Global Decay Chain Vertex Fitting at B-Factories

J.-F. Krohn\textsuperscript{a,} \textsuperscript{a}, P. Urquijo\textsuperscript{a}, F. Abudiné\textsuperscript{d}, S. Cunliffe\textsuperscript{b}, T. Ferber\textsuperscript{b}, M. Gelb\textsuperscript{c}, J. Gemmler\textsuperscript{c}, P. Goldenzweig\textsuperscript{c}, T. Keck\textsuperscript{c}, I. Komarov\textsuperscript{b}, T. Kuhn\textsuperscript{c}, L. Ligiòi\textsuperscript{d}, M. Lubej\textsuperscript{f}, F. Meier\textsuperscript{c}, F. Metzner\textsuperscript{c}, C. Pulvermacher\textsuperscript{c}, M. Ritter\textsuperscript{e}, U. Tamponi\textsuperscript{b}, F. Tenchini\textsuperscript{b}, A. Zupanc\textsuperscript{f}

\textsuperscript{a}University of Melbourne, Melbourne, Australia
\textsuperscript{b}Deutsches Elektronen-Synchrotron, Hamburg, Germany
\textsuperscript{c}Karlsruhe Institute of Technology, Karlsruhe, Germany
\textsuperscript{d}Max-Planck-Institut für Physik, Munich, Germany
\textsuperscript{e}Ludwig Maximilians Universität, Munich, Germany
\textsuperscript{f}Jožef Stefan Institute, Ljubljana, Slovenia
\textsuperscript{g}University of Sydney, Sydney, Australia
\textsuperscript{h}INFN - Sezione di Torino, Torino, Italy

Abstract

We present a particle vertex fitting method designed for B factories. The presented method uses a Kalman Filter to solve a least squares estimate to globally fit decay chains, as opposed to traditional methods that fit each vertex at a time. It allows for the extraction of particle momenta, energies, vertex positions and flight lengths, as well as the uncertainty estimates of these quantities. Furthermore, it allows for the precise extraction of vertex parameters in complex decay chains containing neutral final state particles, such as $\gamma$ or $K_{S}^{0}$, which cannot properly be tracked due to limited spatial resolution of longitudinally segmented single-layer crystal calorimeters like the Belle II ECL. The presented technique can be used to suppress combinatorial background and improve resolutions on measured parameters. We present studies using Monte Carlo simulations of collisions in the Belle II experiment, where modes with neutrals are crucial to the physics analysis program.

1. Introduction

Particle vertex fitting techniques are widely used in particle and nuclear physics. Beyond the suppression of background, applications range from the improvement of particle momentum resolution (under the assumption they originate from some vertex point), to the determination of the presence of intermediate particles and the precision determination of decay vertex positions. One

\textsuperscript{a}Corresponding author.
Email: jkrohn@student.unimelb.edu.au

Preprint submitted to Elsevier Science February 1, 2019
can, for example, combine the measurements of two charged pion tracks originating from the decay of a $K_S^0$ to extract the decay vertex position, flight length and four-vector and their uncertainties. By performing a kinematic fit, one obtains an improvement of the pion track momenta and can use the $\chi^2$ probability of the fit result to suppress background.

In order to construct more complex decay topologies, one usually combines cascades of these fits starting with long lived stable particles such as electrons, muons, pions and photons, forming intermediate resonances and finally the full decay of interest. For example, in the decay $B^0 \to J/\Psi K_S^0$, where $J/\Psi \to \mu^+\mu^-$ and $K_S^0 \to \pi^+\pi^-$, one would first fit the $J/\Psi$ and $K_S^0$ candidates and then use these to construct the $B^0$ candidates, as depicted in Fig. 1a and Fig. 1b. However, this only works well if the final state particles are charged and leave traces in the tracking detectors. Neutral particles cannot be tracked; they are only measured by their energy deposition in the calorimeter. Single layer crystal calorimeters do not offer directional information on where the particle associated to an energy deposition originated from. This means that the decay vertex cannot be extracted from the fit. In order to obtain the momentum vector, it has to be assumed that the particle originates from the primary interaction point and travels directly into the cluster’s center of gravity. This can introduce a large bias on the momentum components. Consider, for example, the decay $B^0 \to J/\Psi K_S^0$, where the kaon instead decays to neutral final states, $K_S^0 \to \pi^0\pi^0$, and the pions decay, $\pi^0 \to \gamma\gamma$, as displayed in Fig. 2. Weakly decaying intermediate particles such as $K_S^0$ have relatively long flight lengths, up to 15 cm for $K_S^0$ at Belle II. Thus for a neutral particle in such a particle decay chain, the assumption that it originates from the primary $e^+e^-$ collision is not sufficient.

The method we present in this paper overcomes these issues by globally fitting the entire decay tree in a single fit, taking into account all intermediate particles, extracting all involved particle’s four-momenta, vertex positions, flight lengths and their covariance matrices, using a Kalman Filter as described in Ref. [1]. We use the software environment of Belle II and rely on the C++ template library EIGEN [2] for matrix operations, which provides a fast execution time for the fit algorithm. We furthermore present physics applications of the fitter with Belle II Monte Carlo samples.

