Vertex Reconstruction

M. Regler

28. - 29. August 2003
CERN School of Computing
Donau-Universität Krems

Contents

1 Motivation

- The goal of *experimental particle physics*:

  By investigation of the interactions of elementary particles the experiment provides the data which are required to develop or verify theories (mathematical descriptions of phenomena).

- On a macroscopic time scale the result of a collision, i.e. an interaction of two elementary particles, is again particles which are charaterized by their momenta, charges and masses.

- The aim of *'vertex reconstruction'* is to achieve optimal knowledge of the momenta and the vertex position in space of charged particles at the vertex.

- In most collision experiments, *reconstruction of tracks of all charged particles* is performed first and individually for all tracks.

Motivation

- However, to study the properties of the underlying physics of the reaction *best knowledge of the momenta of the particles at the interaction vertex* is required.
- *Energy and momentum conservation* contribute valuable *constraints* for a better understanding of the reaction dynamics by studying the properties of the reaction products (kind of particles, angular distribution, etc.)

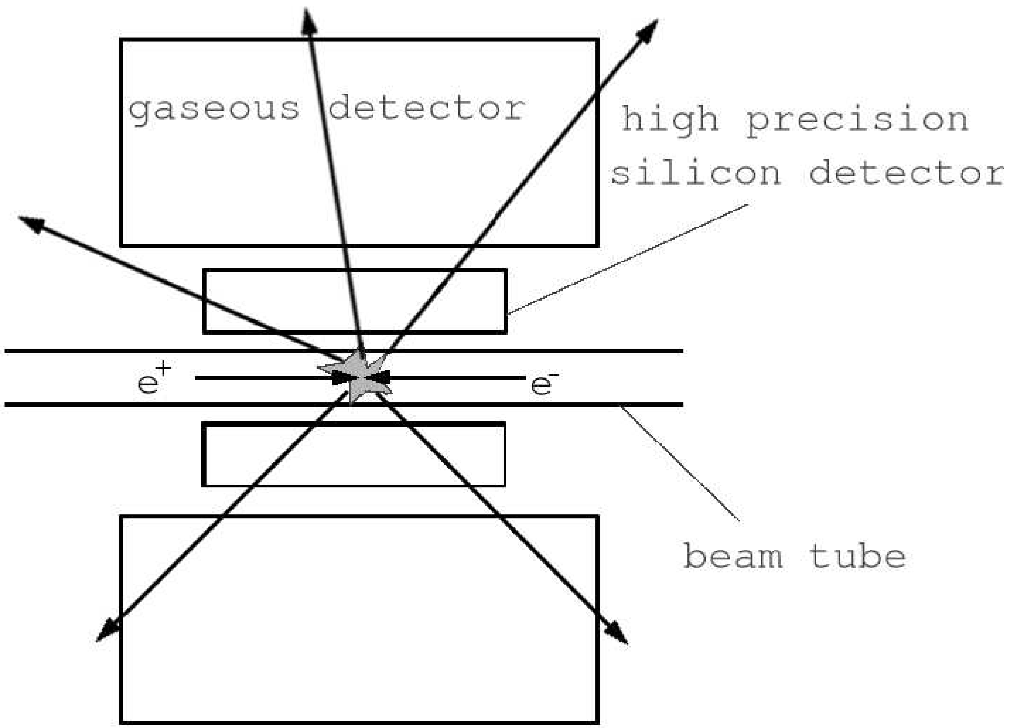


Figure 1: An electron-positron collider experiment (schematic not to scale)

2 The reconstruction of individual charged tracks

2.1 From detector signals to tracks

- *Charged particles* can be measured in a non-destructive way by *creation of free charges*, which, after suitable multiplication and/or amplification, can be registered by the front end electronics.
- The creation of free charges is achieved either by ionisation (in gaseous detectors) or by the creation of electron-hole pairs (in semiconductor detectors)
- The result can be interpreted as the measurement of a coordinate or a pair of *coordinates*, which, together with the knowledge of the detector surface, results in the knowledge of a space point.

The reconstruction of individual charged tracks

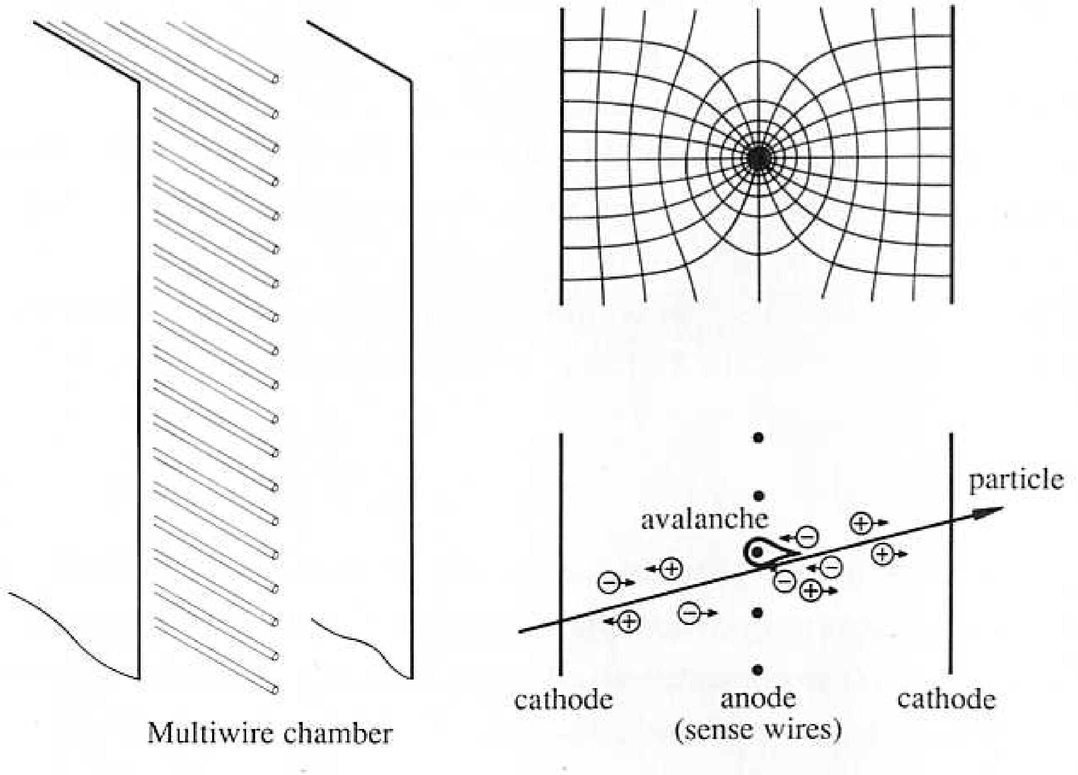


Figure 2: A multiwire proportional chamber with the field lines and the equipotential lines from the high voltage. Charge multiplication occurs exponentially in the region around the sense wire where the field grows like $\frac{1}{r}$.

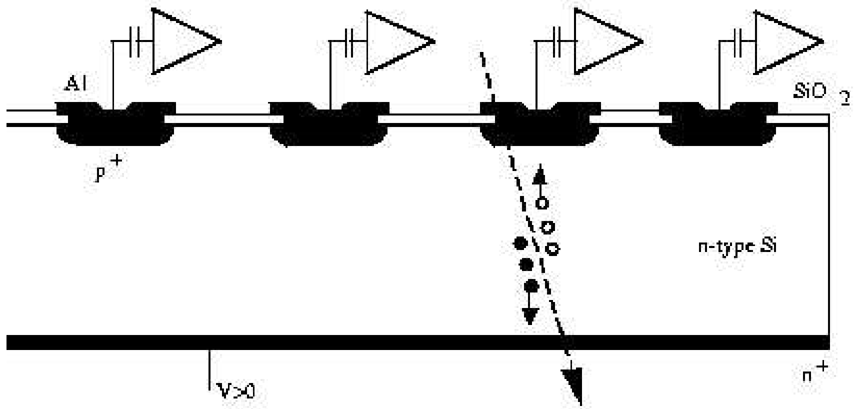


Figure 3: A silicon-microstrip detector. The electric field of the depletion voltage acts as driving force for electrons and holes. A typical thickness of the silicon wafer is 200μm

The reconstruction of individual charged tracks

- *Pattern recognition*, track search and track fitting are often performed as properly separated tasks. In pattern recognition the coordinates are linked to track canditates, and finally an optimal set of compatible tracks will be selected. Some partial iteration procedure might be necessary after the subsequent track fitting modul.

- The aim of *track fitting* is to obtain the *ultimate geomtrical resolution* from the measured coordinates. This involves the knowledge of a well defined track model, see paragraph 3.2. The hypothesis that the set of track coordinates corresponds to a real physical track is submitted to its *final confidence test*. It has turned out that in most cases the *least-squares method* (LSM) meets best the requirements of track fitting, giving an *optimal estimate* for the initial track parameters and some helpful test quantities.

