A Gaussian-sum Filter for vertex reconstruction

R. Frühwirth - HEPHY Vienna
T. Speer - University of Zurich

IX International Workshop on Advanced Computing and Analysis Techniques in Physics Research
Tsukuba
4th December 2003
Vertex reconstruction

- Standard tool for vertex reconstruction is the Kalman Filter (also implemented in the reconstruction software of the CMS experiment at LHC, CERN)
- The Kalman Filter is mathematically equivalent to a global least square minimization (LSM)
- If the model is linear and random noise is Gaussian:
  - LS estimators are unbiased and have minimum variance
  - Residuals and pulls of estimated quantities are also Gaussian
- For non-linear models or non-Gaussian noise, it is still the optimal linear estimator
- Non-Gaussian measurement errors degrade results!
The Gaussian-sum Filter

- **Gaussian-sum Filter (GSF)**
  - Measurement error distributions modelled by **mixture of Gaussians**:
    - Main component of the mixture would describe the core of the distribution
    - Tails would be described by one or several additional Gaussians.
  - First proposed by R. Frühwirth for track reconstruction
    (Computer Physics Communications 100 (1997) 1.)
  - Successfully implemented in the CMS reconstruction software for **electron track reconstruction**:
    - Bethe-Heitler energy loss distribution modeled by a mixture of Gaussians
  - GSF for vertex reconstruction now also implemented in the CMS reconstruction software.
The Gaussian-sum Filter for vertex reconstruction

- Track parameter error distributions modeled by a mixture of Gaussians
- Vertex State vector $x$, is also distributed according to a mixture of Gaussians
- Iterative procedure: estimate of the vertex is updated with one track at the time
- Add new track to vertex, each component of the Vertex State is updated with each component of the track (Combinatorial combination of all track components)
- The new Vertex State $x_k$ is therefore distributed according to a mixture of $N_k$
  
  \[ ( = N_{\text{track} - k} \ast N_{\text{vertex} - k - 1}) \text{ Gaussians} \]
- The filter is a weighted sum of several Kalman Filters
  - GSF is implemented as a number of Kalman filters run in parallel
  - The weights of the components are calculated separately
- Non-linear estimator: weights depend on the measurements
Simulation

Simplified simulation in a fully controlled environment:

- Tracks generated at a common vertex
- No track reconstruction
- Track parameters are smeared according to known distributions:
  - E.g. 2 component Gaussian mixture:
    - Narrow component: 90 % Relative weight
      (Standard deviation of Impact parameter = 100µm)
    - Wide component: 10 % Relative weight
      Std dev. 10x larger (Impact parameter = 1000µm)
      \( \Rightarrow \) Ratios of Standard deviation = 10

- For the Kalman Filter:
  - tracks smeared according to two-component mixture
  - single component used in the fit:
    \( \Rightarrow \) track parameter variance of dominating component
    \( \Rightarrow \) estimated position independent of scaling of variance (but not position uncertainty or \( \chi^2 \))
Kalman Filter fit

Four track-vertex fit with the Kalman Filter:

- Non-Gaussian tails in the distributions of residuals and pulls
- Large number of fits with $P(\chi^2) < 0.01$

<table>
<thead>
<tr>
<th>Mean</th>
<th>0.1194E-03</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMS</td>
<td>0.2110E-01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mean</th>
<th>0.5084E-01</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMS</td>
<td>2.590</td>
</tr>
</tbody>
</table>

$\chi^2 / \text{NDF} = 862.6 / 94$
Mean = 377.0
Sigma = 0.8686E-02

$\chi^2 / \text{NDF} = 764.4 / 97$
Mean = 263.4
Sigma = 0.4318E-02
Gaussian-sum Filter fit

Four track-vertex fit with the GSF (using the full Gaussian mixture)

Residuals: smaller tails than with the Kalman Filter, smaller resolution.
The remaining tails are due to events with several outliers.
No outliers in the pull distributions: error on the outliers correctly taken into account.