1.1. The Belle II Experiment

The algorithm described in this paper was developed for the analysis software framework of Belle II. The Belle II experiment takes place at the asymmetric $e^+e^-$ collider, SuperKEKB. SuperKEKB provides a beam energy slightly above the mass of the $Y(4S)$ resonance (10.58 GeV/$c^2$) at an instantaneous luminosity of $8 \cdot 10^{35}$ cm$^{-2}$s$^{-1}$. The $Y(4S)$ resonance decays into pairs of $B$-mesons just above production threshold, hence this type of experiment is referred to as a $B$-factory. The asymmetric beam energy gives the $B$-meson a relativistic boost along a direction close to the detector’s axis of symmetry, increasing its flight length in the lab frame, which makes it easier to study the time evolution of $B$ decays - a key observable in the study of CP symmetry violation.
Figure 1: a) Depiction of a $J/\Psi \rightarrow \mu^+\mu^-$ decay. The red lines show the track helix approximations obtained from the tracking detectors, the blue dashed lines show the decaying particle momentum vectors found by the fit. Since the decay length of the $J/\Psi$ is too short to be seen in the detector, its decay vertex is taken to be the one of the $B^0$. b) Depiction of a $B^0 \rightarrow J/\Psi (\rightarrow \mu^+\mu^-) K_S^0 (\rightarrow \pi^+\pi^-)$ decay. Measured track helices do not necessarily overlap in three dimensions. The depicted length ratios are not to scale.

Figure 2: Depiction of a the decay $B^0 \rightarrow K_S^0 J/\Psi$. The red lines show the track helix approximations obtained by the tracking detectors, the blue dashed lines show the composite particle momentum vectors found by the fit. The dashed black lines depict the photon momenta found by the fit. Note that these can only be extrapolated by the fit as the directional information of the calorimeter is not sufficient. The initial guess is that they point from the interaction point towards the calorimeter cluster. The decay lengths of the $J/\Psi$ and $\pi^0$ are too short to be seen in the detector therefore the vertex positions are taken from the particle above them in the hierarchy.
The Belle II detector has a cylindrical structure designed to study the decays of \( B^- \) and \( D^- \)-mesons, \( \tau \)-leptons and other processes produced in \( e^+e^- \) collisions. Six layers of silicon vertex detectors (2 layers of silicon pixels (PXD), and 4 layers of double sided silicon detectors (SVD)) are located in the central volume of the detector, designed to accurately track the flight paths of charged particles. The following layers are, a central drift chamber (CDC) used to measure track trajectories within a solenoid magnetic field, Cherenkov light based particle identification devices surrounding the CDC in the barrel (TOP) and forward regions (ARICH), followed by the CsI(Tl) electromagnetic calorimeter (ECL). The outermost layers are composed of, a magnet solenoid, a \( K_0^L \) and muon detector system (KLM), which is also used as the flux return yoke of the magnet. The magnetic field is aligned along the detector’s axis of symmetry. For a more detailed description of the detector see Ref. [3].

Fig. 3 and Fig. 4 show an event display depicting simulated particles traversing the Belle II detector. In the decay \( \Upsilon(4S) \to B^0\bar{B}^0 \), one meson decays as \( \bar{B}^0 \to D\omega\pi\pi \), and the other as \( B^0 \to K_S J/\Psi \), with \( J/\Psi \to \mu^+\mu^- \) and \( K_S \to \pi^0\pi^0 \) with \( \pi^0 \to \gamma\gamma \). Fig. 3 depicts the full detector geometry and Fig. 4 shows a close-up of the inner vertex detectors. In this example we show that the decay vertex of the \( K_S^0 \) can be very displaced.

2. Extended Kalman Filter

Vertex fitting is a least squares minimization problem. The best least squares estimator (LSE) is the solution to this problem. The computational challenge in finding a LSE lies in matrix inversions, which naively scale as \( O(n^3) \), where \( n \) is the dimension of the matrix. In a naive approach, this is equal to the number of parameters extracted in the fit. A Kalman Filter is an iterative approach to find the LSE by defining a series of constraints (knowledge of parameters from measurements and symmetries) on a hypothesis (a set of particle parameters in this case). The hypothesis has different states during the stages of the filtering process. The filtering process is an iterative algorithm applying the constraints sequentially and updating the state with respect to the constraints. The sequence is repeated until convergence is reached or divergence is observed. In the case of convergence, the last state describes the best hypothesis for the parameters that can be found. We use an Extended Kalman Filter in the gain matrix formulation [4, 5, 1] in our algorithm. The state vector \( \mathbf{x} \) holds the particle parameters to be optimized by the fit. The parameters depend on the type of particle. The most general parametrization takes the form

\[
\mathbf{x} = \{x_1, y_1, z_1, \theta_1, p_{x1}, p_{y1}, p_{z1}, E_1, \ldots, x_n, y_n, z_n, \theta_n, p_{xn}, p_{yn}, p_{zn}, E_n\},
\]