2.2 Trackfitting and the least squares method

- For the transition from fitting individual tracks in general to the fit of a common vertex one or several *reference surfaces* in the neighbourhood of the vertex are chosen. A frequent choice is the inner boundary of the beam-tube surface.
- If the magnetic field is known, a trajectory (track) is defined by its initial conditions which consist of 5 parameters at the reference surface:
  - the intersection of the trajectory with the reference surface (p_1, p_2)
  - 2 directions (p_3, p_4) and the absolute value of the inverse of the momentum or the momentum vector (p_3, p_4, p_5) itself.

  We shall call these n-tupel of five parameters the *'parameter vector'* $\boldsymbol{p}$.

The reconstruction of individual charged tracks

- The next step is to construct the *track model*: The *track model* $\boldsymbol{f}(\boldsymbol{p})$ describes the intersection impact of a trajectory defined by the initial parameters $\boldsymbol{p}$ with the detector surfaces which may be obtained from pattern recognition. In a homogeneous magnetic field the tracectory would be a helix.

- If pattern recognition did a good job, i.e. $\boldsymbol{p_0}$ is sufficiently near to the true initial conditions $\boldsymbol{p}_t$, the track model can be linearized by a *Taylor expansion of first order*:

$$\boldsymbol{f}(\boldsymbol{p}) = \boldsymbol{f}(\boldsymbol{p_0}) + \left.\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{p}}\right|_{\boldsymbol{p_0}} (\boldsymbol{p} - \boldsymbol{p_0}) + O\left(\boldsymbol{p} - \boldsymbol{p_0}\right)^2 \quad \text{with} \quad \left.\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{p}}\right|_{\boldsymbol{p}_0} \equiv D \qquad (1)$$

The reconstruction of individual charged tracks

- The corresponding *measurement vector* is the sum of *true intersections* $\boldsymbol{f}(\boldsymbol{p_t})$ smeared by the *measurement error* $\boldsymbol{\epsilon_r}$ which is a random quantity:

$$\boldsymbol{c_r} = \boldsymbol{f}(\boldsymbol{p_t}) + \boldsymbol{\epsilon_r} \qquad (2)$$

(whenever it is not necessary to distinguish between the random property of a measurement error and a realisation of a measurement the subscript 'r' is omitted.)

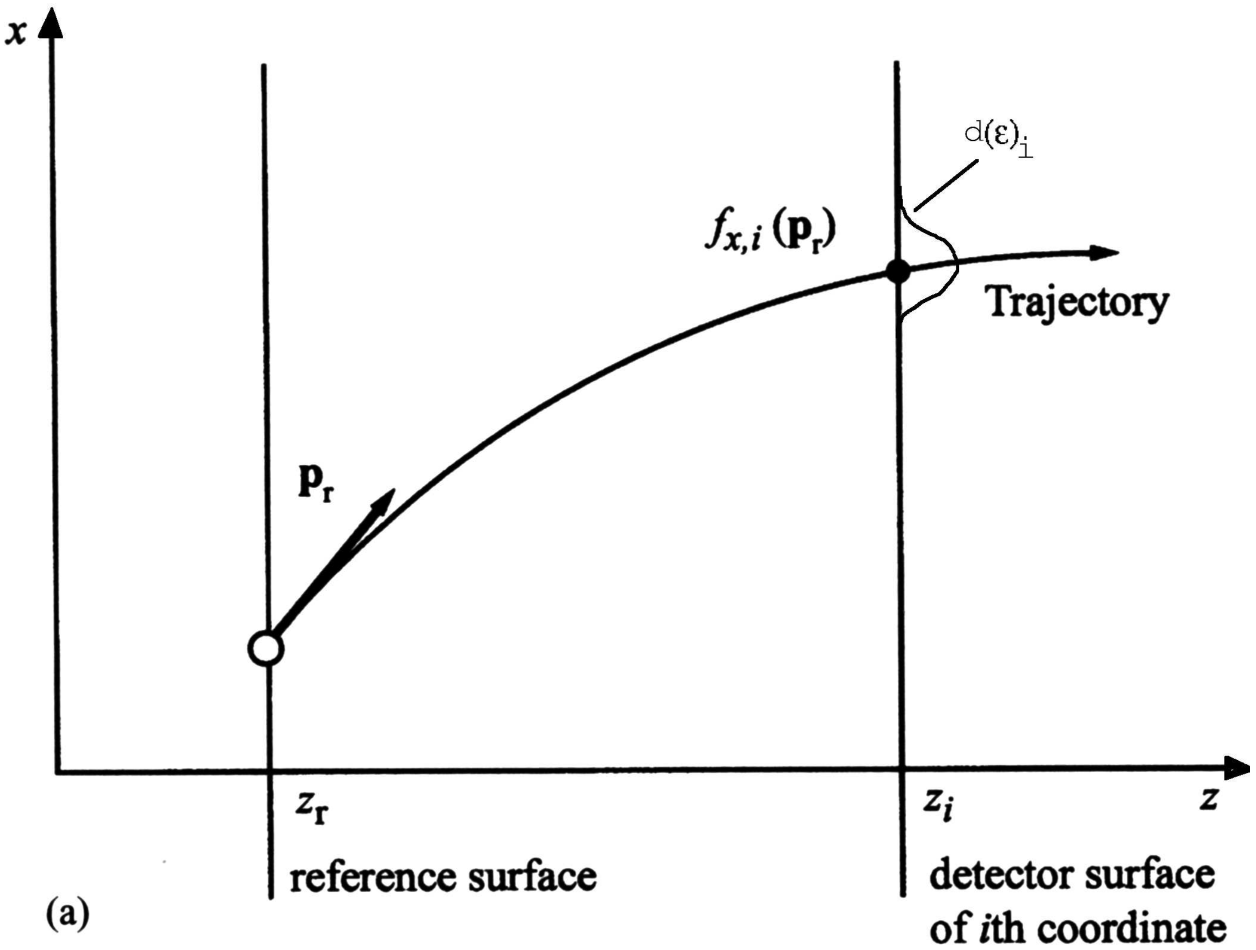


Figure 4: Reference plane, the five dimensional track parameter $\boldsymbol{p_r}$ and the 'impact functions' $f_i(\boldsymbol{p_r})$, $\quad i = 1, ..., n$

The reconstruction of individual charged tracks

- The least squares method (LSM) applied in track fitting tries to *minimize the function*

$$\mathcal{M}(\boldsymbol{p}) = [\boldsymbol{f}(\boldsymbol{p_0}) + D(\boldsymbol{p} - \boldsymbol{p_0}) - \boldsymbol{m}]^T \cdot \\ \cdot \; W[\boldsymbol{f}(\boldsymbol{p_0}) + D(\boldsymbol{p} - \boldsymbol{p_0}) - \boldsymbol{m}] \quad (3)$$

where $\boldsymbol{m}$ is a 'realisation' of the random quantity $\boldsymbol{c}$, i.e. a specific set of measurement coordinats belonging to one track[a].

[a] V is the covariance matrix of the measurement errors $V = < \boldsymbol{\epsilon}_r^T \boldsymbol{\epsilon_r} >$, while W, the 'weight matrix' is its inverse.

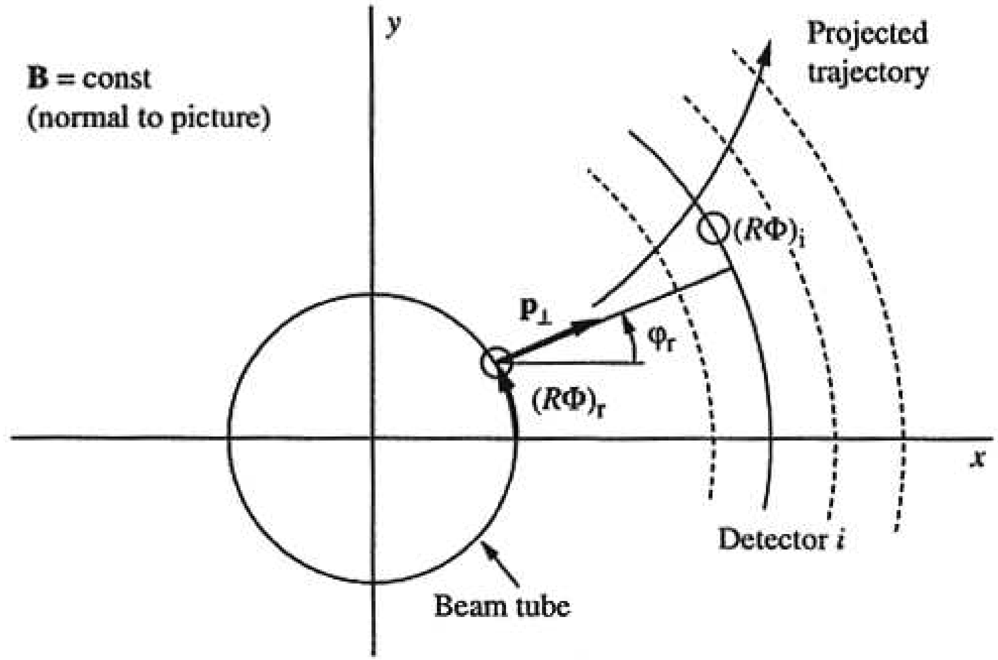


Figure 5: In many storage ring experiments, the magnetic field is rotationally symmetric around the beam axis. In this case the reference surface is often chosen to be a cylinder

- Differentiating $\mathcal{M}$ with respect to $\boldsymbol{p}$ and putting $\frac{\partial \mathcal{M}}{\partial \boldsymbol{p}} = 0$ yields for the 'fitted' value of $\tilde{\boldsymbol{p}}$:

$$\tilde{\boldsymbol{p}} = \boldsymbol{p_0} + (D^T W D)^{-1} D^T W \, (\boldsymbol{m} - \boldsymbol{f}(\boldsymbol{p_0})) \tag{4}$$

This can be interpreted as 'pulling back' the measurement to a closest point compatible with the 'constraint surface' in the coordinate space (estimation).

