Flavoured jet algorithms

Giovanni Stagnitto

LHCP 2023, Belgrade, 22-26.05.2023
(Usual) experimental definition of flavoured jet

“An (anti-\(k_t\)) jet is flavoured if it contains at least one heavy hadron within \(\Delta R < R\) with \(p_T > p_{T,\text{cut}}\)”

This definition is adopted as “true” label in MC samples.

These samples are then used to train ML architectures (“high-level taggers”), which exploit low-level variables as inputs.
(Usual) experimental definition of flavoured jet

“An (anti-\(k_t\)) jet is flavoured if it contains at least one heavy hadron within \(\Delta R < R\) with \(p_T > p_{T,\text{cut}}\)”

This definition is both soft and collinear (IRC) unsafe
(in massless perturbative QCD calculations)

i.e. arbitrary soft and/or collinear emissions alter the flavour of jets
(Usual) experimental definition of flavoured jet

“An (anti-\(k_t\)) jet is flavoured if it contains at least one heavy hadron within \(\Delta R < R\) with \(p_T > p_{T,\text{cut}}\)”

\[ g \rightarrow q\bar{q} \text{ is always flavoured even in the collinear limit} \]

An even-tag veto in calculations is enough to fix this issue
(Usual) experimental definition of flavoured jet

“An (anti-\(k_t\)) jet is flavoured if it contains at least one heavy hadron within \(\Delta R < R\) with \(p_T > p_{T,\text{cut}}\)”

\[
q \rightarrow qg \text{ collinear with a hard gluon leads to a flavourless jet}
\]

With \(p_{T,\text{cut}}\), it requires a fragmentation function, as we are identifying a particle

Without \(p_{T,\text{cut}}\), any IRC safe flavour-agnostic algorithm will recombine the \(qg\) pair
(Usual) experimental definition of flavoured jet

“An (anti-\(k_t\)) jet is flavoured if it contains at least one heavy hadron within \(\Delta R < R\) with \(p_T > p_{T,\text{cut}}\)”

\[
\begin{align*}
\hat{\sigma} & \quad p_\ell \\
& \quad p_{\bar{\ell}} \\
& \quad p_q \\
& \quad p_{\bar{q}} \\
& \quad p_j
\end{align*}
\]

Soft large-angle \(g \rightarrow b\bar{b}\) polluting the flavour of other jets

No way of fixing this issue within a flavour-agnostic jet algorithm!
**Solution: the flavour-$k_t$ algorithm**

[Banfi, Salam, Zanderighi (hep-ph/0601139)]

**Flavour-aware distance:**

$$d_{ij}^{(F, \alpha)} = \frac{\Delta y_{ij}^2 + \Delta \phi_{ij}^2}{R^2} \times \begin{cases} \max(k_{ti}, k_{tj})^\alpha \min(k_{ti}, k_{tj})^{2-\alpha}, & \text{softer of } i, j \text{ is flavoured,} \\ \min(k_{ti}^2, k_{tj}^2), & \text{softer of } i, j \text{ is flavourless} \end{cases}$$

at the price of jets with different kinematics i.e. not anti-$k_t$ jets.

Comparison with experimental data not straightforward

[Gauld et al. (2005.03016)]

[Czakon et al. (2011.01011)]
In the past year, several alternative proposals!

[Caletti, Larkoski, Marzani, Reichelt (2205.01109)]
[Caletti, Larkoski, Marzani, Reichelt (2205.01117)]
[Czakon, Mitov, Poncelet (2205.11879)]
[Caola, Grabarczyk, Hutt, Salam, Scyboz, Thaler (to appear)]
[Gauld, Huss, GS (2208.11138)]

I will briefly introduce them, by then focusing on the last one.
Use Soft Drop to remove soft quarks, by using JADE as reclusters.

Flavour of jet = flavour of particle(s) lying along the Winner-Take-All (WTA) axis.

Soft safe, but **collinear unsafe**: requires usage of suited fragmentation functions.

