Vertex reconstruction with tracks described by a Gaussian mixture

R. Frühwirth, W. Mitaroff and **W. Waltenberger**

(presented by **W. Mitaroff**)

Austrian Academy of Sciences, Inst. of High Energy Physics
A-1050 Vienna, Austria, Europe

Reconstructed electron tracks, with proper treatment of the energy loss by bremsstrahlung, can be described by a mixture of Gaussian state vectors. These are the virtual measurements of a subsequent vertex reconstruction, which has to be performed by a Gaussian sum filter (GSF).

Algorithms which have been developed for the CMS experiment exist, and are to be implemented in our detector-independent vertex reconstruction toolkit RAVE. They represent an extension of the Kalman filter and smoother, and can be combined with the robust adaptive vertex filter.

An inherent problem of the GSF is the exponential growth of its number of Gaussian components. This can be limited by collapsing "similar" pairs, as defined by an appropriate "distance measurement".

Results presented are based on CMS simulations.

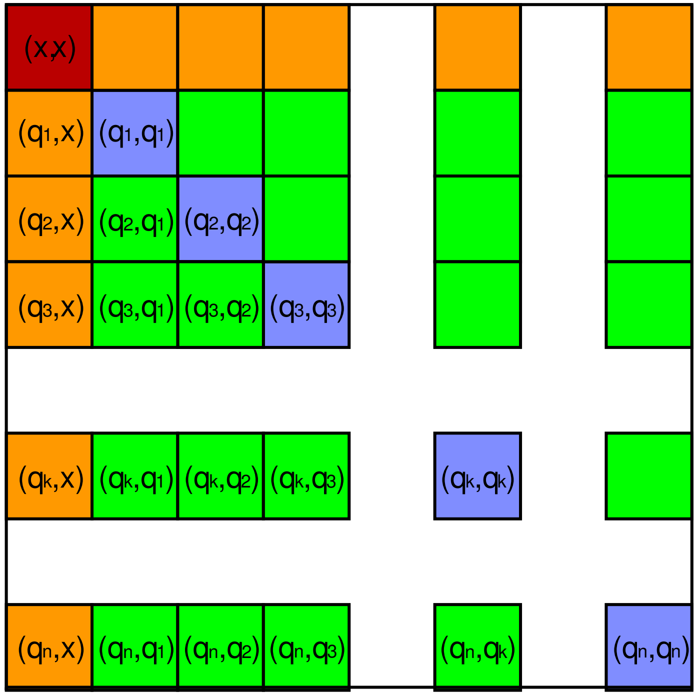
Reminder: vertex fit by a Kalman filter

Virtual measurements (Gaussian errors):

- Reconstructed tracks: $\mathbf{p}_k$, $\mathbf{V_k}$ $[k = 1 \ldots n]$
  - $\mathbf{p}_k$ = 5-vectors of fitted track parameters,
  - $\mathbf{V_k} \equiv \mathrm{cov}(\mathbf{p}_k)$ = symmetric 5×5 matrices.
- Beam interaction profile (optional): $\mathbf{v}$, $\mathbf{V_o}$
  - $\mathbf{v}$ = 3-vector of centre of the beam int. profile,
  - $\mathbf{V_o} \equiv \mathrm{cov}(\mathbf{v})$ = symm. or diag. 3×3 matrix.

Fit results described by Gaussian errors:

- Reconstructed vertex position: $\mathbf{x}$, $\mathbf{C}$
  - $\mathbf{x}$ = 3-vector of fitted space coordinates,
  - $\mathbf{C} \equiv \mathrm{cov}(\mathbf{x})$ = symmetric 3×3 matrix.
- Re-fitted tracks at vertex: $\mathbf{q}_k$, $\mathbf{D}_k$ $[k = 1 \ldots n]$
  - $\mathbf{q}_k$ = 3-vectors of smoothed track parameters,
  - $\mathbf{D}_k \equiv \mathrm{cov}(\mathbf{q}_k)$ = symmetric 3×3 matrices.