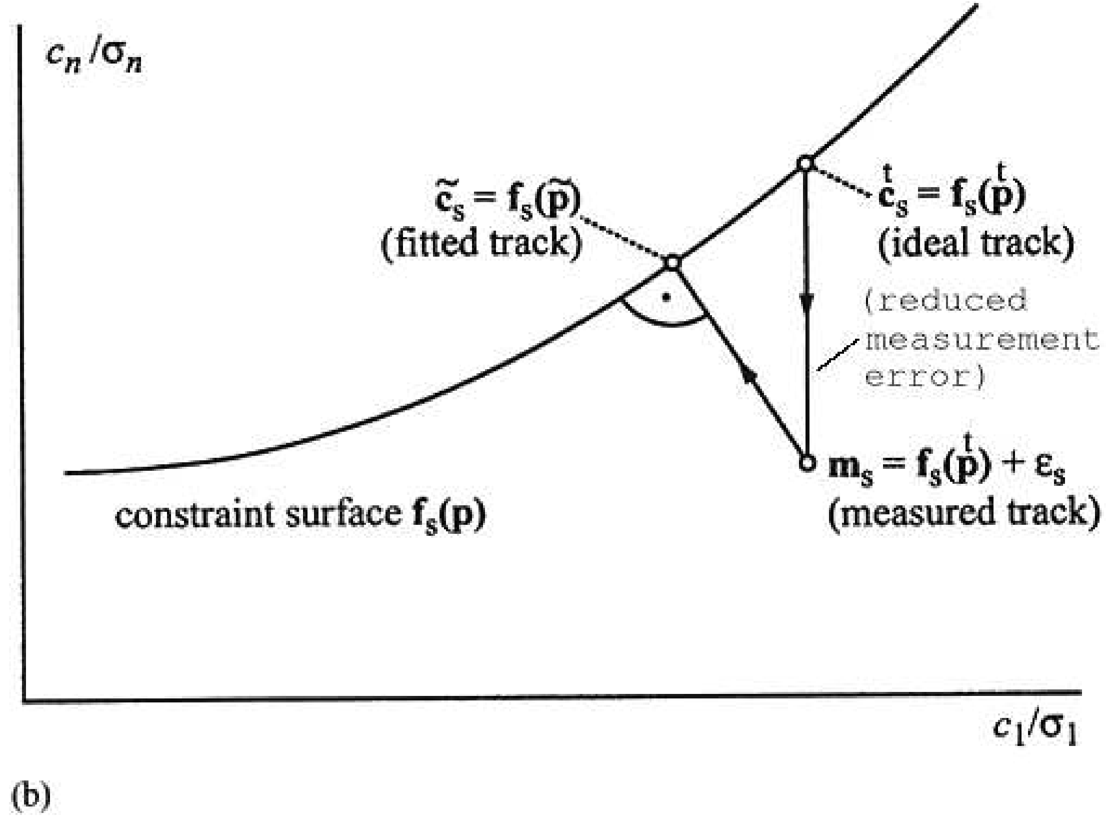


Figure 6: The fitted value $\tilde{\boldsymbol{p}}$ on the five dimensional constraint surface (the subscript 'S' denotes scaled by the measurement error σ. Uncorrelated errors are assumed for the purpose of graphical representation

The reconstruction of individual charged tracks

- With a linear model, the LSM is the *linear unbiased estimator with least variance* within the class of linear estimations ('Gauss Markov theorem').
- The covariance matrix ('*error matrix*') of $\tilde{\boldsymbol{p}}$ is obtained by *error propagation*:

$$\boldsymbol{V} = cov(\tilde{\boldsymbol{p}}) =< (\tilde{\boldsymbol{p}} - \boldsymbol{p}_t)^T \cdot (\tilde{\boldsymbol{p}} - \boldsymbol{p}_t) >= (D^T W D)^{-1} \quad (5)$$

- The set $\tilde{\boldsymbol{p}}_{\boldsymbol{j}}(j = 1...n$, number of charged tracks) and the corresponding covariance matrices $\boldsymbol{V}_j(\tilde{\boldsymbol{p}}_{\boldsymbol{j}})$ is a compressed form of the detector information. In case of Gaussian measurement errors this compression is even sufficient, and the *full information is passed to the subsequent vertex fit.*

3 The vertex fit

3.1 The standard vertex fit

- From a formal point of view, the results of the single track fit, the set $\tilde{\boldsymbol{p}}_{\boldsymbol{k}}$ $(k = 1 \ldots n)$ can be considered as *virtual measurements and input to the vertex fit.*
- For each reconstructed track there are 5 fitted track parameters $\boldsymbol{p}_{\boldsymbol{k}}$ and the corresponding covariance matrix $\boldsymbol{V}_{\boldsymbol{k}} = \boldsymbol{G}_{\boldsymbol{k}}{}^{-1}$, defined at some reference surface. Since the tracks have been measured independently, the errors are also independent, with the consequence that the global $(5n \times 5n)$ covariance matrix of all $\boldsymbol{p}_{\boldsymbol{k}}$ is (5×5) block-diagonal.
- When passing from the outer side of the beam tube to its inner side, one has still to correct for the *multiple scattering occuring in the beam tube*:

$$V(\tilde{\boldsymbol{p}}) \leftarrow V(\tilde{\boldsymbol{p}}) + V_{MS} \tag{6}$$

where V_{MS} accounts for the effect of multiple scattering which acts mainly on the covariance matrix of the directions.

The vertex fit

- The aim is the estimation of the $(3n+3)$-dimensional 'parameter vector' composed of the *vertex position* $\boldsymbol{x}$ and of the *momentum vectors* $\boldsymbol{q_k}$ of all tracks *at the common vertex*

- The functional dependence of the track parameters $\boldsymbol{p_k}$ on the vertex parameters $\boldsymbol{x}$ and $\boldsymbol{q_k}$ are again given by nonlinear functions $\boldsymbol{p_k} = \boldsymbol{h_k}(\boldsymbol{x}, \boldsymbol{q}_k)$. The functions $\boldsymbol{h_k}$ are therefore *Taylor-expanded* to first order at some point $(\boldsymbol{x}_0, \boldsymbol{q}_{k,0})$:

$$\boldsymbol{h}_k(\boldsymbol{x}, \boldsymbol{q}_k) \cong \boldsymbol{h}_k(\boldsymbol{x}_0, \boldsymbol{q}_{k,0}) + \boldsymbol{A}_k \cdot (\boldsymbol{x} - \boldsymbol{x}_0) + \boldsymbol{B}_k \cdot (\boldsymbol{q}_k - \boldsymbol{q}_{k,o}) = \boldsymbol{c}_k + \boldsymbol{A}_k \boldsymbol{x} + \boldsymbol{B}_k \boldsymbol{q}_k \quad (7)$$

where $\boldsymbol{A}_k = \left.\frac{\partial \boldsymbol{h}_k}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}_0, \boldsymbol{q}_{k,0}}$ and $\boldsymbol{B}_k = \left.\frac{\partial \boldsymbol{h}_k}{\partial \boldsymbol{q}_k}\right|_{\boldsymbol{x}_0, \boldsymbol{q}_{k,0}}$ are the respective (5×3) *derivative matrices* computed at the expansion point, and the $\boldsymbol{c}_k$ are constant.

- If the position of the vertex is known a priori to some extent as is the case for the *interaction region* of a storage ring, this knowledge can be added as an *independent measurement of the vertex position*, with its proper covariance matrix, e.g. form the variance of the beam particle defolded from the measured interaction profile:

$$\mathcal{M} \leftarrow \mathcal{M} + \frac{(x - <x>)^2}{\sigma^2_{IP,x}} + \frac{(y - <y>)^2}{\sigma^2_{IP,y}} \left(+ \frac{(z - <z>)^2}{\sigma^2_{IP,z}} \right)^{\text{a}} \tag{8}$$

which is valid only for the primary vertex.

- This leads to the following *linear model*:

$$\begin{pmatrix} \boldsymbol{x}_I \\ \boldsymbol{p}_1 \\ \cdot \\ \cdot \\ \cdot \\ \boldsymbol{p}_n \end{pmatrix} = \underbrace{\begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} & \cdot & \cdot & \boldsymbol{0} \\ \boldsymbol{A}_1 & \boldsymbol{B}_1 & \boldsymbol{0} & \cdot & \cdot \\ \boldsymbol{A}_2 & \boldsymbol{0} & \boldsymbol{B}_2 & \boldsymbol{0} & \cdot \\ \cdot & \boldsymbol{0} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \boldsymbol{A}_n & \boldsymbol{0} & \cdot & \cdot & \boldsymbol{B}_n \end{pmatrix}}_{\mathcal{D}} \cdot \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{q}_1 \\ \cdot \\ \cdot \\ \cdot \\ \boldsymbol{q}_n \end{pmatrix} + \begin{pmatrix} \epsilon_0 \\ \epsilon_1 \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_I \end{pmatrix} \tag{9}$$

- The first row is only needed if the interaction profile is included in the LSM-Ansatz, (see equation (9))

[a] optional; can also be suppressed by zero weight assignment.