---

this system has the smallest invariant mass and passes SD

soft quark can alter the flavour
“Flavour anti-$k_t$”: modify anti-$k_t$ distance when flavoured particles involved

\[ d_{ij} = R^2 \min(k_{T,i}^{-2}, k_{T,j}^{-2}) \cdot S_{ij}^a, \quad d_B = k_{T,i}^{-2} \]

where \( S_{ij} \neq 1 \) only when \( i \) and \( j \) are of opposite flavour

\[ S_{ij}^a = 1 - \theta(1 - \kappa)\cos\left(\frac{\pi}{2} \kappa\right), \quad \kappa = \frac{1}{a} \frac{k_{T,i}^2 + k_{T,j}^2}{2 k_{T,max}^2} \]

One recovers (IRC flavour unsafe) anti-$k_t$ jets when \( a \to 0 \)
“Flavour neutralisation”

neutralise \equiv \text{remove the (opposite) flavour of both 1 \& 2 while maintaining kinematics}

\[
\begin{align*}
    u_{ik} &= \max \left( p_{ti}^2, p_{tk}^2 \right)^p \min \left( p_{ti}^2, p_{tk}^2 \right)^q \\
    &\times 2 \left[ \frac{1}{a^2} \left( \cosh(a \Delta y_{ik}) - 1 \right) - \left( \cos \Delta \phi_{ik} - 1 \right) \right]
\end{align*}
\]

from Ludovic Scyboz slides at Moriond QCD 2023
“Flavour dressing”
Flavour assignment factorised from jet reconstruction (exact anti-$k_t$ kinematics by construction)

**Inputs**: flavour-agnostic jets (jets obtained with any IRC safe algorithm) and flavour inputs (e.g. b- or c-quarks, stable heavy-flavour hadrons, …)

**Preliminary step**: we first build flavour clusters to recombine flavour inputs with radiation close in angle, but without touching the soft particles (thanks to a Soft Drop condition [Larkoski, Marzani, Soyez, Thaler 1402.2657]):

$$\frac{\min(p_{t,a}, p_{t,b})}{(p_{t,a} + p_{t,b})} > z_{cut} \left( \frac{\Delta R_{ab}}{\delta R} \right)^\beta$$

**Dressing step**: in order to assign flavour to jets, we run a sequential recombination algorithm with flavour-$k_t$-like distances between jets and flavour clusters.
IRC safety test in $e^+e^- \rightarrow$ jets

= vanishing “bad” identification of flavours in the fully unresolved regime

only soft and/or collinear radiation

Any gen-$k_t$ algo is safe (no additional flavour in the event)
IRC safety test in $e^+e^- \rightarrow$ jets

= vanishing “bad” identification of flavours in the fully unresolved regime

only soft and/or collinear radiation

Naive dressing unsafe, flavour dressing safe!
IRC safety test in $e^+e^- \to$ jets

= vanishing “bad” identification of flavours in the fully unresolved regime

only soft and/or collinear radiation

![Graph showing IRC safety test results]

Naive dressing unsafer, flavour dressing still safe!
Systematic IRC safety tests

Numerical framework developed by Caola et al. has allowed to discover potentially problematic configurations at higher orders (CMP = “flavour anti-$k_t$”; GHS = “flavour dressing”)

→ as for GHS, work in progress to fix them
In principle, massive calculations do not require an IRC safe flavour algorithm (screening effect due to $m_q$).

However, presence of large logarithms $\log(Q^2/m_q^2)$, spoiling the convergence of the perturbative series ($\alpha_s \log(m_Z^2/m_c^2) \sim 1$).

**Benefits of massless calculations with IRC safe jet tagging:**
- in the initial-state, a massless calculation allows for a resummation of $\log(Q^2/m_q^2)$ by PDF evolution (crucial in some cases e.g. when probing non-perturbative charm PDF)
- in the final-state, an IRC safe prescription implies a suppressed sensitivity on $\log(Q^2/m_q^2)$, both in fixed order and resummed calculations / parton showers.
Test flavour dressing in a realist scenario: $Z + b$-jet

[same setup of Gauld, Gehrmann-De Ridder, Glover, Huss, Majer (2005.03016)]

Remarkable agreement between (N)NLO and NLO+PS
→ for most distributions
largely insensitive to all-order corrections
First new result with flavour dressing:

$Z + c$-jet at LHCb

Measurement sensitive to intrinsic charm in the proton

LHCb data at 13TeV for ratio

$\frac{d\sigma_{Z+c}}{dy_Z} / \frac{d\sigma_{Z+j}}{dy_Z}$ [2109.08084]

(With flavour dressing, both the numerator and the denominator feature the same sample of anti-$k_t$ jets!)
**Ratio** $\sigma(Z + c \jet) / \sigma(Z + \jet)$ at **NNLO**

[Gauld, Gehrmann-De Ridder, Glover, Huss, Rodriguez Garcia, GS (2302.12844)]