(2)

where vertex coordinates of the \( i \)-th particle are denoted as \( \{x_i, y_i, z_i\} \), its decay length is denoted as \( \theta_i \) and the four momentum is \( \{p_{x,i}, p_{y,i}, p_{z,i}, E_i\} \), and \( n \) is the number of particles in the fitted topology. This implies that \( n \approx 8 \times \) number of particles. The problem is split into \( k \) constraints given by measurements and other knowledge of parameters (see Sec. 3 for the definitions
Figure 3: Event display projected onto the $x - y$ plane depicting the process of $e^+e^- \rightarrow \Upsilon(4S) \rightarrow B^0\bar{B}^0$, with $B^0 \rightarrow D\omega\pi$ and $\bar{B}^0 \rightarrow K_S J/\Psi$, where $J/\Psi \rightarrow \mu^+\mu^-$ and $K_{S0} \rightarrow \pi^0\pi^0$ with $\pi^0 \rightarrow \gamma\gamma$. All particles of the signal $B^0$ decay are indicated by black lines. The rest of the event correspond to the decay of the $\bar{B}^0$. All distances are measured in centimetres. For a close-up of the vertex detector see Fig. 4.
Figure 4: Close-up of Fig. 3. In this event, the $K^0_S$ travels until the edge of the vertex detectors before decaying.
of the measurement $m$ and hypothesis $h$ for each of the different constraints). The minimized $\chi^2$ can be expressed as a weighted sum of $k$ sets of equations, which are the constraints. For a given iteration, $\alpha$, of the Kalman Filter, one can write
\[
\chi^2_\alpha = (m_1 - h_1(x^\alpha_0))^T V_1^{-1}(m_1 - h_1(x^\alpha_0)) + \ldots + (m_k - h_k(x^\alpha_{k-1}))^T V_k^{-1}(m_k - h_k(x^\alpha_{k-1})),
\]
using the measurement covariance matrix $V_k^{-1}$ transported to the $m - h$ system as a weight, such that the hypothesis of each constraint $k$ depends on the outcome of the previous constraint.

Minimizing Eq. (3) yields a rule to find a new state $x^\alpha_k$ for the current iteration and constraint, that is
\[
x^\alpha_k = x^\alpha_{k-1} - K^\alpha_k r^\alpha_k, \quad (4)
\]
where $K$ is a gain matrix, which will be defined below. However, we first define the residual $r$ as the distance between measurement and hypothesis
\[
r^\alpha_k = m_k - h_k(x^\alpha). \quad (5)
\]
The hypothesis $h_k(x^\alpha)$ can be linearised around a reference state of the previous iteration $x^{\alpha-1}$ using the Jacobian
\[
H_{k}^{\alpha-1} = \frac{\partial h_k(x^{\alpha-1})}{\partial x^{\alpha-1}}, \quad (6)
\]
such that
\[
h_k(x^\alpha) = h_k(x^{\alpha-1}) - H_{k}^{\alpha-1} \cdot (x^\alpha - x^{\alpha-1}). \quad (7)
\]
Eq. (5) thus becomes
\[
r^\alpha_k = m_k - h_k(x^\alpha) + H_{k}^{\alpha-1} \cdot (x^\alpha - x^{\alpha-1}). \quad (8)
\]
The gain matrix is calculated for every constraint in every iteration and is defined as
\[
K_k = C_{k-1} H_k^T R_k^{-1}. \quad (9)
\]
The only matrix to be inverted is of the dimension of the constraint and not the dimension of the statespace. The current state’s covariance matrix is obtained via propagation of uncertainties
\[
C_k = (1 - K_k H_k) C_{k-1} (1 - K_k H_k)^T = C_{k-1} - C_{k-1} H_k R_k^{-1} H_k^T C_{k-1} \quad (10)
\]
\[
= C_{k-1} - C_{k-1} H_k^T K_k^T H_k.
\]
The covariance matrix of the residual system can be found to be
\[
R_k = V_k + H_k C_{k-1} H_k^T. \quad (11)
\]
For constraints where no measurement is available, for example the mass constraint, we set $V_k = 0$. 

7
Linearising the hypothesis around a reference state, see Eq. 7, makes the fits more stable than a normal Kalman Filter and reduces the overall number of signal candidates for which the fit diverges. In this formulation of the LSE problem, the update of the covariance matrix, Eq. 10, is the computational bottleneck, as it is a dense $m \times m$ matrix with the dimension of the state vector $m$ and must be calculated and filled $k$ times for each iteration.

3. Parametrising and constraining the decay chain

To parametrise the decay chain, we use a set of parameters that describe the properties of the particles. We perform a number of reductions on these properties to reduce the dimensionality of the problem. For final state particles, we only save the momenta, as they do not have decay vertices. For their production vertices we use the decay vertices of their respective composite particles, later referred to as “mothers”. The energy of each each final state particle is calculated using the momenta and it’s nominal mass hypotheses, taking the mass values provided by the particle data group (PDG) [6]. Intermediate particles are classified in two categories: particles that decay dominantly via the strong force and via the weak force. Strongly decaying particles with a lifetime of less than $10^{-14}$ s, which corresponds to boosted flight lengths of the order of less than 1 µm, are treated as if they instantly decay (the detector has a vertex position resolution in the x-y plane for charged particles of 20 – 30 µm). The hypothesised quantities are their energy and momenta, while the production and decay vertices coincide and are taken from their mother particle’s decay vertex. For weakly decaying particles, we additionally measure a decay vertex and a flight length, defined as the distance between the production and decay vertices in three dimensions.