The vertex fit

Estimates of $\boldsymbol{x}$ and all $\boldsymbol{q}_k$ can be determined by the LSM

$$\mathcal{M}_{vertex}(\boldsymbol{x}, \boldsymbol{q}_1, \dots \boldsymbol{q}_n) = \left(\Delta\boldsymbol{x}^T, \Delta\boldsymbol{q}_1^T, \dots \Delta\boldsymbol{q}_n^T\right)\left(\mathcal{D}^T\mathcal{G}\mathcal{D}\right)^{-1}(\dots) \tag{10}$$

(see also equation (4)) and the solution of

$$\begin{pmatrix} \tilde{\boldsymbol{x}} \\ \tilde{\boldsymbol{q}}_1 \\ . \\ . \\ . \\ \tilde{\boldsymbol{q}}_n \end{pmatrix} = \boldsymbol{M}^{-1} \cdot \boldsymbol{N} \cdot \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{p}_1 - \boldsymbol{c}_1 \\ . \\ . \\ . \\ \boldsymbol{p}_n - \boldsymbol{c}_n \end{pmatrix} \tag{11}$$

where $\boldsymbol{M}$ is a $(3n+3) \times (3n+3)$ matrix and $\boldsymbol{N}$ is a $(3n+3) \times (5n+3)$ matrix, both of which contain $\boldsymbol{A}_k$,$\boldsymbol{B}_k$ and $\boldsymbol{G}_k = \boldsymbol{V}_k^{-1}$.

- The evaluation of $\boldsymbol{M}^{-1}$ (which is in fact the *covariance of the estimate*) and of $\boldsymbol{N}$ is strongly facilitated by the *block stucture of the derivative matrix* $\mathcal{D}$ (see Equation (9)) *and of the weight matrix* $\mathcal{G}$ without interaction profile:

The vertex fit

$$\mathcal{G} = \begin{pmatrix} G_I & 0 & & & & 0 \\ 0 & G_1 & & & & \\ & & G_2 & & & \\ & & & \cdot & & \\ & & & & \cdot & 0 \\ 0 & & & & 0 & G_n \end{pmatrix} = \begin{pmatrix} V_I^{-1} & 0 & & & & 0 \\ 0 & V_1^{-1} & & & & \\ & & V_2 & & & \\ & & & \cdot & & \\ & & & & \cdot & 0 \\ 0 & & & & 0 & V_n^{-1} \end{pmatrix} \tag{12}$$

with

$$M = (\mathcal{D}^T \mathcal{G} \mathcal{D})^{-1}, N = \mathcal{D}^T \mathcal{G} \tag{13}$$

and the covariance for the interaction profile (only if available and used, and again only for the primary vertex):

$$V_I = \begin{pmatrix} \sigma_{I,x}^2 & 0 & 0 \\ 0 & \sigma_{I,y}^2 & 0 \\ 0 & 0 & \sigma_{I,z}^2 \end{pmatrix} \tag{14}$$

3.2 The track model

- In most of modern collider experiments a solenoid creates a quite homogeneous magnetic field $\boldsymbol{B}$ in the central part of the detector; and in the vertex region inside the beam tube the assumption of an homogeneous field is a very benefitial approximation for vertex reconstruction.
- If $\boldsymbol{B} = (0, 0, B_z)^T$ is chosen to be parallel to the z-axis, and in the equations of motion the time is replaced by the path length s:

$$\ddot{\boldsymbol{x}} \approx \dot{\boldsymbol{x}} \times \boldsymbol{B}(\boldsymbol{x}) \quad , \quad \dot{\boldsymbol{x}}\ddot{\boldsymbol{x}} = (\dot{\boldsymbol{x}} \cdot (\dot{\boldsymbol{x}} \times \boldsymbol{B}(\boldsymbol{x}))) = 0 \quad \rightarrow \quad \dot{\boldsymbol{x}} = \boldsymbol{v} = \text{constant} \tag{15}$$

and the equations of motion takes the form (see also fig. 2.2):

$$\left.\begin{aligned} \frac{d^2x}{ds^2} &= \frac{\kappa q}{P}\frac{dy}{ds}B_z \\ \frac{d^2y}{ds^2} &= -\frac{\kappa q}{P}\frac{dx}{ds}B_z \\ \frac{d^2z}{ds^2} &= 0 \end{aligned}\right\} \tag{16}$$

where κ is a constant, q the signed charge in units of the elementary charge, P the absolute value of the momentum, and $\boldsymbol{B}$ the magnetic field.

The vertex fit

- For the virtual measurements (i.e the results of the single track fit) it is common practice, at least in the barrel region of a collider detector, to choose the following quantities (possibly after some transformations with 'error propagation')

$$\left.\begin{array}{l} p_1 = (R\Phi)_r \quad , \qquad p_2 = z_r \\ p_3 = \varphi_r \quad \text{or}^{\text{b}} \quad (\varphi_r - \Phi_r) \quad , \qquad p_4 = \tan\lambda_r \\ p_5 = (\frac{1}{hR_H})_r \end{array}\right\} \tag{17}$$

with $(R\Phi)_r, z_r$ the cylinder coordinates of the intersection point with the reference zylinder, φ_r, λ_r the azimuth and dip angle of the direction and R_H the radius of the projected helix. Note that R_H is proportional to the transverse momentum. $h = \pm 1$ is the sense of rotation.

---

[b] for better use of the rotational invariance

The vertex fit

- The *track model* used during the vertex fit is a linear variation around the track defined by the expansion point $\boldsymbol{x}_0, \boldsymbol{q}_{k,0}$ following the fomula for a *helix*. The expansion point should be chosen in the vincinity of the possible intersection of all helices from the same vertex:

$$\left.\begin{aligned} x(s) &= x_0 + R_H \left[\cos\left(\Psi_0 + \tfrac{s\cos\lambda}{hR_H}\right) - \cos\Psi_0\right] \\ y(s) &= y_0 + R_H \left[\sin\left(\Psi_0 + \tfrac{s\cos\lambda}{hR_H}\right) - \sin\Psi_0\right] \\ z(s) &= z_0 + s\sin\lambda \end{aligned}\right\}$$

$$\text{or with} \quad \varphi(s) = \frac{s\cos\lambda}{hR_H}:$$

$$\left.\begin{aligned} x(\varphi) &= x_0 + hR_H\left(\sin\varphi - \sin\varphi_0\right) \\ y(\varphi) &= y_0 - hR_H\left(\cos\varphi - \cos\varphi_0\right) \\ z(s) &= z_0 + hR_H \cdot \tan[\lambda \cdot (\varphi - \varphi_0)] \end{aligned}\right\} \qquad (18)$$

The vertex fit

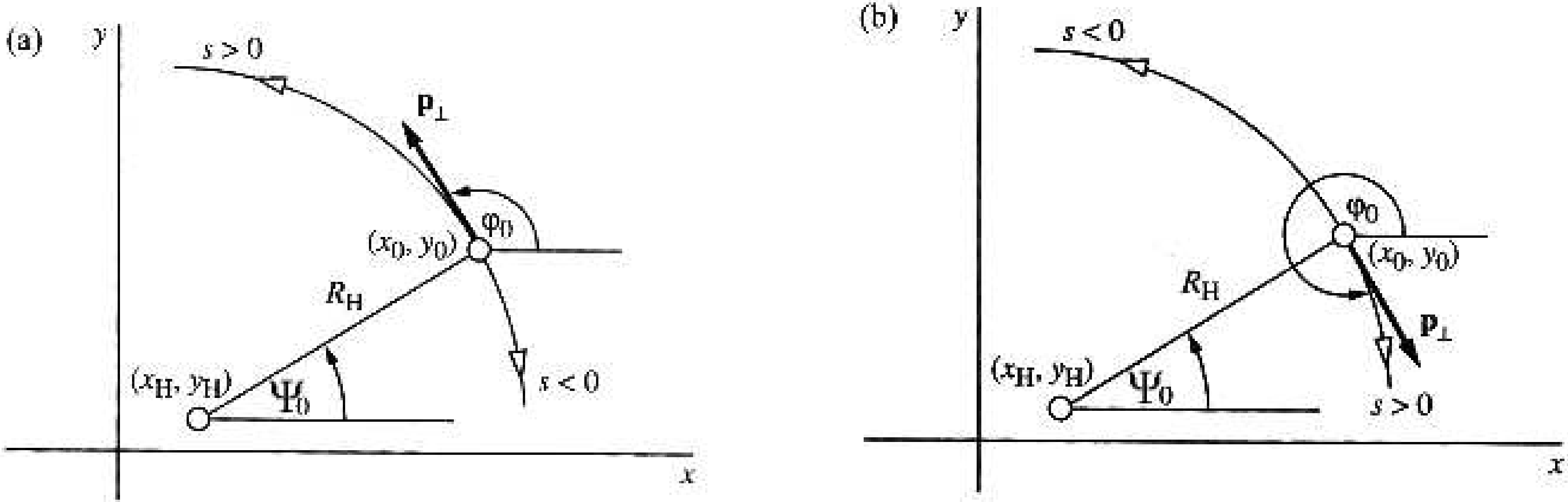


Figure 7:
$\varphi = \Psi + h\frac{\pi}{2}$
(a)For a constant magnetic field parallel to the z axis, the solution of the equation of motion is a helix. The projection on the xy plane is a circle, which is chosen here with a positive sense of rotation $(h = +1)$.
(b)Projected helix with negative sense of rotation $(h = -1)$. The projected curvature is $\frac{1}{R_H}$ and the curvature in space $\left(\equiv \left|\frac{d^2x}{ds^2}\right|\right)$ is $\cos^2 \frac{\lambda}{R_H}$