**NNLO** lies between **NLO**+PS predictions with different PS, but **reduction of theory uncertainties** by a factor of 2. Similar for other distributions.
Final remarks

- **At lot of recent proposals** trying to solve the longstanding issue of a proper definition of flavoured jet

- **IRC-safe definition** allows for massless fixed-order calculations to be directly compared to experimental data (and a suppressed sensitivity on mass logarithms)

- A comparison between the different approaches would be beneficial, as well as a study of their experimental feasibility
BACKUP
## LHCb fiducial cuts

Very unique fiducial region of the measurement:

<table>
<thead>
<tr>
<th>$Z$ bosons</th>
<th>$p_T^Z &gt; 20$ GeV, $2.0 &lt; \eta(\mu) &lt; 4.5$, $60 &lt; m(\mu^+\mu^-) &lt; 120$ GeV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jets</td>
<td>$20 &lt; p_T^j &lt; 100$ GeV, $2.2 &lt; \eta(j) &lt; 4.2$</td>
</tr>
<tr>
<td>Charm jets</td>
<td>$p_T(c \text{ hadron}) &gt; 5$ GeV, $\Delta R(j, c \text{ hadron}) &lt; 0.5$</td>
</tr>
<tr>
<td>Events</td>
<td>$\Delta R(\mu, j) &gt; 0.5$</td>
</tr>
</tbody>
</table>
LHCb fiducial cuts

Very unique fiducial region of the measurement:

<table>
<thead>
<tr>
<th>Z bosons</th>
<th>$p_T(\mu) &gt; 20 \text{ GeV}, 2.0 &lt; \eta(\mu) &lt; 4.5, 60 &lt; m(\mu^+\mu^-) &lt; 120 \text{ GeV}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jets</td>
<td>$20 &lt; p_T(j) &lt; 100 \text{ GeV}, 2.2 &lt; \eta(j) &lt; 4.2$</td>
</tr>
<tr>
<td>Charm jets</td>
<td>$p_T(c \text{ hadron}) &gt; 5 \text{ GeV}, \Delta R(j, c \text{ hadron}) &lt; 0.5$</td>
</tr>
<tr>
<td>Events</td>
<td>$\Delta R(\mu, j) &gt; 0.5$</td>
</tr>
</tbody>
</table>

We explore a theory-driven cut:

$$p_T(Z + \text{jet}) < p_{T,jet}$$

At Born level, the $p_T$ of the $Z+$jet system vanishes, hence the cut limits the hard QCD radiation outside the LHCb acceptance in a dynamical way.
We refrain from making a comparison to the LHCb data

1) definition of flavoured jet not IRC safe

2) significant contamination from MPI
Results: $p_T^{c-jet}$

[Gauld, Gehrmann-De Ridder, Glover, Huss, Rodriguez Garcia, GS (2302.12844)]

\[ \frac{\mathrm{d}\sigma}{\mathrm{d}p_T^{c-jet}} \]

\( \sqrt{s} = 13 \text{ TeV} \)

LHCb cuts, PDF4LHC21

Flavour dressing

\[ \frac{\sigma}{\sigma_{\text{NLO}}} \]

\[ \frac{\sigma}{\sigma_{\text{NNLO}}} \]

LHCb cuts, PDF4LHC21

Flavour dressing

\( p_T^{Z + \text{jet}} < p_T^{\text{jet}} \)

Ratio to NLO

Ratio to NNLO

<table>
<thead>
<tr>
<th>$p_T^{c-jet}$ [fb]</th>
<th>LO</th>
<th>NLO</th>
<th>NNLO</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>10</td>
<td>1.02</td>
<td>1.02</td>
<td>1.02</td>
</tr>
<tr>
<td>20</td>
<td>1.04</td>
<td>1.04</td>
<td>1.04</td>
</tr>
<tr>
<td>30</td>
<td>1.06</td>
<td>1.06</td>
<td>1.06</td>
</tr>
<tr>
<td>40</td>
<td>1.08</td>
<td>1.08</td>
<td>1.08</td>
</tr>
<tr>
<td>50</td>
<td>1.10</td>
<td>1.10</td>
<td>1.10</td>
</tr>
<tr>
<td>60</td>
<td>1.12</td>
<td>1.12</td>
<td>1.12</td>
</tr>
<tr>
<td>70</td>
<td>1.14</td>
<td>1.14</td>
<td>1.14</td>
</tr>
<tr>
<td>80</td>
<td>1.16</td>
<td>1.16</td>
<td>1.16</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$p_T^{c-jet}$ [fb]</th>
<th>NLO + Py8</th>
<th>NLO + Hw7</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>10</td>
<td>1.02</td>
<td>1.02</td>
</tr>
<tr>
<td>20</td>
<td>1.04</td>
<td>1.04</td>
</tr>
<tr>
<td>30</td>
<td>1.06</td>
<td>1.06</td>
</tr>
<tr>
<td>40</td>
<td>1.08</td>
<td>1.08</td>
</tr>
<tr>
<td>50</td>
<td>1.10</td>
<td>1.10</td>
</tr>
<tr>
<td>60</td>
<td>1.12</td>
<td>1.12</td>
</tr>
<tr>
<td>70</td>
<td>1.14</td>
<td>1.14</td>
</tr>
<tr>
<td>80</td>
<td>1.16</td>
<td>1.16</td>
</tr>
</tbody>
</table>
Results: $y^Z$

