It can be shown that the matrix, M, of the transformation is orthogonal $M^T = M^{-1}$ and $\det M = 1$, which defines a rotation in Euclidean space.

B Dictionary of variables

B.1 Vertex isolation

We take the mother particle of the decay and add other tracks to its daughters 1 by 1, building a new vertex χ^2 (see Fig. 91).

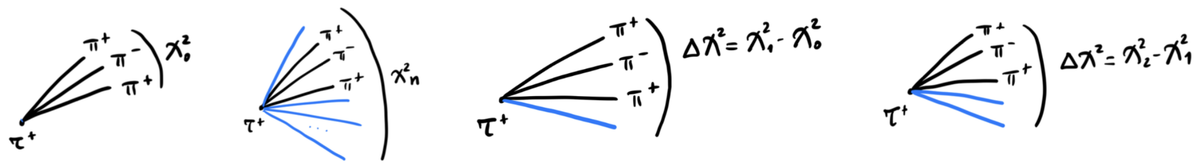

Figure 91: Schematic representation of the vertex isolation method.

The following variables are built:

- VTXISONUMVTX: number of other tracks in the event for which the new vertex fit $\chi^2 < 9$;
- VTXISODCHI2MASSONETRACK: smallest difference in χ^2 when adding 1 track;
- VTXISODCHI2MASSTWOTRACK: smallest difference in χ^2 when adding 1 track to the combination that had the smallest $\Delta\chi^2$ when adding 1 track;
- VTXISODCHI2MASSONETRACK: invariant mass of the tracks used to build the VTXISODCHI2MASSONETRACK variable;
- VTXISODCHI2MASSTWOTRACK: invariant mass of the tracks used to build the VTXISODCHI2TWOTRACK variable.

B.2 Cone isolation

An isolation cone is built around a head particle from tracks whose angular distance from the head particle is, $R = \sqrt{\eta^2 + \phi^2} = 0.5, 1.0, 1.5, 2.0$, where ϕ is the azimuthal angle and η is the pseudorapidity. From the isolation cone all of the particle's daughters are excluded (see Fig. 92) and isolation variables are built with the remaining particles. Charged-cone (CC) variables are concerned with the charged tracks within the isolation cone while neutral-cone (NC) variables are concerned with the neutral objects inside this cone.

The following variables are used:

- DELTAETA, DELTAPHI: $\Delta\eta$ and $\Delta\phi$ between the vector sum momentum of the isolation cone tracks and the head of the cone. For a cone size of 0.5, the DELTAETA variable peaks at 0 and at around 3. The peak around 3 is caused by cases in which the isolation cone is empty, for which DELTAETA is simply the eta of the head

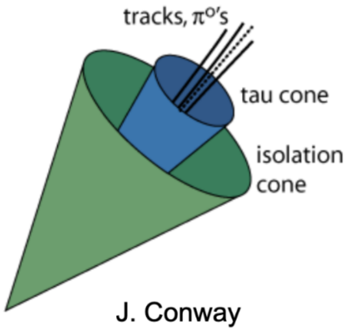

Figure 92: Schematic representation of the isolation cone.

particle. The eta of the head particle peaks around 3. For larger cone sizes, the peak at 3 is reduced because the probability of having an empty isolation cone decreases.;

- IT: $\frac{p_{T,head}}{p_{T,head+cone}}$. When the isolation cone is empty, IT is 1.;
- MAXPT_E,PT,PX,PY,PZ,Q: properties of the object inside the isolation cone with the maximum pT;
- MULT: multiplicity of the objects inside the cone;
- PASYM,PTASYM,PXASYM,PYASYM,PZASYM: $\frac{p_{head}-p_{cone}}{p_{head}+p_{cone}}$, where p is replaced by pT, p_x, p_y and p_z, accordingly. When the isolation cone is empty, the PASYM variables are 1.;
- PX,PY,PZ: total momentum of the objects in the isolation cone;
- SPT: scalar summed pT of the objects inside the isolation cone;
- VPT: vector summed pT of the objects inside the isolation cone.