- The full information (e.g. needed by a subsequent kinematics fit) includes also these covariances:
  - $\mathbf{E}_k^{\mathrm{T}} \equiv \mathrm{cov}(\mathbf{q}_k, \mathbf{x})$ $[k = 1 \ldots n]$ — n asymmetric 3×3 matrices,
  - $\mathrm{cov}(\mathbf{q}_k, \mathbf{q}_\ell)$ $[k, \ell = 1 \ldots n$ with $k \neq \ell]$ — $n \cdot (n-1)$ asymmetric 3×3 matrices.

In case of **non-Gaussian errors**, the $\mathbf{V}_k$ do not fully describe the measurements. The KF is still the optimal linear estimator, and its results are equivalent to Gaussian w.r.t. "mean squared quantities" (Gauss-Markov theorem). Improvement can only be achieved by a non-linear estimator, like the GSF.

Reference: *R. Frühwirth, P. Kubinec, W. Mitaroff, M. Regler*: Computer Physics Comm. **96** (1996) 189.

Tracks described by a Gaussian mixture

- Energy loss of electrons and positrons is dominated by bremsstrahlung. It is a stochastic process which can be modeled by the Bethe-Heitler formula.
- A track $\mathbf{p}_k$ reconstructed with proper treatment of bremsstrahlung is described by a mixture of M_k Gaussian measurement vectors $\mathbf{p}_k^i$: its p.d.f. is

$$\wp(\mathbf{p}_k) = \sum_{i=1}^{M_k} \gamma_k^i \cdot \Gamma(\mathbf{p}_k; \mathbf{p}_k^i, \mathbf{V}_k^i), \qquad \sum_{i=1}^{M_k} \gamma_k^i = 1$$

  with $\Gamma(\mathbf{p}_k; \ldots)$ being a multivariate Gaussian p.d.f. of mean $\mathbf{p}_k^i$ and covariance matrix $\mathrm{cov}(\mathbf{p}_k^i, \mathbf{p}_k^i) \equiv \mathbf{V}_k^i$. In general the means need not to be equal.

- Each component $i = 1 \ldots M_k$ of the mixture corresponds to one hypothesis on the virtual measurement, with the weight γ_k^i being its probability.
- In practice, a number of components $M_k \leq 6$ is sufficient.

References: *R. Frühwirth*: Computer Physics Comm. **154** (2003) 131.
W. Adam, R. Frühwirth, A. Strandlie, T. Todorov: CMS note 2005/001, CERN.

Vertex fit by a Gaussian sum filter (1)

- The **measurement equation** maps the state vector to the measurement vector. It is linearized by a 1^{st} order approximation at some "expansion point" $\mathbf{e}_k = (\mathbf{x}_e, \mathbf{q}_k^e)$, and homogenized by transforming away the constant term:

$$\mathbf{p}_k^i = \mathbf{h}_k(\mathbf{x}, \mathbf{q}_k) + \epsilon_k^i, \qquad \mathrm{cov}(\epsilon_k^i) \equiv \mathbf{V}_k^i = (\mathbf{G}_k^i)^{-1}$$

$$\mathbf{h}_k(\mathbf{x}, \mathbf{q}_k) \approx \mathbf{A}_k\mathbf{x} + \mathbf{B}_k\mathbf{q}_k, \qquad \mathbf{A}_k \equiv [\partial\mathbf{h}_k/\partial\mathbf{x}]_{\mathbf{e}_k}, \quad \mathbf{B}_k \equiv [\partial\mathbf{h}_k/\partial\mathbf{q}_k]_{\mathbf{e}_k}$$

- P.d.f. of measurement hypothesis $\mathbf{p}_k^i$, conditional on a state vector $(\mathbf{x}, \mathbf{q}_k)$:

$$\wp(\mathbf{p}_k^i \mid \mathbf{x}, \mathbf{q}_k) = \Gamma(\mathbf{p}_k^i; \mathbf{h}_k(\mathbf{x}, \mathbf{q}_k), \mathbf{V}_k^i)$$

- The estimated state vector $\mathbf{x}_{k-1}$ of the vertex position, based on $(k-1)$ tracks $\{\mathbf{p}_1, \ldots \mathbf{p}_{k-1}\}$, is assumed as mixture of N_{k-1} Gaussian $\mathbf{x}_{k-1}^j$, with p.d.f.

$$\wp(\mathbf{x}_{k-1}) = \sum_{j=1}^{N_{k-1}} \pi_{k-1}^j \cdot \Gamma(\mathbf{x}_{k-1}; \mathbf{x}_{k-1}^j, \mathbf{C}_{k-1}^j), \qquad \sum_{j=1}^{N_{k-1}} \pi_{k-1}^j = 1$$