3.3 Tests

- All previously said concerning the properties of the LSM requires of course that the assumption made for the *'ingredients'* to be *sufficiently close to reality*:
  - Virtual measurements
    - no bias
    - 'errors' (in the sense of their covariance matrix) well understood
  - Track model
    - correct basic model (including the precise knowledge of the magnetic field)
    - proper choice of variables appropriate for a linear expansion, and of course a reasonable expansion point

The vertex fit

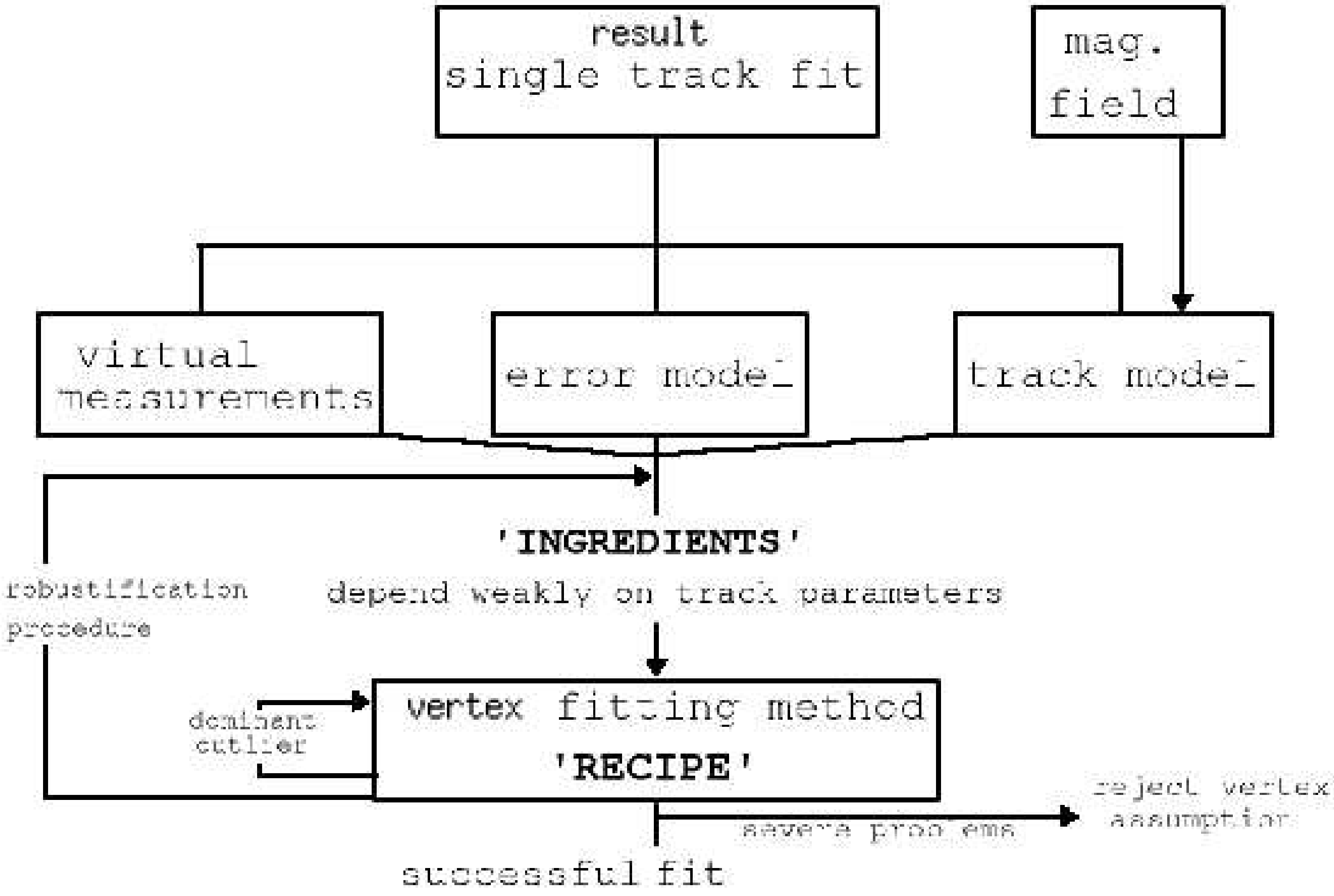


Figure 8: 'Ingredients' and 'recipe' of the vertex fit

The vertex fit

- Test 1: The *'χ^2' and its distribution*: The mathematical definition of the 'χ^2'-distribution with n degrees of freedom is the probability function of the square of the radius vector of an n-dimensional normal standard distribution. In the case of an estimation of a linear model and normally distributed errors, the χ^2-distribution has $n - m$ degrees of freedom, where m is the number of free parameters, i.e. $\mathcal{M}(\tilde{\boldsymbol{p}})$ (see also equation (3)) *obeys a χ^2-distribution.* If the errors are not nomally distributed, but the weight is chosen as the inverse of the covariance matrix, the distribution 'pseudo-χ^2' will be a distorted χ^2-distribution, with an expectation of the $< \chi^2 >$, still corresponding to the number of degrees of freedom, namely $n - m$ (*equivalence class up to second order averages*).

- The χ^2 is defined as (see also (3), (9) and (11)):

$$\mathcal{M}_{vertex}\left(\boldsymbol{x}, \boldsymbol{q}_1, \ldots \boldsymbol{q}_n\right) = \left(\Delta\boldsymbol{x}^T, \Delta\boldsymbol{q}_1^T, \ldots \Delta\boldsymbol{q}_n^T\right)\left(\mathcal{D}^T\mathcal{G}\mathcal{D}\right)^{-1}(\ldots) \tag{19}$$

- Test 2: The 'pull quantities' (in general called stretch quantities).

  While the global χ^2-test is defined to test the hypothesis for the set of tracks emerging forme one interaction vertex, the pull quantities give some indication of which are the most incompatible virtual measurements. The residuals between the fitted tracks and the true values are accessible only in simulated data, while the residuals with respect to the measured values are of course available from data, too.

The covariance matrix for the resulting residuals is calculated by error propagation

$$\mathcal{V} = \mathcal{G}^{-1} - \mathcal{D}(\mathcal{D}^T\mathcal{G}\mathcal{D})^{-\infty}\mathcal{D}^T \tag{20}$$

with the pull quantities

$$\frac{(\text{virtual measurement - impact of fitted track})_i}{\sqrt{\mathcal{V}_{ii}}} \tag{21}$$

- The test quantities have a twofold meaning: The overall correctness of the fit procedure (and therefore also of the 'ingredients') are confirmed by the correct probability density functions, while for the individual event large χ^2s or excessive pull quantities are considered as a warning, as most of the possible mistakes (going back as much as to a background fit in the single track reconstruction) increase these quantities. Other characteristics than excessive pull quantities for low moments can be also a hint that multiple scattering is underestimated in the covariance matrix.

4 The Kalman filter

4.1 Advantages of the Kalman filter

- In the classical approach, a global vertex fit involves the inversion of a $(3n+3) \times (3n+3)$ matrix, requiring $O(n^3)$ arithmetic operations and thus becoming prohibitive in case of high track multiplicities. However, this fact is not so dramatic as in the single track fit procedure where the block structure of the covariance matrix is lost because of the off diagonal terms introduced by multiple scattering.
- The *Kalman filter and smoother require only* $O(n)$ *operations* for the fit, resp. $O(n^2)$ operations if the full covariance matrix is needed. In addition, they provide a good test criterium for discriminating between tracks originating from this vertex against tracks belonging to another vertex or to background ('outlier problem').

