Primary Vertex Reconstruction at LHCb

M. Kucharczyk$^{1,2}$, P. Morawski$^3$, M. Witek$^1$.

$^1$Henryk Niewodniczanski Institute of Nuclear Physics PAN, Krakow, Poland
$^2$Sezione INFN di Milano Bicocca, Milano, Italy
$^3$AGH - University of Science and Technology, Faculty of Physics and Applied Computer Science, Krakow, Poland

Abstract

The paper describes the algorithm for the primary vertex finding in the offline reconstruction of the LHCb data. The reconstruction is performed in two steps: seeding and fitting. In the first step the primary vertex candidates are selected by searching for the space points where an accumulation of track trajectories may be observed. In the second step the weighted least square method is employed to find the final vertex position. The performance of the primary vertex reconstruction is presented.
1 Introduction

The forward geometry of the LHCb spectrometer and hadronic collisions at LHC result in a relatively high density of tracks observed in the detector. Accurate estimation of the proton-proton interaction point, the primary vertex (PV), is essential to conduct precise measurements in the LHCb experiment [1]. The events with $b\bar{b}$ pair can be distinguished from the large background of light quarks production by observation of decay vertices of b-hadrons displaced by a few millimeters from the corresponding primary vertex. The properties of b-hadron decay products are used to efficiently suppress the background: relatively large impact parameter IP (defined as a minimum distance of the track to PV) and relatively large transverse momentum $p_T$ with respect to the beam axis. In addition the compatibility between the momentum vector of b-hadron and its flight direction determined from the production and decay vertices is checked.

Before collecting first data, the primary vertex reconstruction was developed using simulated events. The quality of real data and detailed data taking conditions were not known. Therefore, the main assumption was to deliver a bias free and robust algorithm in terms of both efficiency and precision. Since the beam spot stability was unknown before the LHC start, the requirements was to run the PV finding algorithm without any prior estimation of the PV position. In addition any bias related to the inclusion of tracks originated from the long lived particles should be kept minimal. Another issue is the multiple proton-proton interactions. The optimal working point has to be selected to compromise the contradictory requirements: high efficiency and low rate of false PV. This is also related to the reconstruction of low multiplicity PV and a separation between two PVs. Moreover an attention has to be paid to prevent the reconstruction of 5 and 6 prong decays of b-hadrons which may be classified as a PV and therefore rejected in the analysis. Various samples of fully simulated events and data collected at $\sqrt{s} = 7$ TeV has been used to determine the performance of the reconstruction algorithm.

2 Primary vertex reconstruction

The primary vertex finding is based on tracks [2] reconstructed in the LHCb detector. It consists of two subsequent steps: seeding and fitting, which are described in Sec. 2.1 and Sec. 2.2. Reconstructed seeds are sorted according to descending multiplicity. In this way high multiplicity PVs are reconstructed first. This is particularly useful to reduce incorrect reconstruction of secondary vertices as primary vertices. Otherwise, the low multiplicity secondary vertex could absorb tracks from the corresponding PV acquiring enough tracks to fulfill a minimum PV multiplicity condition. The PV fit is performed for every seed. The successfully fitted vertex is demanded to be separated from previously reconstructed ones. Loose separation conditions were used, $|\Delta z|/\sigma_{\Delta z} > 9$ if multiplicity of the closest PV is grater than 9 and $|\Delta z|/\sigma_{\Delta z} > 81$ otherwise, where $\Delta z$ denotes the difference between $z$ coordinates of two primary vertices. This loose separation does not protect against a large false PV rate for events with a high number of proton-proton interactions. For such events the probability to form a false PV of high multiplicity ($\sim 10$)
with a combination of tracks originating from two close PVs is relatively high. This is because the tracks with $\chi^2_{IP} < 9$ are assigned to the vertex, while the others, worse measured halo tracks, are not. The $\chi^2_{IP}$ is defined as the increase in $\chi^2$ of the PV fit if the track is included. In order to solve this problem, the tracks with $\chi^2_{IP} < 25$ around the reconstructed and accepted PV are removed from the set of tracks used for further PV search in the event. The seeding and fitting steps are repeated until no new PV is found in the event.

2.1 Seeding

The purpose of primary vertex seeding is to find candidates for primary vertices - points at which a sufficient number of tracks pass close to each other. Two seeding algorithm were developed: the 3D seeding which is able to find seeds in the detector volume and fast seeding which seeks for the PV seed along the beam line.