- Dummy state vector of momentum at $\mathbf{x}_{k-1}$: $\wp(\mathbf{q}_k^o) = \Gamma(\mathbf{q}_k^o; \mathbf{q}_k^e, \mathbf{D}_k^o \to \infty)$

Vertex fit by a Gaussian sum filter (2)

- First, in a **partial GSF step**, hypothesis i of the next track $= \mathbf{p}_k^i$ is added, yielding an updated estimate of the full state vector $= (\mathbf{x}_k^i, \mathbf{q}_k^i)$.
- With $\wp(\mathbf{x}_{k-1})$ and $\wp(\mathbf{q}_k^o)$ as prior p.d.f.s, the posterior p.d.f. of this new state vector, conditional on the observation $\mathbf{p}_k^i$, follows from Bayes' theorem as

  $$\wp(\mathbf{x}_k^i, \mathbf{q}_k^i \mid \mathbf{p}_k^i) \propto \wp(\mathbf{p}_k^i \mid \mathbf{x}_k^i, \mathbf{q}_k^i) \cdot \wp(\mathbf{x}_{k-1}) \cdot \wp(\mathbf{q}_k^o),$$
  resulting in
  $$\wp(\mathbf{x}_k^i, \mathbf{q}_k^i \mid \mathbf{p}_k^i) = \sum_{j=1}^{N_{k-1}} \omega_k^{ij} \cdot \Gamma\left((\mathbf{x}_k^i, \mathbf{q}_k^i); (\mathbf{x}_k^{ij}, \mathbf{q}_k^{ij}), \mathbf{X}_k^{ij}\right), \qquad \sum_{j=1}^{N_{k-1}} \omega_k^{ij} = 1$$

  with means $(\mathbf{x}_k^{ij}, \mathbf{q}_k^{ij})$ and 6×6 covariance matrix $\mathbf{X}_k^{ij}$ obtained from a **Kalman filter** for component j being updated by component i. Posterior weights ω_k^{ij} correspond to prior weights π_{k-1}^j, conditional on the measurement hypothesis $\mathbf{p}_k^i$. Explicit formulae are given on the next slide.
- The **complete GSF step** is a weighted sum over all hypotheses:

  $$\wp(\mathbf{x}_k, \mathbf{q}_k \mid \mathbf{p}_k) = \sum_{i=1}^{M_k} \sum_{j=1}^{N_{k-1}} \gamma_k^i \, \omega_k^{ij} \cdot \Gamma\left((\mathbf{x}_k, \mathbf{q}_k); (\mathbf{x}_k^{ij}, \mathbf{q}_k^{ij}), \mathbf{X}_k^{ij}\right)$$

Formulae for one Kalman filter step

$$\mathbf{x}_k^{ij} = \mathbf{C}_k^{ij} \left((\mathbf{C}_{k-1}^{j})^{-1} \, \mathbf{x}_{k-1}^{j} + \mathbf{A}_k^{\mathrm{T}} \, \mathbf{H}_k^{i} \, \mathbf{p}_k^{i} \right)$$

$$\mathbf{q}_k^{ij} = \mathbf{W}_k^{i} \, \mathbf{B}_k^{\mathrm{T}} \, \mathbf{G}_k^{i} \left(\mathbf{p}_k^{i} - \mathbf{A}_k \, \mathbf{x}_k^{ij} \right)$$

$$\mathbf{X}_k^{ij} \equiv \begin{pmatrix} \mathbf{C}_k^{ij} & \mathbf{E}_k^{ij} \\ \mathbf{E}_k^{ij\mathrm{T}} & \mathbf{D}_k^{ij} \end{pmatrix}$$

$$\mathbf{C}_k^{ij} \equiv \mathrm{cov}(\mathbf{x}_k^{ij}) = \left((\mathbf{C}_{k-1}^{j})^{-1} + \mathbf{A}_k^{\mathrm{T}} \, \mathbf{H}_k^{i} \, \mathbf{A}_k \right)^{-1}$$