4.2 System and measurement equation

- Initially, the state vector consists only of the prior information about the vertex position $\boldsymbol{x}_0$, and its covariance matrix $\boldsymbol{C}_0$.
  - In case of the primary vertex, that information may be obtained from the beam interaction profile.
  - If no such information is available, $\boldsymbol{x}_0$ is arbitrary, and $\boldsymbol{C}_0^{-1}$ is set to zero.
- For each track, the state vector is augmented by the 3-momentum vector $\boldsymbol{q}_k$. For the linearized track model see (9).
- Contrary to the single fit there is no *'process noise'*, and the system equation for the vertex position is simply the identity:

$$\boldsymbol{x}_k(\boldsymbol{x}_{k-1}) = \boldsymbol{x}_{k-1} =: \boldsymbol{x} \tag{22}$$

- The measurement equation, after substracting the inhomogeneous constants $\boldsymbol{c}_k$ is

$$\boldsymbol{p}_k = A_k \boldsymbol{x} + B_k \boldsymbol{q}_k \tag{23}$$

see equ. (9)

The Kalman filter

- Adding a new track to a vertex fitted with $k-1$ tracks so far:

$$
\begin{aligned}
\tilde{\boldsymbol{x}}_k &= \boldsymbol{C}_k[\boldsymbol{C}_{k-1}^{-1}\tilde{\boldsymbol{x}}_{k-1} + \boldsymbol{A}_k^T\boldsymbol{G}_k^B\boldsymbol{p}_k], \quad ^{a} \\
\mathrm{cov}(\tilde{\boldsymbol{x}}_k) &= \boldsymbol{C}_k = (\boldsymbol{C}_{k-1}^{-1} + \boldsymbol{A}_k^T\boldsymbol{G}_k^B\boldsymbol{A}_k)^{-1}, \\
&\text{with} \quad \boldsymbol{G}_k^B = \boldsymbol{G}_k - \boldsymbol{G}_k\boldsymbol{B}_k\boldsymbol{W}_k\boldsymbol{B}_k^T\boldsymbol{G}_k.
\end{aligned} \tag{24}
$$

- Now the tracks can be calculated.

$$
\begin{aligned}
\tilde{\boldsymbol{q}}_k &= \boldsymbol{W}_k\boldsymbol{B}_k^T\boldsymbol{G}_k(\boldsymbol{p}_k - \boldsymbol{A}k\tilde{\boldsymbol{x}}_k), \quad ^{b} \\
\mathrm{cov}(\tilde{\boldsymbol{q}}_k) &= \boldsymbol{D}_k = \boldsymbol{W}_k + \boldsymbol{W}_k\boldsymbol{B}_k^T\boldsymbol{G}_k\boldsymbol{A}_k\boldsymbol{C}_k\boldsymbol{A}_k^T\boldsymbol{G}_k\boldsymbol{B}_k\boldsymbol{W}_k = {}^{c} \\
&= \boldsymbol{W}_k + \boldsymbol{E}_k^T\boldsymbol{C}_k^{-1}\boldsymbol{E}_k, \\
\mathrm{cov}(\tilde{\boldsymbol{x}}_k, \tilde{\boldsymbol{q}}_k) &= \boldsymbol{E}_k = -\boldsymbol{C}_k\boldsymbol{A}_k^T\boldsymbol{G}_k\boldsymbol{B}_k\boldsymbol{W}_k, \\
\text{with} \quad \boldsymbol{W}_k &= (\boldsymbol{B}_k^T\boldsymbol{G}_k\boldsymbol{B}_k)^{-1}
\end{aligned}
$$

1. The block structure of the total covariance matrix of all virtual measurements facilitates to add a new track k to a vertex already fitted with $k-1$ tracks, allowing feasible strategies to deal with complex topologies:

This is in fact just a weighted mean. The first term is trivial and contains the vertex information from $k-1$ tracks. The second term adds the information as obtained from the new track (contained in the kth virtual measurement) and used for the improvement of the vertex position.

2. The next step evaluates the momentum $\boldsymbol{q}_k$ at the vertex (note that $\tilde{\boldsymbol{x}}_k$ and $\tilde{\boldsymbol{q}}_k$ are not the final estimates, but only the information of the $k-1$ previous tracks plus this one that have been implemented so far. The implementation of the full information is done by the smoother (see below).
3. As usual the covariance matrices are obtained by error propagation. Again, the formula are drastically simplified by the block structure of the global error matrices of the virtual measurements.

- The residuals with respect to the virtual measurements are:

$$\boldsymbol{r}_k = \boldsymbol{p}_k - \boldsymbol{A}_k\tilde{\boldsymbol{x}}_k - \boldsymbol{B}_k\tilde{\boldsymbol{q}}_k,$$

- There is a useful quantity which checks the compatibility of the added track with the

The Kalman filter

previous ones, the 'Filter χ^2' of this track.

$$\chi^2_{k,F} = (\tilde{\boldsymbol{x}}_k - \tilde{\boldsymbol{x}}_{k-1})^T \boldsymbol{C}^{-1}_{k-1} (\tilde{\boldsymbol{x}}_k - \tilde{\boldsymbol{x}}_{k-1}) + \boldsymbol{r}_k^T \boldsymbol{G}_k \boldsymbol{r}_k,$$

- The *total* χ^2 is evaluated, it would also have been obtained *from a global vertex fit* with the k tracks considered so far:

$$\chi^2_k = \chi^2_{k-1} + \chi^2_{k,F}$$

- The chi-squares of the filter step $\chi^2_{k,F}$ have two degrees of freedom. They are independent random variables. The current total chi-square of the fit, χ^2_k, has $2k$ degrees of freedom if there is prior information on the vertex position, and $2k-3$ degrees of freedom otherwise.
- The Kalman filter step is repeated until the last track n has been added to the vertex fit.
- Due to the simple form of the system equation, the final estimate of the vertex position after inclusion of n tracks can actually be computed in "one go":

$$\tilde{\boldsymbol{x}}_n = \boldsymbol{C}_n[\boldsymbol{C}_0^{-1}\boldsymbol{x}_0 + \sum_{k=1}^{n} \boldsymbol{A}_k^T \boldsymbol{G}_k^B (\boldsymbol{p_k} - \boldsymbol{c}_{k,e})], \tag{25}$$

$$\mathrm{cov}(\tilde{\boldsymbol{x}}_n) = \boldsymbol{C}_n = (\boldsymbol{C}_0^{-1} + \sum_{k=1}^{n} \boldsymbol{A}_k^T \boldsymbol{G}_k^B \boldsymbol{A}k)^{-1}. \tag{26}$$

(27)

4.3 The smoother

- As there is no process noise, it requires no more than recomputing momentum vectors and covariance matrices with the final estimate of the vertex position (upperscript n means that the information of all n tracks has been implemented):

The Kalman filter

$$\tilde{\boldsymbol{q}}_k^n = \boldsymbol{W}_k \boldsymbol{B}_k{}^T \boldsymbol{G}_k (\boldsymbol{p}_k - \boldsymbol{A}_k \tilde{\boldsymbol{x}}_n), \quad (28)$$

$$\mathrm{cov}(\tilde{\boldsymbol{q}}_k^n) = \boldsymbol{D}_k^n = \boldsymbol{W}_k + \boldsymbol{W}_k \boldsymbol{B}_k{}^T \boldsymbol{G}_k \boldsymbol{A}_k \boldsymbol{C}_n \boldsymbol{A}_k{}^T \boldsymbol{G}_k \boldsymbol{B}_k \boldsymbol{W}_k =$$

$$= \boldsymbol{W}_k + \boldsymbol{E}_k^{nT} \boldsymbol{C}_n^{-1} \boldsymbol{E}_k^n, \quad (29)$$

$$\mathrm{cov}(\tilde{\boldsymbol{x}}_n, \tilde{\boldsymbol{q}}_k^n) = \boldsymbol{E}_k^n = -\boldsymbol{C}_n \boldsymbol{A}_k{}^T \boldsymbol{G}_k \boldsymbol{B}_k \boldsymbol{W}_k. \quad (30)$$

- If required, the full covariance matrix and the total chi-square can be computed after smoothing:

$$\mathrm{cov}(\tilde{\boldsymbol{q}}_k^n, \tilde{\boldsymbol{q}}_j^n) = \boldsymbol{W}_k \boldsymbol{B}_k{}^T \boldsymbol{G}_k \boldsymbol{A}_k \boldsymbol{C}_n \boldsymbol{A}_j{}^T \boldsymbol{G}_j \boldsymbol{B}_j \boldsymbol{W}_j = \boldsymbol{E}_k^{nT} \boldsymbol{C}_n^{-1} \boldsymbol{E}_j^n, \quad k \neq j, \quad (31)$$

$$\boldsymbol{r}_k^n = \boldsymbol{p}_k - \boldsymbol{A}_k \tilde{\boldsymbol{x}}_n - \boldsymbol{B}_k \tilde{\boldsymbol{q}}_k^n, \quad (32)$$

$$\chi_n^2 = (\boldsymbol{x}_0 - \tilde{\boldsymbol{x}}_n)^T \boldsymbol{C}_0^{-1} (\boldsymbol{x}_0 - \tilde{\boldsymbol{x}}_n) + \sum_{k=1}^{n} \boldsymbol{r}_k^{nT} \boldsymbol{G}_k \boldsymbol{r}_k^n, \quad (33)$$

5 Efficient robustification against outliers and background

5.1 Definition of outliers for single track fit and vertex fit

- For a single track we can define two reasons for outliers:
  - *'outliers'* which have been *generated by the particle* under investigation, but with an error much larger than expected from the individual detector. The most common process to generate such outliers is the presence of energetic electrons (δ-rays) in gaseous and semiconductor detectors.
  - The measured coordinate is a *noise signal* that was wrongly associated with the track by the pattern recognition; it can be either genuine detector noise, or picked up from another particle's set of measurements.