[Gauld, Gehrmann-De Ridder, Glover, Huss, Rodriguez Garcia, GS (2302.12844)]
Test in a realist scenario: $Z + b$-jet

[same setup of Gauld, Gehrmann-De Ridder, Glover, Huss, Majer (2005.03016)]

Some sensitivity observed in $p_T^Z$, likely due to:

Even if IRC finite, it leads to large migration of (unflavoured)-jet into the $b$-jet sample.
Test in a realist scenario: $Z + b$-jet

[same setup of Gauld, Gehrmann-De Ridder, Glover, Huss, Majer (2005.03016)]

Some sensitivity observed in $p_T^Z$, likely due to:

Effect captured at NNLO
Results: $\eta^{c-jet}$

[Gauld, Gehrmann-De Ridder, Glover, Huss, Rodriguez Garcia, GS (2302.12844)]

\[\eta^{c-jet}\]
The flavour-\(k_t\) algorithm

[Banfi, Salam, Zanderighi (hep-ph/0601139)]

1. Introduce a distance measure \(d_{ij}^{(F)}\) between every pair of partons \(i, j\):

\[
d_{ij}^{(F,\alpha)} = (\Delta \eta_{ij}^2 + \Delta \phi_{ij}^2) \times \begin{cases} 
\max(k_{ti}, k_{tj})^{\alpha} \min(k_{ti}, k_{tj})^{2-\alpha}, & \text{softer of } i, j \text{ is flavoured,} \\
\min(k_{ti}^2, k_{tj}^2), & \text{softer of } i, j \text{ is flavourless,}
\end{cases}
\]

as well as distances to the two beams,

\[
d_{iB}^{(F,\alpha)} = \begin{cases} 
\max(k_{ti}, k_{tB}(\eta_i))^{\alpha} \min(k_{ti}, k_{tB}(\eta_i))^{2-\alpha}, & i \text{ is flavoured,} \\
\min(k_{ti}^2, k_{tB}(\eta_i)^2), & i \text{ is flavourless,}
\end{cases}
\]

and an analogous definition of \(d_{iB}^{(F,\alpha)}\) involving \(k_{tB}(\eta_i)\) instead of \(k_{tB}(\eta_i)\) (both defined as in eqs. (11) and (12)). As in section 2 we have introduced a class of measures, parametrised by \(0 < \alpha \leq 2\).

2. Identify the smallest of the distance measures. If it is a \(d_{ij}^{(F,\alpha)}\), recombine \(i\) and \(j\); if it is a \(d_{iB}^{(F,\alpha)}\) (\(d_{iB}^{(F,\alpha)}\)) declare \(i\) to be part of beam \(B\) (\(\bar{B}\)) and eliminate \(i\); in the case where the \(d_{iB}^{(F,\alpha)}\) and \(d_{iB}^{(F,\alpha)}\) are equal (which will occur if \(i\) is a gluon), recombine with the beam that has the smaller \(k_{tB}(\eta_i), k_{tB}(\eta_i)\).