B.3 Track isolation BDT

Track isolation discriminates signal against partial reconstructed background by checking that the underlying tracks in each event are not coming from a selected candidate.

Tracks in each event are divided into 3 categories which are schematically represented in Fig. 93:

- Selected tracks: tracks coming from a selected candidate (red);
- Isolating tracks: any track that is not coming from a selected candidate (black);
- Non-isolating tracks: tracks that are coming from the same vertex as the selected candidate (purple)

Track isolation is implemented with machine learning techniques: a BDT is trained with geometric variables that help discriminating isolating from non-isolating tracks. The BDT is trained on the $B_s \to K\mu\nu$ decay with 5 input features:

- $FC = \frac{p_{Niso\_track+\mu/K}\ \theta}{p_{Niso\_track+\mu/K}\ \theta + p_{T,Niso\_track} + p_{T,\mu/K}}$;

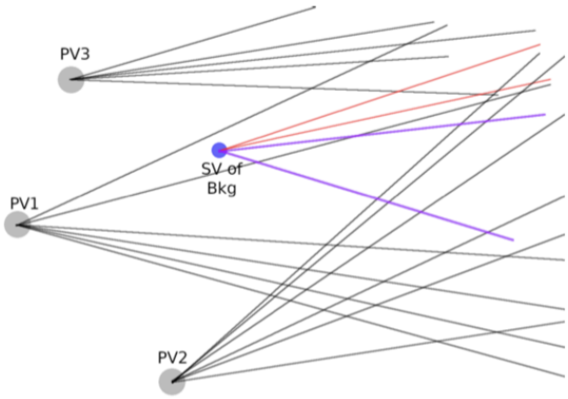

Figure 93: Schematic representation of the track definition in the track isolation tool.

- θ: angle between the non-isolating track and the muon/kaon track;
- LOG(DOCA): logarithm of the distance of closest approach (DOCA) between the non-isolating track and the muon/kaon track;
- PVDIS: distance between the PV and the non-isolating track vertex;
- distance between the vertex of the mother of the non-isolating track and the PV.

The BDT is trained on 3 samples with 3 different backgrounds:

- $Bs\mu\nu \to D_s\mu\nu \to KK\pi\mu\nu$ (1);
- $B^{\pm} \to J/\psi K^{\pm} \to \mu\mu K^{\pm}$ (2);
- $B_s\mu\nu \to K^{*+}\mu^-\nu \to K^+\pi^0\mu^-\nu$ (3).

The BDT is then applied to the 7 tracks of our decay (6 pions + kaon) to check whether there are non-isolating tracks coming from the same vertex as the signal tracks. The following variables are created:

- Bp_TRKISOBDTFIRSTVALUE: BDT output when using background sample 1;
- Bp_TRKISOBDTSECONDVALUE: BDT output when using background sample 2;
- Bp_TRKISOBDTTHIRDVALUE: BDT output when using background sample 3.

B.4 Vertex isolation BDT

A BDT is trained to evaluate how isolated the daughter tracks of a certain particle are from other tracks in the event. The BDT is trained on 3 different decays:

- $\Lambda_c^* \to \Lambda_c\pi^+\pi^-$ (1);
- $B^+ \to D^*\tau^+$ (2);
- $\Lambda_b \to p\mu$ (3);

where the Λ_c daughters, the D^* daughters and the proton form the candidate in cases (1), (2), and (3), respectively, and the $\pi^+\pi^-$, the τ^+ and the μ form the extra track, respectively. The background sample is form by every other track n the event. The BDT can either be trained on hard or soft extra track(s).

