégory Soyez

Presented at the International Workshop on Deep Inelastic Scattering & Related Subjects (DIS)
London, United Kingdom
April 7 to 11, 2008

**Physics Department
Nuclear Theory Group**

**Brookhaven National Laboratory
P.O. Box 5000
Upton, NY 11973-5000
www.bnl.gov**

Notice: This manuscript has been authored by employees of Brookhaven Science Associates, LLC under Contract No. DE-AC02-98CH10886 with the U.S. Department of Energy. The publisher by accepting the manuscript for publication acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

This preprint is intended for publication in a journal or proceedings. Since changes may be made before publication, it may not be cited or reproduced without the author's permission.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or any third party's use or the results of such use of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof or its contractors or subcontractors. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

The SISCone and anti- k_t jet algorithms

Gregory Soyez^{1*}

Brookhaven National Laboratory - Physics Department
Building 510, Upton, NY 11973 - USA

We illustrate how the midpoint and iterative cone (with progressive removal) algorithms fail to satisfy the fundamental requirements of infrared and collinear safety, causing divergences in the perturbative expansion. We introduce SISCone and the anti- k_t algorithms as respective replacements that do not have those failures without any cost at the experimental level.

The general picture. Jets are an important tool in hadronic physics and they will play a predominant role at the LHC. By defining jets one aims at accessing, from the final-state particles, the underlying hard parton-level processes. Since a parton is not a well-defined object, a jet definition is also not unique.

Two broad classes of jet definitions exist. The first one works by defining a distance between pairs of particles, performing successive recombinations of the pair of closest particles and stopping when all resulting objects are too far apart. Algorithms within that class differ by the definition of the distance, frequent choices being $d_{ij}^2 = \min(k_{t,i}^2, k_{t,j}^2)(\Delta y_{ij}^2 + \Delta\phi_{ij}^2)$ for the k_t algorithm [1], and $d_{ij}^2 = (\Delta y_{ij}^2 + \Delta\phi_{ij}^2)$ for the Cambridge-Aachen algorithm [2].

Cone algorithms make up the second class, where jets are defined as dominant directions of energy flow. One introduces the concept of *stable cone* as a circle of fixed radius R in the $y-\phi$ plane such that the sum of all the momenta of the particles within the cone points in the same direction as the centre of the circle. Cone algorithms attempt to identify all the stable cones. Most implementations use a seeded approach to do so: starting from one seed for the centre of the cone, one iterates until the cone is found stable. The set of seeds can be taken as the set of initial particles (sometimes over a p_t threshold) or as the midpoints between previously-found stable cones. As we shall see, this iterative method fails to identify *all* the stable cones, leading to infrared or collinear unsafety in the perturbative computations.

Cone algorithms can be split in two subclasses according to how they deal with the fact that stable cones may overlap. Cone algorithms with split-merge, identify the hardest overlapping pair of stable cones and merge (split) them if they share more (less) than a fraction f of the hardest cone. JetClu and midpoint are typical representatives of that subclass. Cone algorithms with progressive removal start with the hardest unclustered particle, iterate from there until a stable cone is found and call it a jet. Its contents are removed and one starts again with the remaining particles. The iterative cone is the typical example of a subclass with the particular feature that hard jets are fully conical.

The SNOWMASS accords have established a series of requirements that any jet algorithm has to fulfil. These are basically that one can use the algorithm for theoretical computations, *e.g.* it gives finite perturbative results, as well as for experimental purposes, *e.g.* it runs fast enough and has small corrections from hadronisation and the underlying event.

We illustrate in these proceedings [3] that midpoint and the iterative cone fail to give finite perturbative results. We introduce SISCone and the anti- k_t algorithms as solutions to those problems that do not spoil the experimental usability.

*Work done under contract No. DE-AC02-98CH10886 with the US Department of Energy.

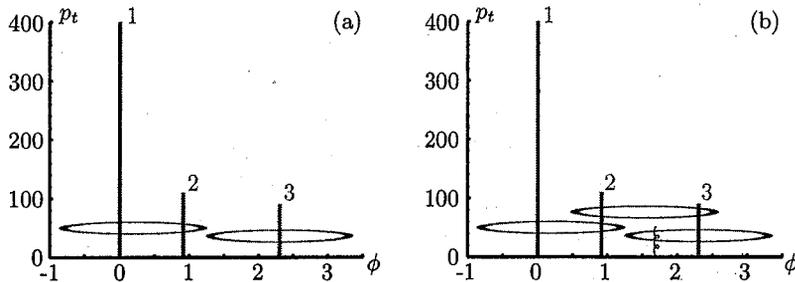


Figure 1: Stable cones found by the midpoint algorithm for a 3-particle event (left) and for the same event with an additional infinitely soft gluon (right).