3.1. Parametrising the constraints

Constraints are defined by Eq. 7. The resulting Jacobians take the form of $m \times n$ matrices where $m$ is the dimension of the state vector and $n$ is the dimension of the respective constraint. Thus, only few of its elements are non-zero. For example, for a three dimensional point constraint $k$, the hypothesis of particle number $j$ with $h_j = \{x_j, y_j, z_j\}$ and $x$ as in Eq. 2, only the $j$-th diagonal block is non-zero

$$\frac{\partial h}{\partial x} = H = \begin{pmatrix} 0 & \mathbb{1}_3 & 0 \end{pmatrix}.$$ (12)

The blocks filled with zero correspond to the parameters of particle $x_i \neq x_j$. We will omit the columns filled with zeros throughout this section, for brevity. In the following we list the definitions of constraints that have been implemented in the Belle II software, based on the specific geometry of Belle II.
3.1.1. Reconstructed track

A track can be parametrised with a five parameter helix. In Belle II it was chosen to use a perigee-parametrised helix, such that the helix is defined at the perigee, the point of closest approach of the helix to the origin of the coordinate system. The corresponding transformations to transport a helix to that point are discussed in Ref. [7]. A description of the parameters can be found in Table 1 and a depiction of the helix is in Fig. 5. We parametrise tracks such that we can express the model’s dependence on Cartesian parameters as

\[ h_{\text{track}}(x) = \begin{pmatrix} d_0 \\
\phi_0 \\
\omega \\
z_0 \\
tan \lambda \end{pmatrix} = \begin{pmatrix} A(1 + U)^{-1} \\
atan2(p_y, p_x) - atan2(\omega \cdot \Delta_{\parallel}, 1 + \omega \cdot \Delta_{\perp}) \\
z + l \cdot tan \lambda \\
p_z / p_t \end{pmatrix} \cdot (13) \]

Where \( atan2 \) refers to the phi domain corrected inverse tangent function. We use the same parametrisation for the hypothesis and the measurement. We label the measurement quantities with the index \( m \). The residuals of iteration \( \alpha \) then become

\[ r_{\alpha}^{\text{track}}(x) = \begin{pmatrix} d_{0,m} - d_0 \\
\phi_{0,m} - \phi_0 \\
\omega_m - \omega \\
z_{0,m} - z_0 \\
tan \lambda_m - tan \lambda \end{pmatrix} + H^\alpha - 1 \cdot (x^\alpha - x^{\alpha-1}) \cdot (14) \]

We define the Jacobian block \( A := \partial h / \partial x \) as the derivatives with respect to the vertex position, and \( B := \partial h / \partial p \) as the derivatives with respect to momentum. The positions of these blocks in the Jacobian depend on the topology fitted and the particle represented by the track. We choose to order the state vector hierarchically. This means that the decay vertex parameters come before its momentum, followed by the daughter particle’s parameters. The full Jacobian \( H \) then takes the following form

\[ H = \begin{pmatrix} \ldots & \ldots & \ldots \\
\ldots & A & \ldots & B & \ldots \\
\ldots & \ldots & \ldots \\
\ldots & \ldots & \ldots \\
\ldots & \ldots & \ldots \end{pmatrix} \cdot (15) \]

For the non-zero elements of the Jacobian blocks, denoted by \( \partial d_0 / \partial x = A_{d_0,x} \), we derive the spatial components as

\[ A_{d_0,x} = \frac{p_y p_0}{p_{t0}}, \quad A_{d_0,y} = -\frac{p_x p_0}{p_{t0}}, \]

\[ A_{\phi_{0,x}} = \frac{a \cdot q \cdot p_{x0}}{p_{t0}^2}, \quad A_{\phi_{0,y}} = -\frac{a \cdot q \cdot p_{y0}}{p_{t0}^2}, \]

\[ A_{z_{0,x}} = -\frac{p_x \cdot p_{x0}}{p_{t0}^2}, \quad A_{z_{0,y}} = -\frac{p_y \cdot p_{y0}}{p_{t0}^2}, \quad A_{z_{0,z}} = 1, \quad (16) \]
Table 1: Definitions of the Belle II helix parametrisation and their dependencies.

| Parameter | Definition |
|-----------|------------|
| $d_0$ | The distance of closest approach to the $z$ axis (POCA) signed with the $z$ component of the angular momentum w.r.t. to the origin. |
| $\phi_0$ | Azimuthal angle of the momentum at the POCA. |
| $\omega$ | Scaled inverse of the track momentum. |
| $z_0$ | The pivotal point is the perigee that is the POCA. |
| $\tan \lambda$ | The angle of the momentum at the POCA with respect to the $x - y$ plane. |

\[
\phi = \text{atan2}(p_y, p_x)
\]

Angle of the $x - y$ plane to the helix.

\[
\Delta_\parallel = -x \cdot \cos \phi - y \cdot \sin \phi
\]

Quantities used to transport the coordinate system, such that $\{x, y, z\}$ points to the perigee.

\[
\Delta_\perp = -y \cdot \cos \phi + x \cdot \sin \phi
\]

\[
A = 2 \cdot \Delta_\perp + \omega \cdot (\Delta_\parallel^2 + \Delta_\perp^2)
\]

\[
U = \sqrt{1 + \omega \cdot A}
\]

\[
\lambda = \text{atan2}(\omega \cdot \Delta_\parallel, 1 + \omega \cdot \Delta_\perp)
\]