$$\mathbf{D}_k^{ij} \equiv \mathrm{cov}(\mathbf{q}_k^{ij}) = \mathbf{W}_k^{i} + \mathbf{E}_k^{ij\mathrm{T}} \, (\mathbf{C}_k^{ij})^{-1} \, \mathbf{E}_k^{ij} \qquad \mathbf{W}_k^{i} = \left(\mathbf{B}_k^{\mathrm{T}} \, \mathbf{G}_k^{i} \, \mathbf{B}_k \right)^{-1}$$

$$\mathbf{E}_k^{ij} \equiv \mathrm{cov}(\mathbf{x}_k^{ij}, \mathbf{q}_k^{ij}) = -\, \mathbf{C}_k^{ij} \, \mathbf{A}_k^{\mathrm{T}} \, \mathbf{G}_k^{i} \, \mathbf{B}_k \, \mathbf{W}_k^{i} \qquad \mathbf{H}_k^{i} = \mathbf{G}_k^{i} - \mathbf{G}_k^{i} \, \mathbf{B}_k \, \mathbf{W}_k^{i} \, \mathbf{B}_k^{\mathrm{T}} \, \mathbf{G}_k^{i}$$

$$(\chi^2_{\mathrm{F}})_k^{ij} = \left(\mathbf{x}_k^{ij} - \mathbf{x}_{k-1}^{j} \right)^{\mathrm{T}} (\mathbf{C}_{k-1}^{j})^{-1} \left(\mathbf{x}_k^{ij} - \mathbf{x}_{k-1}^{j} \right) + \mathbf{r}_k^{ij\mathrm{T}} \, \mathbf{G}_k^{i} \, \mathbf{r}_k^{ij}$$

$$(\chi^2_{\mathrm{F}})_k = \sum_{i=1}^{M_k} \sum_{j=1}^{N_{k-1}} \gamma_k^{i} \, \omega_k^{ij} \cdot (\chi^2_{\mathrm{F}})_k^{ij} \quad [\text{“pseudo”}] \qquad \mathbf{r}_k^{ij} = \mathbf{p}_k^{i} - \left(\mathbf{A}_k \, \mathbf{x}_k^{ij} + \mathbf{B}_k \, \mathbf{q}_k^{ij} \right)$$

Calculation of the posterior weights (prior quantities have a bar):

$$\bar{\mathbf{r}}_k^{ij} = \mathbf{p}_k^{i} - \left(\mathbf{A}_k \, \mathbf{x}_{k-1}^{j} + \mathbf{B}_k \, \mathbf{q}_k^{e} \right), \qquad \bar{\mathbf{R}}_k^{ij} \equiv \mathrm{cov}(\bar{\mathbf{r}}_k^{ij}) = \mathbf{V}_k^{i} + \mathbf{A}_k \, \mathbf{C}_{k-1}^{j} \, \mathbf{A}_k^{\mathrm{T}} + \mathbf{B}_k \, \mathbf{D}_k^{0} \, \mathbf{B}_k^{\mathrm{T}}$$

$$\omega_k^{ij} \propto \pi_{k-1}^{j} \cdot \left| \bar{\mathbf{R}}_k^{ij} \right|^{-1/2} \cdot \mathrm{e}^{-(\bar{\chi}^2)_k^{ij}/2} \Leftarrow \Sigma_{j=1}^{N_{k-1}} \omega_k^{ij} = 1, \qquad (\bar{\chi}^2)_k^{ij} = \bar{\mathbf{r}}_k^{ij\mathrm{T}} \, (\bar{\mathbf{R}}_k^{ij})^{-1} \, \bar{\mathbf{r}}_k^{ij}$$

Note: choose $\mathbf{D}_k^{0}$ big enough to avoid a bias on ω_k^{ij}, but not too big in order to avoid it becoming singular.