2.1.1 3D seeding

The 3D seeding algorithm does not use any constraint on the position and is able to find PV seeds displaced from the beam line. This is done in a loop over all tracks. For each track (base track) a number of close tracks is determined. A track is defined as close if its distance of closest approach (DOCA) to the base track is less than 1 mm. At least four close tracks are required, otherwise the next base track is considered. For every track pair the point of closest approach (POCA) is calculated. Then the average position over POCA is calculated in a two-step procedure. First, the truncated mean is calculated iteratively. The POCA with maximum distance to this mean is determined and removed if the distance is larger than 5 mm. If enough POCA entries are left after truncating, the weighed average is then calculated. The rough uncertainties on coordinates of every POCA are calculated by assigning typical track parameters and taking into account an opening angle of the tracks. Weighed $x$, $y$ and $z$ coordinates are calculated and the point with such coordinates is added to a seed collection. All tracks used in the calculation of this weighed average are marked as used. Then the loop over tracks is continued and the tracks marked as used are skipped. Obviously, a PV cannot be reconstructed if its seed is not found. Seeding should ideally have 100% efficiency. Therefore, it is important to maximize the seeding efficiency allowing for a certain amount of false seeds. The false seeds do not affect much the performance since the subsequent PV fit predominantly fails. The negative effect is small increase of the time of PV reconstruction.

2.1.2 Fast seeding

The fast seeding procedure is based on finding the clusters of tracks along the beam line. The cluster is defined by the $z_{\text{clus}}$ coordinate and its uncertainty $\sigma_{z_{\text{clus}}}$. The procedure starts with a set of initial clusters determined by the closest approaches of the tracks with respect to the beam axis. The uncertainty of the cluster is parametrized depending on the track angle $\theta$ with respect to the beam axis, the uncertainty decreasing with increasing
The general principle of the method is as follows. In each iteration a list of all pairs of clusters are considered, and the pair with the minimum distance is selected. The distance itself is defined as:

\[ D_{\text{pair}} = \frac{|z^{\text{clu}_1} - z^{\text{clu}_2}|}{\sqrt{(\sigma_{z^{\text{clu}_1}})^2 + (\sigma_{z^{\text{clu}_2}})^2}}. \]  

(1)

The selected pair is merged into one cluster under condition that \( D_{\text{pair}} < 5 \). The weighted mean method for clusters from the pair is employed to get \( z^{\text{clu}} \) and \( \sigma_{z^{\text{clu}}} \) of the merged cluster. Then the next iteration is performed. The procedure stops if no more pair of clusters remains to be merged. The final clusters are the candidates for the primary vertex seeds. In the final step the quality of these candidates are checked. The clusters with low track multiplicity are removed. Additional quality conditions are also applied to reduce the rate of false clusters. Main source of false seeds comes from clusters with relatively low multiplicities, which are formed around tracks with small uncertainties pointing towards a high-multiplicity primary vertex. A stable working point has been found, balancing between high efficiency and the rate of false seeds.

### 2.2 Fitting

The standard least square iterative procedure for primary vertex fitting is not optimal, in particular for events enriched in secondary particles which do not originate directly form the PV, like in the case of \( b\bar{b} \) production. The reason is that at the reconstruction stage the signal tracks coming form the \( b \)-hadron decays are not yet known and they are likely assigned to the PV. This leads to a systematic shift towards the secondary vertex. Furthermore, there is always a fraction of badly measured tracks (ghost tracks or tracks with large multiple scattering). They decrease the resolution of the PV position tending to pull the position of the PV due to an underestimated error of the track parameters. The fitting procedure has been developed in order to limit such effects. The adaptive weighted least square method has been employed. The Tukey biweight method [3] is used to assign a weight to the track according to the value of its \( \chi^2_{\text{IP}} \). The Tukey biweight dependence on the track \( \chi^2_{\text{IP}} \) is given by the following expressions:

\[
W_T = (1 - \chi^2_{\text{IP}}/C_T^2)^2 \quad \text{if} \quad \chi_{\text{IP}} < C_T, \\
W_T = 0 \quad \text{if} \quad \chi_{\text{IP}} \geq C_T, 
\]