Arc length from the perigee to a point.

\[
l = \text{atan2}(\omega \cdot \Delta_\parallel, 1 + \omega \cdot \Delta_\perp)
\]

Moments along the $x$, $y$, $z$ axes.

\[
q
\]

Charge of the particle.

\[
B_z
\]

Magnetic field strength in $z$-direction.

\[
a = B_z/c
\]

Magnetic field strength in the $z$ direction divided by the speed of light.

\[
p_t = \sqrt{p_x^2 + p_y^2}
\]

Transverse momentum.

\[
p_{t0} = \sqrt{p_{tx0}^2 + p_{ty0}^2}
\]

Pseudo transverse momentum.

\[
p_{x0} = p_x - a \cdot q \cdot y
\]

Pseudo momentum along the $x$ direction.

\[
p_{y0} = p_y + a \cdot q \cdot x
\]

Pseudo momentum along the $y$ direction.

\[
r^2 = x^2 + y^2
\]

Radius squared.

\[
\beta = 1 + \frac{m}{p_t}
\]

Quantity used for reading convenience.
Figure 5: The perigee parametrisation of the track helix, depiction adapted from [8]. A description of the parameters can be found in Tab. 1.

and for the momenta

\[
B_{d_0,p_x} = - \frac{y((aq)^2r + 2aqp_yx + 2p_y^2\beta)}{p_0p_t^2\beta^2} - \frac{p_x(2p_yx\beta + aq(y^2(-2 + \beta) + x^2\beta))}{p_0p_t^2\beta^2},
\]

\[
B_{d_0,p_y} = \frac{2p_x^2x\beta + 2p_y^2y(p_y - aqx + p_yp_0/p_t)}{p_0p_t^2\beta^2} + \frac{aq(aqx - p_y(x^2(-2 + \beta) + y^2\beta))}{p_0p_t^2\beta^2},
\]
For photons we measure the position of the calorimeter cluster and its energy and can infer the vertex parameters. The geometry, depicted in Fig. 6, gives

\[ 0 = \mathbf{u}_{\text{mother}} + \mathbf{\delta} - \mathbf{m}, \]  

(19)

substituting \( \mathbf{\delta} = \tau \cdot \mathbf{p} \) and inserting the energy relation, we get

\[
\mathbf{h}_{\text{photon}}(\mathbf{x}) = \begin{pmatrix} u_x + \tau \cdot p_x \\ u_y + \tau \cdot p_y \\ u_z + \tau \cdot p_z \end{pmatrix} \quad \text{and} \quad \mathbf{m}_{\text{photon}}(\mathbf{x}) = \begin{pmatrix} m_x \\ m_y \\ m_z \\ E_m \end{pmatrix},
\]

(20)

where \( \{u_x, u_y, u_z\} \) are the production vertex coordinates, \( \{p_x, p_y, p_z\} \) are the parameters of the momentum vector pointing from the production vertex to the mother.
calorimeter cluster, \( \{m_x, m_y, m_z, E_m\} \) are the position and measured energy of
the corresponding ECL cluster. The parameter \( \tau \) is the decay time, which can
be eliminated when writing down the residual to reduce the dimensionality of
the equation system and avoid a trivial local minimum of \( r_\gamma \) at \( \tau = 0 \) when
taking \( \{u_x, u_y, u_z\} = 0 \) as the starting point of the first iteration. Since the
game of the detector is cylindrical, we can not simply eliminate any of the
dimensions as this could introduce a pole in the residual equations. Thus, we
sort the momenta and eliminate the dimension with the highest momentum.

Such that we get a 3-dimensional equation system

\[
\mathbf{r}^\alpha_{\text{photon}}(\mathbf{x}) = \left( \begin{array}{c} (m_i - u_i) - (m_k - u_k) \frac{p_i}{p_k} \\ (m_j - u_j) - (m_k - u_k) \frac{p_j}{p_k} \\ E_m - \sqrt{p_i^2 + p_j^2 + p_k^2} \end{array} \right) + \mathbf{H}^{\alpha-1} \cdot (\mathbf{x}^{\alpha} - \mathbf{x}^{\alpha-1}),
\]

(21)

where the indices \( i, j, k \) indicate the dimensions by order of increasing momen-
tum \( p_k \geq p_i \geq p_j \). We define \( A_{i,u_k} := \frac{\partial h'_i}{\partial u_k} \) and \( B_{i,p_k} := \frac{\partial h'_i}{\partial p_k} \) with the hypothesis of the reduced system \( r' \). Thus, the non-zero entries are

\[
\begin{align*}
A_{0,u_k} &= \frac{p_i}{p_k}, & A_{0,u_i} &= -1, \\
A_{1,u_k} &= \frac{p_j}{p_k}, & A_{1,u_j} &= -1, \\
B_{0,p_k} &= p_k^{-2}, & B_{0,p_i} &= \frac{u_k - m_k}{p_k}, \\
B_{1,p_k} &= p_k^{-2}, & B_{1,p_j} &= \frac{u_k - m_k}{p_k}, \\
B_{2,p_k} &= \frac{-p_k}{|p|}, & B_{2,p_i} &= \frac{p_i}{|p|}, & B_{2,p_j} &= \frac{p_j}{|p|}.
\end{align*}
\]