Efficient robustification against outliers and background

In the framework of the LSM the outlier problem could be handled globally by modifying the error matrix by including the contribution of outliers, thus ensuring the proper propagation of errors. However, in most cases this would spoil the overall resolution, and the χ^2 would be distorted, so that a χ^2-cut would be non-transparent and could spoil the overall normalization and - if correlated with $\tilde{\boldsymbol{p}}$ - introduce a bias.

- In case of a single track fit the optimal resolution is obtained by a *careful balance between removal and downweigthing*. Note that this is a subtle *deviation from the linearity of the estimation*, and the 'Gauss Markov' theorem is no more applicable.
- For the vertex fit the situation of outliers is different. In general - if the detector works well and the pattern recognition program is well tuned - all virtual measurements should belong to tracks from real particles. Outliers originate by:
  - *Virtual measurements* distorted by the single track fit procedure by bad detector information and/or by the single track algorithm
  - *Genuine outliers* are now the tracks belonging to another vertex than the one under investigation. Therefore - if we need the full exclusive topology - removal implies the obligation to associate it with another vertex!

5.2 Robustification strategies

5.2.1 The combinatorial approach

- *A symmetric test uses the chi-square of the smoother*, which has two degrees of freedom as well as the chi-square of the filter. The estimate results from *removing* track k from the fitted vertex by an "inverse Kalman filter", using $-\boldsymbol{G}_k$ instead of $\boldsymbol{G}_k$ [a]:

$$\tilde{\boldsymbol{x}}_k^{n*} = \boldsymbol{C}_k^{n*}[\boldsymbol{C}_n^{-1}\tilde{\boldsymbol{x}}_n - \boldsymbol{A}_k{}^T\boldsymbol{G}_k{}^B\boldsymbol{p}_{\boldsymbol{k}}], \tag{34}$$

$$\mathrm{cov}(\tilde{\boldsymbol{x}}_k^{n*}) = \boldsymbol{C}_k^{n*} = (\boldsymbol{C}_n^{-1} - \boldsymbol{A}_k{}^T\boldsymbol{G}_k{}^B\boldsymbol{A}_k)^{-1}, \tag{35}$$

$$\chi^2_{k,S} = (\tilde{\boldsymbol{x}}_n - \tilde{\boldsymbol{x}}_k^{n*})^T(\boldsymbol{C}_k^{n*})^{-1}(\tilde{\boldsymbol{x}}_n - \tilde{\boldsymbol{x}}_k^{n*}) + \boldsymbol{r}_k^{n\,T}\boldsymbol{G}_k\boldsymbol{r}_k^n. \tag{36}$$

[a] note the full effect on W_k and G_k^B as well!

Efficient robustification against outliers and background

- In contrast to the filter, the *smoothed chi-squares* $\chi^2_{k,S}$ *are no longer independent* and therefore do not sum up to the total χ^2. But they are a more powerful test and offer a more symmetric test criterium than the filtered $\chi_{k,F}$. If there are several outlier tracks however, the estimates $\tilde{\boldsymbol{x}}_k^{n*}$ are biased by the remaining outliers, and the power of the test decreases.
- The basic but costly strategy for the reconstruction of a multivertex topology would be:
  - assembling the tracks to vertices
  - fitting the vertices by a Kalman filter
  - defining a set of tracks with large smoothed χ^2
  - trying to assemble the tracks from the set chosen above to the closest competitive vertex if lying on the same side as the residual, or some other trivial checks
  - making a final decision
- The bottleneck of this method is that the vertices fitted in a first approach might have been strongly distorted by one or more wrongly associated tracks, therefore the costly combinatorial approach must be iterated.

- The method could be made more efficiently by *downweighting the outliers during the first run*, thus making the first run more stable against wrongly associated tracks. A well known method is the M-Estimator.

5.2.2 The M-estimator

- If *only one vertex is expected* straight forward methods can be applied to get a *robust estimation* of the vertex position. This implies of course that sufficient tracks are available to *compete with the outliers*. The knowledge of the interaction profile is further important information stabilising the estimate of the position of a (primary) vertex and the correlated momenta.
- The idea is to *modify the objective Function* $\mathcal{M}$ (equ. (3) for single track fit, equ. (8) and (19) for the vertex fit) so that outliers have less influence on the estimate by *downweighting* the measurements of *large residuals*. Note that, although the track model is still linear, *the downweighting procedure is a deviation from the class of linear estimators*.

Efficient robustification against outliers and background

- For the M-estimator the error matrix must be diagonalised. Therefore the *M-estimator is especially suitable for the vertex fit*, because no (virtual) measurements from the individual tracks transform into each other[b].

- The M-estimator is based on a generalized objective function:

$$\mathcal{M}(\boldsymbol{x}) = \sum_{i=1}^{n} \Psi(r_i/\sigma_i), \tag{37}$$

- $\Psi(t)$ is a function specified in terms of its derivative $\psi(t) = d\Psi/dt$:
  - If $\psi(t) = t$, we obtain the L_2- (least-squares) estimator.
  - If $\psi(t) = \text{sgn}(t)$, we obtain the L_1-estimator.
  - If $\psi(t) =$ Huber's function, we get "our" M-estimator:

[b] This would be the case for the coordinates of the single track fit in the presence of multiple scattering.

Efficient robustification against outliers and background

$$\psi(t) = \begin{cases} t, & |t| \leq R \\ R \cdot \mathrm{sgn}(t), & |t| > R. \end{cases} \tag{38}$$

($R =$ *constant of robustness*, usually chosen between 1 and 3).

- $\mathcal{M}$ is minimized with respect to $\boldsymbol{x}$:

$$\partial \mathcal{M} / \partial x_j = \sum_{i=1}^{u} (r_i/\sigma_i) \cdot (a_{ij} w_i / \sigma_i) = 0, \quad j = 1, \ldots, m, \tag{39}$$

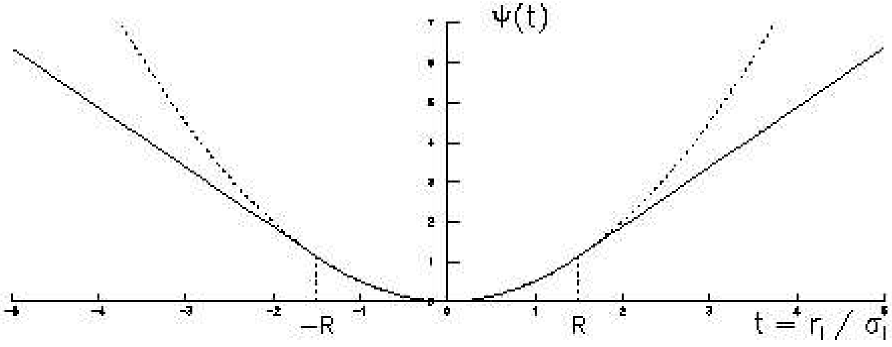
with

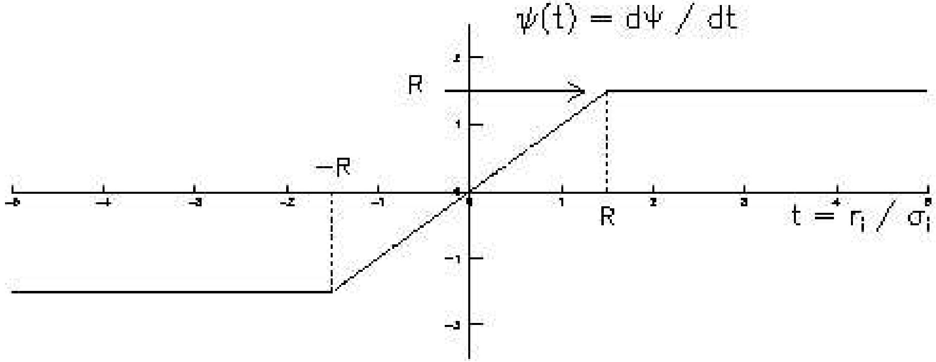
$$w_i = \frac{\psi(r_i/\sigma_i)}{r_i/\sigma_i} = \begin{cases} 1, & |r_i| \leq R\sigma_i \\ R\sigma_i/|r_i|, & |r_i| > R\sigma_i. \end{cases} \tag{40}$$

Efficient robustification against outliers and background

- The M-estimator is formally identical to a least-squares estimator with modified weights w_i/σ_i^2, thus downweighting the contribution of large residuals to the objective function $\mathcal{M}$.
- The estimate cannot be computed explicitely, as the weights depend on the unknown residuals r_i. Therefore one has to resort to an *iterative procedure*:
  - The initial estimate is computed by ordinary least-squares, corresponding to $w_i = 1$ for all i.
  - The estimate of iteration $k - 1$ is used to compute residuals which are then used to compute the weights for iteration k.
  - The iteration is stopped as soon as there is no significant change in the minimum value of the objective function $\mathcal{M}$.