Vertex fit by a Gaussian sum filter (3)

- For the **next GSF step** (adding track $\mathbf{p}_{k+1}$), the marginal p.d.f of the posterior $\wp(\mathbf{x}_k, \mathbf{q}_k \mid \mathbf{p}_k)$ will become the prior p.d.f. of the vertex position $\mathbf{x}_k$:

$$\text{with } (ij) \to j', \qquad \pi_k^{j'} = \gamma_k^i \, \omega_k^{ij}, \qquad N_k = M_k \cdot N_{k-1}$$

$$\wp(\mathbf{x}_k) = \sum_{j'=1}^{N_k} \pi_k^{j'} \cdot \Gamma(\mathbf{x}_k; \mathbf{x}_k^{j'}, \mathbf{C}_k^{j'}), \qquad \sum_{j'=1}^{N_k} \pi_k^{j'} = 1$$

- This p.d.f. is a mixture of N_k Gaussian components. Their number increases exponentially with the number of tracks added. A way to keep it limited is to identify a "similar" pair of components, and to collapse it into one:

$$\pi_k^1 \cdot \Gamma(\mathbf{x}_k; \mathbf{x}_k^1, \mathbf{C}_k^1) + \pi_k^2 \cdot \Gamma(\mathbf{x}_k; \mathbf{x}_k^2, \mathbf{C}_k^2) =: (\pi_k^1 + \pi_k^2) \cdot \Gamma(\mathbf{x}_k; \mathbf{x}_k^c, \mathbf{C}_k^c)$$

$$\mathbf{x}_k^c = p^1 \cdot \mathbf{x}_k^1 + p^2 \cdot \mathbf{x}_k^2, \qquad \text{where } p^1 \equiv \pi_k^1/(\pi_k^1 + \pi_k^2), \;\; p^2 \equiv \pi_k^2/(\pi_k^1 + \pi_k^2)$$

$$\mathbf{C}_k^c = p^1 \cdot (\mathbf{C}_k^1 + \mathbf{x}_k^1 \mathbf{x}_k^{1\mathrm{T}}) + p^2 \cdot (\mathbf{C}_k^2 + \mathbf{x}_k^2 \mathbf{x}_k^{2\mathrm{T}}) - \mathbf{x}_k^c \mathbf{x}_k^{c\mathrm{T}}$$

- The similarity between two Gaussian p.d.f.s may be measured by the **Kullback-Leibler** or by the **Mahalanobis** distance (see the reference).

Is there a Gaussian sum smoother ?

- After iteration over all tracks $\ell = 1 \ldots n$, the vertex position's final estimate $\wp(\mathbf{x}_n)$ contains the complete information from all tracks.

  The estimate of each track k at the vertex (marginal p.d.f. of the posterior) $\wp(\mathbf{q}_k)$ contains information from tracks $\ell = 1 \ldots k$ only.

- A straight-forward GSF smoother would update each component $i = 1 \ldots M_k$ of track k with those N_n/M_k components j^* of the final vertex to which it had contributed, using the Kalman formulae with $(\mathbf{x}_n^{ij^*}, \mathbf{C}_n^{ij^*})$. However, the $k.i$ to j relationship has been destroyed by the collapses.

- Because only the last track fitted by the GSF does contain the full information, a "forward & backward GSF" can perform the smoothing:

  - GSF in ascending $(\ell = 1 \ldots k-1)$ and descending $(\ell' = n \ldots k+1)$ order;
  - Calculate the weighted mean $\wp_{mean}(\mathbf{x}_{k\notin})$ of $\wp_{asc}(\mathbf{x}_{k-1})$ and $\wp_{desc}(\mathbf{x}_{k+1})$;
  - Last GSF step of adding $\mathbf{p}_k$ to prior $\wp_{mean}(\mathbf{x}_{k\notin})$ yields posterior $\wp(\mathbf{x}_n, \mathbf{q}_k^*)$.

- This smoothed estimate $\wp(\mathbf{q}_k^*)$ contains the information from all tracks.

Simulated performance of the GSF

Simulated 2-component Gaussian mixture: 90 % "narrow" ($\sigma_{IP} = 100\mu$m) and 10 % "wide" ($\sigma_{IP} = 1000\mu$m). 4-track vertex fits by the Kalman filter (Fig. 1), and by the GSF without collapsing (Fig. 2).