(22)

The full Jacobian then takes the form

\[
\mathbf{H} = \left( \begin{array}{ccc}
\ldots & A & \ldots \\
\ldots & \ldots & \ldots \\
\ldots & \ldots & \ldots
\end{array} \right).
\]

(23)

We must transform the covariance matrix of the measurement into the reduced
system. For that, we use

\[
\mathbf{V}' = \mathbf{FV}F^T,
\]

(24)

with the transport matrix \( \mathbf{F} = \frac{\partial r'}{\partial m} \), which depends on the sorting of the
momenta such that the non-zero entries are

\[
\begin{align*}
F_{0,m_k} &= \frac{p_i}{p_k}, & F_{0,m_i} &= 1, \\
F_{1,m_k} &= \frac{p_j}{p_k}, & F_{1,m_j} &= 1, \\
F_{2,E_k} &= 1.
\end{align*}
\]

(25)

We do not parametrise this constraint in \( p_k \) in order to keep the derivatives in
Eq. [22] as computationally simple as possible.
3.1.3. Reconstructed $K^0_L$

We treat $K^0_L$ in the same way as photons, except that we use the nominal mass provided by the PDG in the energy calculation. The KLM detector is used for the cluster position measurement instead of the calorimeter. It cannot provide a precise energy measurement. Instead, it extrapolates the energy deposited by a particle as $E = c \cdot n$, where $n$ is the number of hit cells in the cluster and $c$ is a constant with the units GeV. This approach makes the energy measurement for $K^0_L$ much less resolved than for photons.

3.1.4. Kinematic constraint

The kinematic constraint enforces four-momentum conservation, meaning it fits the four-momentum of the mother as the sum of the daughter momenta

$$r^{\alpha}(x) = p_{\text{particle}} - \sum_i p_{i,\text{daughter}} + H^{\alpha-1} \cdot (x^{\alpha} - x^{\alpha-1}).$$  (26)

The Jacobian for a particle with two daughters, which in this example are taken to be stable particles, can be defined

$$H = \begin{pmatrix} \ldots & \ldots & \ldots & \ldots \\ \ldots & A & \ldots & B_1 \\ \ldots & \ldots & \ldots & \ldots \end{pmatrix},$$  (27)

with the blocks $A, B$ as

$$A = \frac{\partial h}{\partial p_{\text{particle}}} = \mathbb{1}_4,$$  (28)

and

$$B_i = \frac{\partial h}{\partial p_{\text{daughter},i}} = -1 \begin{pmatrix} 1 & 1 & 1 \\ p_{x,i}/E_i & p_{y,i}/E_i & p_{z,i}/E_i \end{pmatrix}.$$  (29)

Note that the energy row of $B_i$ depends on how the particle is parametrised, composite particles, for example, are parametrised with an energy variable in the state vector, resulting in $B = -\mathbb{1}_4$, while for stable particles Eq. 29 is used.

3.1.5. Geometric constraint

The geometric constraint fits the decay length parameter $\theta$ for composite particles, see Fig. [6] Accounting for the geometry we have

$$0 = u_{\text{mother}} + \Delta - u.$$  (30)

Instead of directly extracting a flight vector $\Delta$, we use the unit vector of the momentum as it is well constrained by the previously filtered kinematic constraints, substituting $\Delta = \theta \cdot p/|p|$, allows for a more accurate estimation of $\theta$.

Thus we define the residual as

$$r^{\alpha}(x) = u_{\text{mother}} + \theta \cdot \frac{p}{|p|} - u + H^{\alpha-1} \cdot (x^{\alpha} - x^{\alpha-1}) \cdot (x^{\alpha} - x^{\alpha-1}).$$  (31)
using

\[ A = \frac{\partial h}{\partial u_{\text{mother}}} = \mathbb{1}_3, \quad B = \frac{\partial h}{\partial u} = -\mathbb{1}_3, \quad C = \frac{\partial h}{\partial \theta} = \frac{1}{|p|} \begin{pmatrix} p_x \\ p_y \\ p_z \end{pmatrix}, \] (32)

and

\[ D = \frac{\partial h}{\partial p} = \frac{\theta}{|p|^3} \begin{pmatrix} (p_y^2 + p_z^2) & -p_x p_y & -p_x p_z \\ -p_y p_x & (p_z^2 + p_x^2) & -p_y p_z \\ -p_z p_x & -p_z p_y & (p_x^2 + p_y^2) \end{pmatrix}, \] (33)

such that

\[ H = \begin{pmatrix} \ldots & A & \ldots & B & \ldots & C & \ldots & D & \ldots \end{pmatrix}. \] (34)

### 3.1.6. Mass constraint

The mass constraint requires a particle four-vector to be consistent with its nominal mass. We treat the particle as a measurement with infinite precision and use the mass value provided by PDG such that

\[ r^\alpha(x) = m_{\text{PDG}}^2 - E^2 + |p|^2 - H^{\alpha-1} \cdot (x^\alpha - x^{\alpha-1}), \] (35)

with

\[ H = \begin{pmatrix} \ldots & 2p_x & \ldots \\ \ldots & 2p_y & \ldots \\ \ldots & 2p_z & \ldots \\ \ldots & -2E & \ldots \end{pmatrix}. \] (36)