Efficient robustification against outliers and background





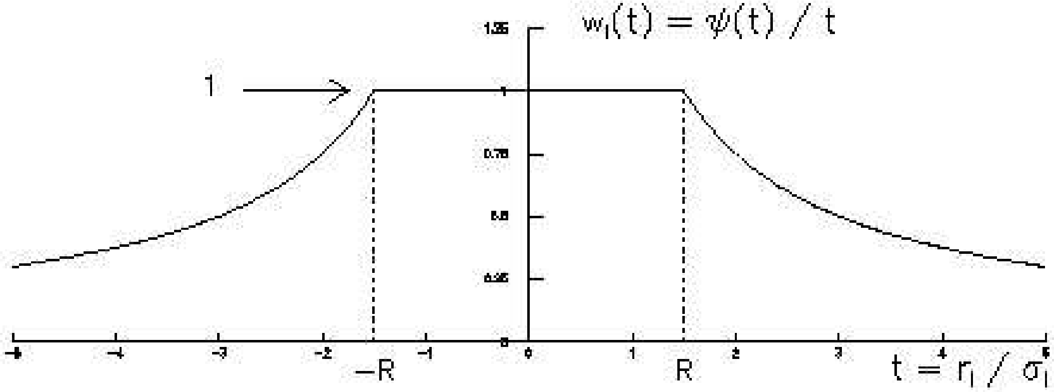


Figure 9: M-estimator functions Ψ, ψ and extra weight $w_i(r_i/\sigma_i)$ for $R = 1.5$

5.2.3 Outlook: Deterministic annealing filter (DAF)

- Tracks which cannot be regarded as belonging to a certain vertex have a large influence on the estimated vertex position. What is therefore needed is a *robust filter which reduces the influence of wrongly associated tracks* on the estimated vertex position and which gives some quantitative indication of the incompatibility between them.
- For a primary vertex the situation is less critical as the *interaction profile robustifies the vertex position.* But for the distinction of the vertices of relatively short lived particles such as charm- and bottom quark mesons a good robustification is essential to avoid large combinatorial overhead.
- the main request is to find a *reweighting procedure* which needs only a small number of iterations before making final decisions.
- One way of checking a set of tracks for the quality of association to a vertex was the 'smoothed' χ^2 (see equ(36)); the correlation with other tracks is there but is relatively small. The problem of this test is that it loses its power if there are several outliers.

Efficient robustification against outliers and background

- The M-estimator is a valid way out to stabilise a single vertex or a high-multiplicity primary vertex against outliers, but for secondary vertices there are two drawbacks:
  - The downweighting is not effective enough
  - There is no assignment probability for a multi-vertex assignment strategy
- A way out is to calculate an 'association probability' as a function of the smoothed χ^2,a 'cutoff' parameter χ^2_{cutoff} and a so-called 'temperature' T; χ^2_{cutoff} and T are tunable. The idea follows an *analogy to thermodynamics*, and in-depth studies for the CMS-detector are under way (R. Frühwirth, Hephy, Vienna, private communication):

$$p_k = \frac{e^{-\frac{\chi^2_k}{2T}}}{e^{-\frac{\chi^2_{cutoff}}{2T}} + e^{-\frac{\chi^2_k}{2T}}} \tag{41}$$

$$V_k^{-1} \leftarrow p_k V_k^{-1}$$

- This weight function has the following properties:
  - $\chi^2_{cutoff} \rightarrow \infty$ means that all $p_k \equiv 1$, and the fit degenerates to the standard LSM procedure

- $T \to 0$ means a fixed cut volume χ^2_{cutoff} (step function) which results in a 'hard' assignment, namely 0 or 1.

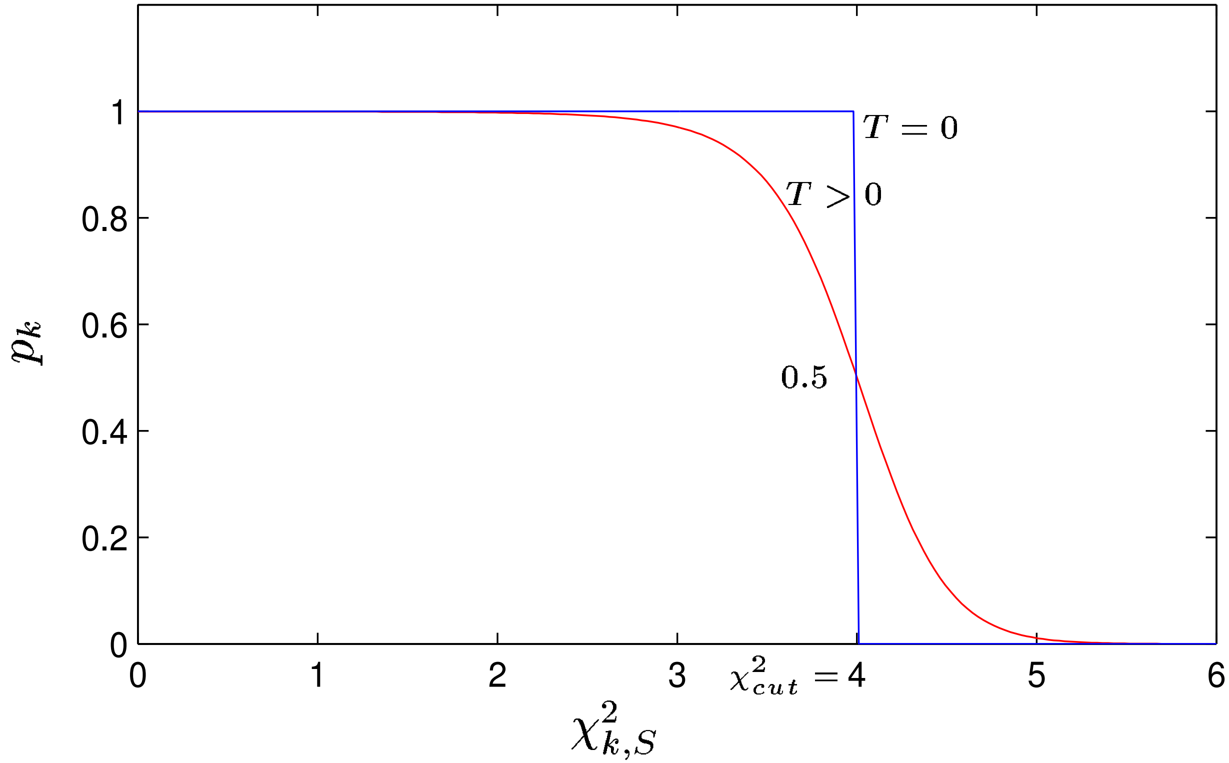


Figure 10: Association probability depending on the $\chi^2_{k,S}$ of a certain track. Here we see the graphs for $T = 0$ and $T \neq 0$.

Efficient robustification against outliers and background

- A possible *strategy for a multi vertex fitter* would consist of several steps:
  - in a first step a classical approach for the experiment of exclusive and exhaustive subsets ('clusters') is performed
  - then a Kalman filter is applied to each subset
  - obvious mis-associations are sorted out by a combinatorial approach, and the content of the subsets will be refined.
- In the next step, the overall set is used again. The essential point is now that according to the number of presumed vertices ($i = 1 \dots n_v$) n_v *overall fits are performed*. The assignment probability (or *weight function) with two indices,*

$$w_{ik} = \frac{e^{-\frac{\chi^2_{ik}}{2T}}}{e^{-\frac{\chi^2_{S,cutoff}}{2T}} + \sum_{k=1}^{n_v} \chi^2_{ik}} \tag{42}$$

where i is the index of the respective vertex. The tracks coming from different vertices are now drastically down-weighted.

Efficient robustification against outliers and background

- A number of iterations is performed. During the iterations, in addition to the cut value $\chi^2_{S,cutoff}$ already chosen, the temperature can also be adjusted. However more experience has to be gained, and the tuning will certainly be very different for different vertex event topologies, as expected e.g. at B factories or for multi-event patterns as expected at LHC.

Further reading

- Regler, M. and Frühwirth, R. (1990): Reconstruction of Charged Tracks. In: *Concepts and Techniques in High Energy Physikcs V*, Plenum Publishing Cooperation, New York, Ed. T. Ferbel
- Regler, M., Frühwirth, R. and Mitaroff, W. (1996): *Filter Methods in Track and Vertex Reconstruction.* International Journal of Modern Physics C*4*, 521
- Avery, P. (1998): *Vertexing and Kinematic Fitting*, Lectures Given at SLAC (http://www.phys.ufl.edu/~avery)
- Frühwirth, R. et al (2000): *Data Analysis Techniques for High Energy Physics* (second edition), Cambridge University Press, Ed. M. Regler and R. Frühwirth
- Waltenberger, W. et al (2003): Talk given at the Aste Intern. Workshop on Advanced Computing and Analysis Techniques in Physics Research, Moscow.
  http://cmsdoc.cern.ch/cms/Physics/btau/management/activities/reconstruction/vertex/acat02.d/Robust.pdf
- Vanlaer P., D'Hondt M., Frühwirt R. and Waltenberger W. (2003): *Sensitivity of Robust Vertex Fitting Algorithms* in preparation (fru@hephy.oeaw.ac.at)