SISCone as a replacement for the midpoint algorithm. Let us consider the 3-particle event displayed in Fig. 1(a). When clustered with the midpoint algorithm, 2 stable cones are found, leading to two jets: one with particles 1 and 2 and a second one with particle 3. If one adds to that hard event an infinitely soft gluon as shown in Fig. 1(b), a third stable cone is found and the three hard particles are clustered in a single jet. This change in the jet structure upon addition of soft particles, a phenomenon which happens with infinite probability in perturbative QCD, gives rise to divergences in the perturbative expansion and proves that the midpoint algorithm is infrared unsafe.

This problem arises from the fact that the seeded approach misses stable cones — here the one containing particles 2 and 3 in Fig. 1(a). The workaround to restore IR safety is thus to find a seedless method that provably identifies all the stable cones. This is notoriously complex: a naive approach testing the stability of all subsets of particles [4] has a complexity of order $N2^N$ for N particles which is much slower than the $\mathcal{O}(N^3)$ complexity of the midpoint algorithm, making this solution unusable for experimental purposes.

The solution [6] is to use the geometrical observation that any enclosure in the $y - \phi$ plane can be moved without changing its contents until it touches two points. Browsing all pairs of particles allows thus to enumerate all possible cones and to check their stability at an overall cost of $\mathcal{O}(N^3)$. Additional efforts can even bring the final complexity to $\mathcal{O}(N^2 \log(N))$ *i.e.* faster than the midpoint algorithm. This is illustrated on Fig. 2 where we observe that in practice SISCone runs faster than the typical implementations of the midpoint algorithm without a seed threshold and at least as fast as when a 1 GeV seed threshold is used.

This has been implemented [6, 7, 5] in a C++ code named SISCone (Seedless Infrared Safe Cone) which is the first cone algorithm to satisfy the SNOWMASS requirements, that is to be at the same time IR and collinear safe, and to be fast enough to be used in experimental analysis.

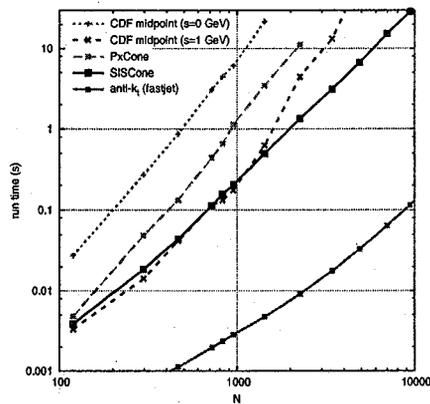


Figure 2: Clustering time for SISCone compared to typical implementations of the midpoint algorithm and the anti- k_t algorithm [5].

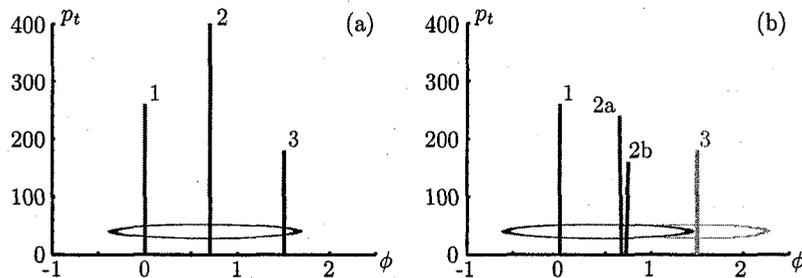


Figure 3: Jets found by the iterative cone for a 3-particle event (left) and for the same event with a collinear splitting (right).

Anti- k_t as a replacement for the iterative cone. As for the midpoint algorithm, we start with an event with three hard particles (see Fig. 3(a)). When clustered with the iterative cone, one stable cone containing all particles is found, resulting in a 1-jet event. If we now split the hardest particle into two collinear particles — a process that also has an infinite probability in perturbative QCD — as shown on Fig. 3(b), clustering with the iterative cone gives a first jet made of particle 1 plus the two collinear ones, then a second jet with particle 3. This example proves that the iterative cone algorithm is collinear unsafe.

Quite surprisingly, we can find a solution to that problem by coming back to the class of the recombination algorithms. The distance measures introduced earlier can be written as

$$d_{ij}^2 = \min(k_{t,i}^{2p}, k_{t,j}^{2p})(\Delta y_{ij}^2 + \Delta\phi_{ij}^2),$$

with $p = 1$ for the k_t algorithm and $p = 0$ for the Cambridge/Aachen algorithm. We can then consider a third case, the one for which $p = -1$ and call it the *anti- k_t algorithm* [8]. Obviously, this algorithm is IR and collinear safe. Furthermore, since its implementation [5] is similar to the one of the k_t algorithm, its speed will be similar too, which certainly makes it usable for experimental purposes as seen on Fig. 2.