(2)

where \( C_T \) denotes the Tukey constant. This dependence for different values of the \( C_T \) constant is illustrated in Fig.1. The three types of tracks that contain VELO segment are used for fitting: Long, Upstream and VELO tracks. The most useful are Long tracks as they contain segment in the tracking stations downstream the magnet and therefore have precise momentum assigned. The Upstream tracks are reconstructed in the region between the interaction point and the magnet. The precision of their momentum measurement is limited. VELO tracks are reconstructed in the VELO detector only with no momentum information. Long and Upstream tracks are extrapolated using full transport service while VELO-only tracks are extrapolated as straight lines.
The position of the reconstructed primary vertex is determined by the least square method minimizing the following $\chi^2_{PV}$:

$$\chi^2_{PV} = \sum_{i=1}^{n_{tracks}} \chi^2_{IP,i} \cdot W_{T,i},$$

where $\chi^2_{IP,i}$ denotes $\chi^2$ of the track impact parameter with respect to the PV, and $W_{T,i}$ denotes the corresponding biweight. The estimation of the PV position is performed iteratively. In every iteration, a new position of the PV is determined. The tracks are extrapolated to the $z$ coordinate of a new PV position. Then the impact parameter $\chi^2_{IP}$ as well as biweight $W$ are calculated. The procedure stops when the convergence conditions are satisfied:

- $|\Delta z| < 0.5 \, \mu m$, where $|\Delta z|$ is the shift of $z$ coordinate of PV after the iteration.
- at least 5 tracks are assigned to the PV; a track is assigned to the PV if its weight is non-zero e.g. $\chi^2_{IP} < 9$

The adaptive method is employed to assure proper convergence. The seeding provides candidates with a limited accuracy. To avoid converging into a local minimum, the $C_T$ is initially set to be large and it is decreasing during the first five iterations down to the nominal value of 3. In addition, the tracks which are assigned zero weight during a
given iteration are not excluded. The weights of all tracks collected around the seed are recalculated during every iteration and contribute to the $\chi^2_{PV}$ if their weights are non-zero. The maximum number of iterations is set to 30. This number is large because the convergence is not as fast as for the standard least square procedure where the weights do not change with iterations.

3 Performance

3.1 Reconstruction efficiency

There is no single number representing the overall efficiency of the PV reconstruction, since the reconstruction efficiency depends on the vertex multiplicity which varies with data samples used. Moreover, the efficiency depends on the luminosity leading to the different number of primary $pp$ interactions per bunch crossing. In the case of multiple $pp$ interactions the efficiency of close vertices is degraded due to two effects. In the case of two overlapping PVs only one PV is reconstructed. If there are two close vertices which are relatively well separated: the one with high multiplicity and the second one with low multiplicity, the higher multiplicity PV is reconstructed first. Due to the migration of tracks from low to high multiplicity PV, the number of remaining tracks may be too low to satisfy the reconstruction criteria.

The primary vertex reconstruction efficiency is defined as a ratio of the number of reconstructed PVs and the number of reconstructible PVs. Reconstructible primary vertex is defined as a vertex which contains at least 5 reconstructed VELO tracks corresponding to the simulated particles coming from a given primary vertex. PVs are divided into two subclasses: isolated reconstructible PVs and close reconstructible PVs. The first one is defined as a reconstructible PV with a distance to the closest reconstructible PV $|\Delta z| > 10$ mm. A reconstructible PV is called “close” if its distance to the closest reconstructible PV $|\Delta z| < 10$ mm. A reconstructed PV is defined as “false” if its distance to any reconstructible PV $|\Delta z| > 5\sigma_z$, where $\sigma_z$ denotes the estimated position error along $z$ axis. A large fraction of such vertices represent reconstructed PVs corresponding to true PVs, which contain reconstructed tracks, but do not fulfill the conditions to be reconstructible.

The PV efficiency studies are based on $b\bar{b}$-inclusive and minimum bias simulated samples with the average number of $pp$ interactions per bunch crossing $\nu = 2.5$. The efficiency plots for all, isolated and close vertices are shown in Fig. 2 for both $b\bar{b}$-inclusive and minimum bias simulated samples. They are plotted as a function of the PV multiplicity, defined as a number of reconstructed tracks coming directly from the PV. As may be observed in Fig. 2 the PV reconstruction efficiency is high for large vertex multiplicity. For isolated PVs of more than 10 tracks the reconstruction is 100% efficient. It is reduced for multiplicities less than 10. There is no significant difference between distributions for $b\bar{b}$-inclusive and minimum bias simulated samples. The average values of PV reconstruction efficiencies for different categories of reconstructible PVs are listed in Table 1. The rate of the false reconstructed PVs is below 3% for simulated $b\bar{b}$-inclusive and minimum bias samples with $\nu = 2.5$, which indicates a proper PV performance for complicated
Table 1: Average PV reconstruction efficiencies and false PV rates for different categories of reconstructible PVs for $b\bar{b}$-inclusive and for minimum bias samples.