3. Repeat the procedure until all the distances are larger than some \(d_{\text{cut}}\), or, alternatively, until one reaches a predetermined number of jets.\(^{10,11}\)

Modified beam distance:

\[
k_{tB}(\eta) = \sum_i k_{ti} \left( \Theta(\eta - \eta_i) + \Theta(\eta_i - \eta) \right) e^{\eta_i - \eta}
\]

IRC flavour safe to all orders, but different kinematics (because new distance)
The **flavour dressing algorithm**

[Gauld, Huss, GS (2208.11138)]

Flavour assignment *factorised* from jet reconstruction: we assign flavour to flavour-agnostic jets in an IRC safe way

Inputs:
flavour agnostic jets \(\{j_k\}\), flavoured clusters \(\{\hat{f}_i\}\), association criterion, accumulation criterion

Run a sequential recombination algorithm with flavour-\(k_i\)-like distances:
- \(d(\hat{f}_i, \hat{f}_j)\) between flavoured clusters;
- \(d(\hat{f}_i, \hat{j}_k)\) if flavoured cluster \(\hat{f}_i\) associated to jet \(j_k\)
- \(d_B(\hat{f}_i)\) if \(\hat{f}_i\) not associated to any jet

Finally, assign flavour to jet \(j_k\) according to collected \(\text{tag}_k\) and *accumulation* criterion
The flavour dressing algorithm: inputs

- **Flavour agnostic jets** \( \{ j_k \} \): set of jets obtained with an IRC safe jet algorithm (e.g. gen-\( k_t \) family), possibly after a fiducial selection.

- **Flavoured clusters** \( \{ \hat{f}_i \} \)

- **Association criterion**

- **Accumulation criterion**
The **flavour** dressing algorithm: inputs

- **Flavour agnostic jets** \( \{ j_k \} \)

- **Flavoured clusters** \( \{ \hat{f}_i \} \): built out of quarks (e.g. c, b) or stable heavy-flavour hadrons (e.g. D, B), by *dressing them with radiation close in angle, but without touching the soft particles*.

Exploiting the Soft Drop criterion [Larkoski, Marzani, Soyez, Thaler 1402.2657]

"Naked" flavoured objects are collinear unsafe

\[
\frac{\min(p_{t,a},p_{t,b})}{(p_{t,a} + p_{t,b})} > z_{\text{cut}} \left( \frac{\Delta R_{ab}}{\delta R} \right)^\beta
\]

- **Association criterion**

- **Accumulation criterion**
The *flavour* dressing algorithm: inputs

- *Flavour agnostic jets* \(\{j_k\}\)

- *Flavoured clusters* \(\{\hat{f}_i\}\)

- **Association criterion**: whether \(\hat{f}_i\) is “associated” to \(j_k\)
  
  At parton-level simply if \(\hat{f}_i\) is a constituent of \(j_k\)

  Other options: \(\Delta R(\hat{f}_i, j_k) < R_{\text{tag}}\), ghost association, …

  Flavour assignment based only on association is soft unsafe

- **Accumulation criterion**
The **flavour** dressing algorithm: inputs

- **Flavour agnostic jets** \( \{ j_k \} \)
- **Flavoured clusters** \( \{ \hat{f}_i \} \)
- **Association criterion**
- **Accumulation criterion**: how to “sum” flavours
  - sum flavoured if unequal number of \( f \) and \( \bar{f} \) (need charge information)
  - sum flavoured if odd number of \( f \) or \( \bar{f} \) (if no charge information)
Definition of flavoured cluster $\hat{f}_i$

1. Initialise a set with all the flavourless objects $p_i$ (particles used as input to jets) and all the flavoured objects $f_i$ (bare flavours), avoiding double counting if necessary.

2. Find the pair with the smallest angular distance $\Delta R_{ab}$:
   - flavourless $p_a, p_b$: combine $p_a$ and $p_b$ into a flavourless $p_{ab}$;
   - flavoured $f_a, f_b$: remove both from the set;
   - flavoured $f_a$, unflavoured $p_b$: remove $p_b$ from the set and check a Soft Drop criterion
     \[
     \frac{\min(p_{t,a}, p_{t,b})}{(p_{t,a} + p_{t,b})} > z_{\text{cut}} \left( \frac{\Delta R_{ab}}{\delta R} \right)^\beta
     \]
     to **recombine collinear while preserving soft**. [default: $\delta R = 0.1$, $z_{\text{cut}} = 0.1$, $\beta = 2$]
     If satisfied, combine $f_a$ and $p_b$ into a flavoured $f_{ab}$.

3. Iterate while there are at least two objects in the set until $\Delta R_{ab} > \delta R$.
   The momentum of $\hat{f}_i$ is given by the accumulated momentum into $f_i$. 
IRC sensitivity in $2 \to 2$ QCD events in $pp$

only soft and/or collinear radiation

Flavour dressing approaches zero faster than a naive flavour tagging as $y_3^{k_t} \to 0$