#### 3.1.7. Beam spot constraint

$B$-mesons have a short lifetime and decay very close to the beam spot, therefore it is useful to constrain the reconstructed particles to a volume within that region. The constraint is implemented by considering the initial $e^+e^-$ collision as an abstract mother particle with the parameters and uncertainties of the beam spot, thus constraining the production vertex position of its daughters to an area within the beam spot’s three dimensional uncertainty region using

\[ r^{\alpha}(x) = s - h_{\text{production}} + H^{\alpha-1} \cdot (x^\alpha - x^{\alpha-1}), \] (37)

with the Jacobian

\[ H = \begin{pmatrix} \ldots & \mathbb{1}_3 & \ldots \end{pmatrix}. \] (38)

Here $s$ denotes the beam spot position vector and $h_{\text{production}}$ is the fitted production vertex of the particle. The beam spot is determined as part of the detector calibration, by averaging the $x$, $y$ and $z$ positions of the $e^+e^-$ collisions over many interactions. The decay vertex can be constrained similarly.
3.1.8. Custom Origin constraint

Similar to the beam spot constraint, one can construct another geometric constraint by defining a custom vertex position and an associated uncertainty to be the origin of the decay chain. This can be very useful when the decay chain contains particles that can not be detected, for example, neutrinos or long lived dark matter particles. In these decay chains it is not useful to fit the full chain, because the missing particle’s four momentum makes the kinematic constraint of the mother particle an incorrect assumption. However, knowing that the particle originates from a \(B\)-meson decay, one can define a geometric constraint corresponding to the volume where \(B\)-mesons decay on average.

A beam energy constraint is not needed, as one can mass constrain the \(\Upsilon(4S)\) particle, since all four-vectors in \(e^+e^- \rightarrow \Upsilon(4S)\) are well known.

4. Fitting complex decay topologies and extracting parameters from the fit

In this section we present two use cases of the algorithm containing one and two \(\pi^0\)-mesons in a decay chain, respectively. There are numerous channels where the phase space is large enough to add \(\pi^0\) to a vertex with two charged tracks, which makes this a very common structure in decay trees and therefore an interesting target to fit in a wide spectrum of analyses. We use Monte Carlo samples from the Belle II experiment and evaluate background rejection and signal resolution improvements, as well as the extraction of various other parameters. The selection criteria used for all analyses in this section are listed in Table 2. We use implicit charge conjugation, meaning that charged particles appearing in the decay chains imply including the opposite charges. We abbreviate decays \(x \rightarrow yz\) as \(x(yz)\).

4.1. Fitting decay chains containing a neutral particle

The first mode, with a single \(\pi^0\), is \(B^0 \rightarrow D^{*+}(D^0(K^-\pi^+\pi^0)\pi^+)\pi^-\). This channel is of interest as it is the most common \(D^0\) decay modes. We use the fitter in this example to suppress background. If we reject combinations for which the fit failed - i.e. those with a p-value \(\leq 0\) - we are able to reject about 20% of the background while rejecting only about 4% signal, as displayed in Fig. 7a. Note that since some of the constraints are nonlinear, the \(\chi^2\) which is used to calculate the p-values can be degenerate, where the effective number of degrees of freedom is smaller than the theoretical one used to calculate the p-values. As a result, the p-values can have non-flat shapes [9].

4.2. Fitting ill-defined decay vertices

We study the decay chain \(B^0 \rightarrow K_{S}^0(\pi^0(\gamma\gamma)\pi^0(\gamma\gamma))J/\Psi(\mu^+\mu^-)\). By performing the fit a large amount of the background can be removed by requiring that it passes the fit, see Fig. 8a. The only well defined vertex in this chain is given by the \(J/\Psi \rightarrow \mu^+\mu^-\) decay. The other vertices are ill-defined, the best assumption that can be made for the production vertex position of the four
photons is that they originate from the interaction point. This assumption biases the reconstructed mass of the $K_0^S$. Performing a fit with mass constrained $\pi^0$-mesons improves the extracted mass of the $K_0^S$, so that after the fit, it is centred around the true value, as depicted in Fig. 8b. It is then possible to further reject background outside the nominal mass window and improve the signal purity when analysing this channel.

4.3. Extracting the decay length of $D^0$-mesons from $D^*$ decays using a beam-spot constraint

The geometric constraints allow for the extraction of all the production and decay vertices of all particles in the decay chain. This allows for the extraction of flight lengths and thus lifetimes of intermediate particles such as $D^0$-mesons. We perform this study on $B^0 \rightarrow D^{(*)} \rightarrow D^0(K^-\pi^+\pi^0)\pi^-\pi^-$ decays and extract the decay length of the $D^0$-meson. The results are depicted in Fig. 9. Since $D^*$-mesons decay almost instantly, we will use the three dimensional distance between the $B^0$ and the $D^0$ decay vertices. To improve the resolution on the $B^0$ vertex we apply a beam-spot constraint.