<table>
<thead>
<tr>
<th>Reconstruction efficiency [%]</th>
<th>$b\bar{b}$-inclusive</th>
<th>minimum bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>All reconstructible</td>
<td>94.2 ± 0.01</td>
<td>93.8 ± 0.01</td>
</tr>
<tr>
<td>Isolated</td>
<td>97.1 ± 0.01</td>
<td>96.1 ± 0.01</td>
</tr>
<tr>
<td>Close</td>
<td>82.3 ± 0.01</td>
<td>82.1 ± 0.01</td>
</tr>
<tr>
<td>False PV rate [%]</td>
<td>1.31 ± 0.03</td>
<td>2.72 ± 0.04</td>
</tr>
</tbody>
</table>

events with multiple $pp$ interactions.

### 3.2 Resolution

Resolution of the PV coordinates can be determined for simulated events and data. In the case of simulation, the resolution is determined from the distribution of the difference between fitted and generated PV coordinates $\Delta x$, $\Delta y$ and $\Delta z$. The resolution depends on the number of tracks used for PV reconstruction. Therefore, similarly to the reconstruction efficiency, there is no global number for the resolution. The resolutions obtained from the Gaussian fit (sigma of the fitted Gaussian) to the $\Delta x$, $\Delta y$ and $\Delta z$ distributions in each multiplicity bin is shown in Fig. 3 for $b\bar{b}$ inclusive events. The resolution improves with increasing PV multiplicity and starts to saturate for larger multiplicities. The corresponding mean values of the fitted Gaussian is presented in Fig. 4. A residual bias of few $\mu m$ may be observed for PV z coordinate. The pull distributions ($\Delta x$, $\Delta y$ and $\Delta z$ divided by their estimated errors) are shown in Fig. 5. The sigma of the distribution for the z coordinate is compatible with 1, while for x and y is slightly higher.

#### 3.2.1 Resolution on data

The resolution is measured from data using the following method. The total number of tracks, $n$, in a vertex is split randomly into two sets of $n_{\text{half}} = n/2$ tracks. If the number of PV tracks $n$ is odd, a random track is discarded. Then the reconstruction algorithm is run separately on those two sets of tracks. Each vertex is an independent measurement of the true interaction point. The distance between two vertices in $x$, $y$ and $z$ divided by $\sqrt{2}$ corresponds to $\Delta x$, $\Delta y$, $\Delta z$ between true and measured positions respectively. This can be used to estimate the resolution of PV containing $n_{\text{half}}$ tracks. The procedure has been validated for simulated data by a comparison to the resolution determined using generated coordinates. The comparison of resolutions measured on the data and predicted from the simulation is shown in Fig. 6.
4 Summary

The primary vertex reconstruction in LHCb has been described. The methods presented were used to find primary vertices since beginning of data taking. The emphasis has been put on robustness, since the performance of the detector and stability of the beam conditions were not known beforehand. Various seeding methods and fitting algorithms were developed to meet the requirements of different analyses which were carried out for gradually increased intensity of the LHC beam. The performance of the primary vertex reconstruction has been presented in terms of reconstruction efficiency and spatial resolution. The resolution predicted from simulation has been compared to the one measured on data.

References


Figure 2: Primary vertex reconstruction efficiencies as a function of multiplicity for: a) all reconstructible; b) isolated and c) close PVs for simulated $b\bar{b}$-inclusive (blue line) and minimum bias (red line) samples.
Figure 3: Primary vertex resolution in a) $x$, b) $y$ and c) $z$ coordinate as a function of PV multiplicity for simulated $b\bar{b}$-inclusive events. The vertex multiplicity is the number of reconstructed tracks corresponding to generated particles coming directly from the primary vertex.
Figure 4: The mean value of a) $\Delta x$, b) $\Delta y$ and c) $\Delta z$ distributions determined from simulation. No bias is observed for $x$ and $y$ coordinates. The residual bias in the $z$ coordinate may be noticed.
Figure 5: The pull distributions determined from simulation for a) $\Delta x$, b) $\Delta y$ and c) $\Delta z$. Gaussian fit to the core component in the range between -2 and 2 is shown.
Figure 6: Ratios of PV resolution measured from data and predicted from simulation as a function of PV multiplicity for a) $x$, b) $y$ and c) $z$ coordinate.