To understand the link with the iterative cone algorithm, we note from the definition of the anti- k_t distance that pairs involving a hard particle will be given small distances. This means that soft particles are recombined with hard ones before recombining among themselves, resulting in regular, soft-resilient, hard jets. This is exactly the hallmark of the iterative cone and it is in that respect that the anti- k_t can be seen as an IR and collinear safe replacement.

To illustrate this property, we show in Fig. 4 the jets resulting from the clustering of an event made with a few hard particles and a large number of very soft ones uniformly distributed. It is clear that the hardest jets are perfectly circular and that, in general, the boundaries between the jets are regular.

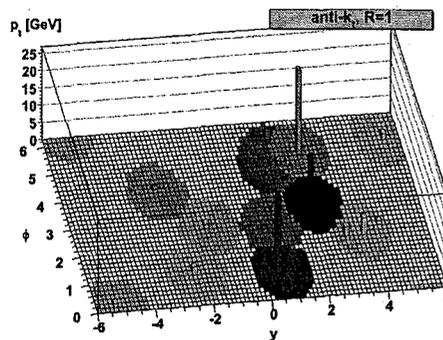


Figure 4: Illustration of the regularity of the jets obtained with the anti- k_t algorithm.

Observable	1st miss cones at	Last meaningful order
Inclusive jet cross section	NNLO	NLO
$W/Z/H + 1$ jet cross section	NNLO	NLO
3 jet cross section	NLO	LO (NLO in NLOJet)
$W/Z/H + 2$ jet cross sect.	NLO	LO (NLO in MCFM)
jet masses in 3 jets	LO	none (LO in NLOJet)

Table 1: Perturbative level at which IR or collinear unsafety arises for various processes.

Physical impact and discussion. As we have seen, the seeded approach to stable cone search suffers from problems w.r.t. perturbative QCD expansion: the algorithms with split-merge are IR unsafe, while the iterative cone (with progressive removal) is collinear unsafe. We have introduced SIScone as a natural replacement of the cone algorithms with split-merge like midpoint, and the anti- k_t algorithm as a candidate to replace the iterative cone. These new algorithms are both IR and collinear safe.

The question one might ask is to what extent these IR and collinear safety issues are important in real measurements. Since the unsafety arises when one has 3 particles in a common vicinity, it becomes important at the order α_s^4 or $\alpha_{EW}\alpha_s^3$ of the perturbative series.

Table 1 summarises for different physical processes, the order at which seeded algorithms stop to be valid. The main message we can get from that table is thus that, if we do not want theoretical efforts to be done in vain, the resort of an IR and collinear safe algorithm like SIScone and the anti- k_t is fundamental. To illustrate the argument more quantitatively, Fig. 5 shows the relative difference, expected to be present at the LO of perturbative QCD, between SIScone and midpoint for the mass of the 2nd hardest jet in 3-jet events. Differences reaching up to 40% are observed, proving that an IR and collinear safe algorithm is mandatory.

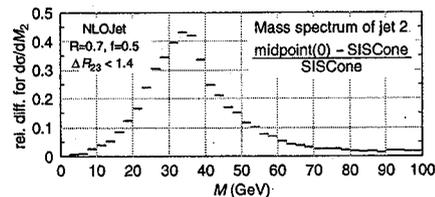


Figure 5: Relative difference between midpoint and SIScone for the mass of the 2nd hardest jet in 3-jet events. The 2nd and 3rd jets are imposed to be distant by at most $2R$.

References

- [1] S. Catani, Y. L. Dokshitzer, M. H. Seymour and B. R. Webber, Nucl. Phys. B **406** (1993) 187 and refs. therein; S. D. Ellis and D. E. Soper, Phys. Rev. D **48** (1993) 3160 [hep-ph/9305266].
- [2] Y. L. Dokshitzer, G. D. Leder, S. Moretti and B. R. Webber, JHEP **9708**, 001 (1997) [hep-ph/9707323]; M. Wobisch and T. Wengler, hep-ph/9907280.
- [3] Slides: <http://indico.cern.ch/contributionDisplay.py?contribId=299&sessionId=13&confId=24657>
- [4] G. C. Blazey *et al.*, hep-ex/0005012.
- [5] M. Cacciari and G. P. Salam, Phys. Lett. B **641** (2006) 57 [arXiv:hep-ph/0512210]. M. Cacciari, G. P. Salam and G. Soyez, FastJet, <http://www.lpthe.jussieu.fr/~salam/fastjet>.
- [6] G. P. Salam and G. Soyez, JHEP **0705** (2007) 086 [arXiv:0704.0292 [hep-ph]].
- [7] G. P. Salam and G. Soyez, SIScone, <http://projects.hepforge.org/siscone>.
- [8] M. Cacciari, G. P. Salam and G. Soyez, JHEP **0804** (2008) 063 [arXiv:0802.1189 [hep-ph]].