4.4. Using a custom origin constraint to improve the $D^{(*)}$ vertex resolution in $B^0 \rightarrow D^{(*)} \rightarrow D^0(K^-\pi^+\pi^0)l^-\bar{\nu}$ decays

In some scenarios one can not fit the entire decay chain e.g. in semi-leptonic decays $B^0 \rightarrow D^*\mu\nu\bar{\nu}$, where the neutrino escapes detection. Since the neutrino’s four vector is unknown it is not possible to use a kinematic constraint to fit the $B$-meson. To circumvent this problem, one can omit fitting the $B^0$, and instead account for its presence with a geometric constraint. In $B^0 \rightarrow D^{(*)} \rightarrow D^0(K^-\pi^+\pi^0)l^-\bar{\nu}$ decays we use a custom origin constraint to constrain the $D^{(*)}$ production vertices to a volume around the median $B$-meson decay vertex position obtained from Monte Carlo simulations. The results are
Figure 8: a) P-value of the fits to $B^0 \to J/\psi K^0_S$. b) Fitted mass of the $K^0_S$ (green) and the mass before the fit (black). The mass distribution is centred around the true value after performing the fits with a $\pi^0$ mass constraint. The resolution has slightly improved as well.

Figure 9: a) The extracted decay length (green) and the generated decay length (black). The negative tail of the extracted decay length is due to the detectors resolution function. b) Pull (measured – generated/uncertainty) of the decay length distribution.
Figure 10: Resolution (= measured - generated) of the $D^*$+ decay vertex position extraction without (a) and with (b) a custom origin constraint. The vertex position and uncertainties for the origin used in this example are extracted from $B^0$ decays.

Table 2: Selection criteria used in the analyses in this section. All particles use the same criteria if appearing in the decay chains. The invariant mass is obtained by summing the particle’s daughter four-momenta before performing the fit. The particle ID for $K^+/\pi^+$ is defined as the likelihood ratio $LR(K^+(\pi^+)) = \mathcal{L}(K^+(\pi^+))/\mathcal{L}(\pi^+)$ and the beam energy constrained mass $m_{bc}$.

| particle | pre selection applied |
|----------|-----------------------|
| $\gamma$ | $E_\gamma > 0.075$ GeV |
| $\pi^0$ | $0.145$ GeV $> m(\gamma\gamma) > 0.125$ GeV |
| $\pi^+$ | $LR(\pi^+) > 0.5$ |
| $K^+$    | $LR(K^+) > 0.5$ |
| $D^0$    | $1.9$ GeV $> m(K^-\pi^+\pi^0) > 1.7$ GeV |
| $D^*$    | $\Delta m(D^*-D^0) < 0.155$ GeV |
| $B^0$    | $m_{bc} = \sqrt{E_{beam}^2/4 - p^2} > 5.27$ GeV |

depicted in Fig. 10. This technique can improve the resolution on the $D^*$+ vertex position.

5. Conclusion

We presented a versatile fitting tool tailored for the environment of $B$ factories with a cylindrical detector geometry. It can be used for various purposes, such as the extraction of particle production and decay vertices, decay lengths, particle four-momenta and rejection of backgrounds, as well as the extraction of the respective uncertainties for all parameters involved in three dimensions. This global fitting technique is particularly powerful in fitting and reducing background in modes that contain neutral particles, and will be very important for the Belle II physics analysis program. As a future extension of the fitter we are looking into the possibility to map the residuals found by the Kalman
Filter to a scalar signal probability $f : r \rightarrow p$. Such a probability may outperform methods using a p-value derived from the fit $\chi^2$ for the purpose of signal-background-separation.

6. Acknowledgements

This work was made possible by the Belle II collaboration and funding from ARC (Australia), ARRS (Slovenia), BMBF (Germany), EXC153 (Germany), HGF (Germany), INFN (Italy), MSMT (Czech Republic), NSERC (Canada) and U.S. DOE. We would like to thank Wouter Hulsbergen for his original paper, the discussion and sharing his code.
References

[1] W. D. Hulsbergen, Decay chain fitting with a Kalman filter, Nucl. Instrum. Meth. A552 (2005) 566–575. arXiv:physics/0503191, doi:10.1016/j.nima.2005.06.078.

[2] G. Guennebaud, B. Jacob, et al., Eigen v3 (2010). URL http://eigen.tuxfamily.org

[3] The Belle II collaboration, Belle ii technical design report arXiv:1011.0352.

[4] R. E. Kalman, A new approach to linear filtering and prediction problems, Transactions of the ASME–Journal of Basic Engineering 82 (Series D) (1960) 35–45.

[5] R. Frühwirth, Application of Kalman filtering to track and vertex fitting. Nucl. Instrum. Methods Phys. Res., A 262 (HEPHY-PUB-503) (1987) 444. 19 p. URL https://cds.cern.ch/record/178627

[6] C. Patrignani, et al., Review of Particle Physics, Chin. Phys. C40 (10) (2016) 100001. doi:10.1088/1674-1137/40/10/100001

[7] V. Karimaki, Effective circle fitting for particle trajectories, Nucl. Instrum. Meth. A305 (1991) 187–191. doi:10.1016/0168-9002(91)90533-V.

[8] N. Dawe. [link] URL https://github.com/ndawe/tikz-track

[9] R. Andrae, T. Schulze-Hartung, P. Melchior, Dos and don’ts of reduced chi-squared, arXiv:1012.3754

[10] E. Kou, et al., The Belle II Physics Book arXiv:1808.10567