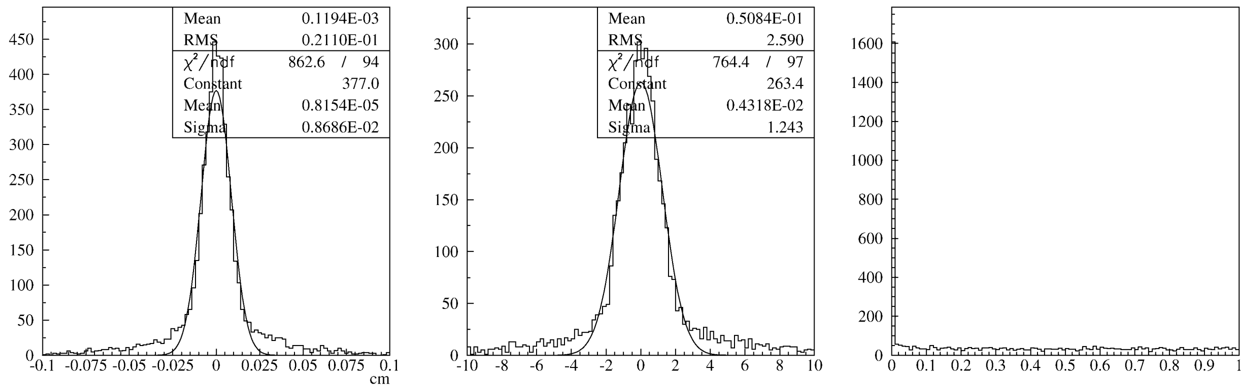


Fig. 1. Residual (left) and pull (middle) of the x-coordinate of the reconstructed vertex and χ^2 probability (right) of the vertex fit using the Kalman filter.

Note that 32% of the Kalman vertex fits have χ^2 probability $< 1\%$, albeit all tracks entered are "good" ones.

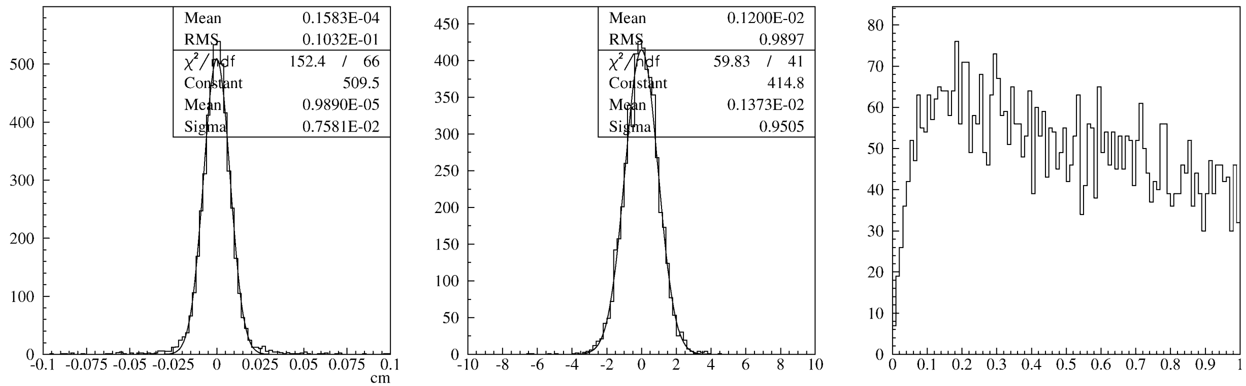


Fig. 2. Residual (left) and pull (middle) of the x-coordinate of the reconstructed vertex and χ^2 probability (right) of the vertex fit using the GSF, without limiting the number of components.

The dip at 0 of the pseudo χ^2 probability is due to the tails of the "narrow" component being well within the range of the "wide" core, and thus are misinterpreted as belonging to the latter.

Conclusions and outlook

- The Gaussian sum filter (GSF) has been implemented and tested in the CMSSW framework of the CMS experiment at LHC, both for the reconstruction of electron and positron tracks, and for the reconstruction of vertices involving tracks described by a Gaussian mixture.

- It has been shown elsewhere that vertex reconstruction by the GSF can successfully be combined with the adaptive vertex fitter (AVF), thus further improving its robustness w.r.t. outlier tracks.

- We plan to implement the GSF in the Vienna fast simulation tool "LiC Detector Toy" (LDT) for the reconstruction of electron and positron tracks, simulated with energy loss by bremsstrahlung.

- We plan to implement the GSF in the Vienna vertex reconstruction toolkit RAVE, and to combine it with the existing AVF.

- Caveat: realisation is subject to the manpower available.

Reference: *T. Speer, R. Frühwirth*: Computer Physics Comm. **174** (2006) 935.