The following variables are used as input features in the BDT:

- $1-(1-\cos\theta)^{0.2}$, where θ is the opening angle between the candidate momentum and the extra track(s). The transformation, $x \to (1-x)^{0.2}$ is done to avoid the peak at 1 in the $\cos\theta$ distribution;
- $\log(p_T^{track})$, where p_T^{track} is the transverse momentum of the extra track(s);
- $\log(\min\chi^2_{IP\_PV})$, where $\chi^2_{IP\_PV}$ is the impact parameter χ^2 of the extra track(s) with respect to the PV. The minimum is taken from all the PVs in the event;
- $\log(\chi^2_{IP\_SV})$, where $\chi^2_{IP\_SV}$ is the impact parameter χ^2 of the extra track(s) with respect to the candidate's mother decay vertex;
- $\log(\Delta\chi^2_{FD\_PV})$, where $\Delta\chi^2_{FD\_PV}$ is the difference in the candidate's mother flight distance from the PV χ^2 with and without the extra track(s);
- $\Delta R = \sqrt{(\Delta\eta)^2+(\Delta\phi)^2}$, where ΔR is the angular distance between the candidate and the extra track(s). $\Delta R^{0.2}$ is used to avoid the peak at 0;
- The flight cosine variable defined as $FC = \frac{|\vec{p}_{track}+\vec{p}_{candidate}|\alpha}{|\vec{p}_{track}+\vec{p}_{candidate}|\alpha+p_{T,candidate}+p_{T,track}}$, where α is the angle between the candidate's flight direction and the total momentum of the candidate and the extra track(s).

The BDT output is computed for each decaying particle in the tau decay (candidate): B^+, τ^+, τ^-, forming the variables:

- Bp_VTXISOBDTHARDFISRTVALUE: BDT output when using sample (1);
- Bp_VTXISOBDTHARDSECONDVALUE: BDT output when using sample (2);
- Bp_VTXISOBDTHARDTHIRDVALUE: BDT output when using sample (3).

B.5 Tau isolation BDT

A BDT is trained on $B^0 \to K^{*0}\tau^+\tau^-$ events giving separation between signal tracks (ST) (from the $B^0 \to K^{*0}\tau^+\tau^-$ decay) and non-signal tracks (NST) (all other tracks in the event). The BDT was originally created for the $B^0_s \to \mu^+\mu^-$ analysis, later to the $B^0_s \to \tau^+\tau^-$ analysis and finally it was re-optimized for the $B^0 \to K^{*0}\tau^+\tau^-$ analysis. For $B^0 \to K^+\tau^+\tau^-$, we used the output of the BDT trained on $B^+ \to K^{*0}\tau^+\tau^-$, whereas for the normalisation channel we use the output of the BDT trained on $B^0_s \to \tau^+\tau^-$. The NST are divided in non-isolating tracks, i.e. tracks coming from a decay vertex that is part of the signal's decay chain, and isolating tracks, i.e. tracks coming from other vertices in the event.

For each NST, a common vertex with the ST, V, is defined as the midpoint along the line of closest approach of the two tracks. For the BDT, the following input features are used:

- $\min d/\sigma_d$, where d is the distance between the NST and the PV and σ_d is the corresponding uncertainty. The minimum is found among all the PVs in the event;
- p_T of the NST;
- angle between the NST and the ST;
- $fc = \frac{|p^{ST}+p^{NST}|\alpha}{|p^{ST}+p^{NST}|\alpha+p_T^{ST}+p_T^{NST}}$, where α is the angle between the total momentum of the NST and the ST and the direction between the PV and V;
- the distance of closest approach (DOCA) between the NST and the ST;
- the distance between V and the B^+ decay vertex;
- the distance between V and the PV.

The isolating tracks are used as the signal proxy and the non-isolating tracks are used as the background proxy. Tracks with high BDT values (isolating tracks) are likely to come from the signal candidate whereas tracks with low BDT values are likely to come from background.

a, b and c are defined as the number of tracks in the event with $BDT < -0.09$, $BDT < -0.05$ and $BDT < 0$, respectively. The following variables are created for τ^+ and τ^- in $B^+ \to K^+\tau^+\tau^-$ / for $\bar{D}^0$ and D_s^+ in the normalisation channel:

- ISOBDTFIRSTVALUE: $a + 100b + 1000c$;
- ISOBDTSECONDVALUE: sum of the BDT values for all tracks with $BDT < -0.05$;
- ISOBDTTHIRDVALUE: sum of ISOBDTSECONDVALUE and the minimum of the BDT values of all tracks in the event.