\( P(\chi^2) \): dip at 0. - in early stages of the fit, bias towards components with a low \( \chi^2 \)
The filters need several iterations (tracks) to stabilise and select the correct vertex component (combination of track components)
Measures of improvement of vertex fits

- Two-component Gaussian mixtures with different ratios of standard deviations and relative weights (4-track vertices)

- Measures:
  - 50% and 90% coverage: half-widths of the symmetric intervals covering 50% and 90% of the residual distribution (x-coordinate)
  - Relative efficiency: ratio of the mean (3D) distances of the estimated vertex from its simulated position, for fits with the Kalman Filter and the GSF
    - For Kalman Filter: estimated position independent of scaling of track parameter variance
    - Fraction of Kalman Filter fits with $P(\chi^2) < 0.01$
      - For Kalman Filter: estimated uncertainty dependent of scaling of track parameter variance

\[
\begin{array}{l|l|l}
\hline
\text{Parameter} & \text{Value} & \text{Error} \\
\hline
\text{Constant} & 137.2 \pm 87 & 4.857 \\
\text{Mean} & 0.276 \pm 0.1086 \times 10^{-3} & 0.1086 \times 10^{-3} \\
\text{Sigma} & 0.753 \pm 0.9231 \times 10^{-4} & 0.9231 \times 10^{-4} \\
\hline
\end{array}
\]
Coverage

Ratio of the standard deviations: 2

50% coverage

Ratio of the standard deviations: 4

90% coverage

Ratio of the standard deviations: 10

Multiplicative factor: 10

Coverage ratios for different standard deviation ratios and multiplicative factors.
Relative efficiency: ratio of the mean distances (in three dimensions) of the estimated vertex from its simulated position, for fits with the KVF and the GSF

- Highest relative efficiency: largest distance between the two-component Gaussian mixture and the single Gaussian
- Larger weight of the tails: tails start to dominate \(\Rightarrow\) lower relative efficiency
Relative efficiency

Kullback-Leibler Distance between a two-component Gaussian mixture and single-Gaussian distribution with identical moments:

\[ D_{KL}(p_1, p_2) = 2 \left( \int_{-\infty}^{\infty} \ln \left( \frac{p_1}{p_2} \right) p_1 \, dx + \int_{-\infty}^{\infty} \ln \left( \frac{p_2}{p_1} \right) p_2 \, dx \right) \]

- \( p \): relative weight of the second Gaussian
- \( f \): ratio of their standard deviations
Fraction of Kalman Filter fits with $P(\chi^2) < 0.01$

- Estimated uncertainty dependent of scaling of track parameter variance
Component limitation

The number of components increases exponentially:

- $n$ measurements, with $m$ components: $n^m$ components at the end!
  - Combinatorial explosion!

- Keep only $M$ components at each step:
  - Keep components with the largest weight, discard the rest
  - Cluster (collapse) components with the smallest 'distance'

2 Distance measurements were used:

- Kullback-Leibler Distance
  $$D_{KL}(p_1, p_2) = \text{tr} \left[ \left(V_1 - V_2\right) \left(V_1^{-1} - V_2^{-1}\right) \right] + \left(\mu_1 - \mu_2\right)^T \left(V_1^{-1} + V_2^{-1}\right) \left(\mu_1 - \mu_2\right)$$

- Mahalanobis Distance
  $$D_M(p_1, p_2) = \left(\mu_1 - \mu_2\right)^T \left(V_1 + V_2\right)^{-1} \left(\mu_1 - \mu_2\right)$$

The GSF vertex filter shows little sensitivity to the number of components kept
Component limitation

2 component Gaussians mixture:
- Narrow comp.: 80% rel. weight
- Wide comp.: 20% rel. weight
- Ratios of Standard deviation = 10

With 4 tracks: up to 16 components

Pulls when a single component is used (Kalman filter)
**Component limitation**

GSF - No limitation of the number of components

\[ P(\chi^2) \]

**Kalman Filter**

\[ \text{Residuals (\mu m)} \]

Mean  0.2073E-03
RMS  0.1576E-01
\[ \chi^2 \]  400.9 / 91
Constant  429.3
Mean  0.5739E-04
Sigma  0.8488E-02

\[ x \text{ Pull} \]

Mean  0.1438E-01
RMS  1.005
\[ \chi^2 \]  83.72 / 52
Constant  435.7
Mean  0.4765E-02
Sigma  0.9001

\[ x \text{ Residuals (\mu m)} \]

Mean  0.2094E-03
RMS  0.2781E-01
\[ \chi^2 \]  1258. / 97
Constant  194.2
Mean  -0.7570E-04
Sigma  0.1505E-01

\[ x \text{ Pull} \]

Mean  0.4029E-01
RMS  3.305
\[ \chi^2 \]  1111. / 97
Constant  148.7
Mean  -0.1308E-01
Sigma  1.929
Component limitation

GSF - Limit of 2 components (using Kullback-Leibler Distance)

GSF - Limit of 4 components (using Kullback-Leibler Distance)
## Component limitation

<table>
<thead>
<tr>
<th>Component Limit</th>
<th>Relative Efficiency 50% coverage</th>
<th>Relative Efficiency 90% coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>No limitation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.91</td>
<td>0.93</td>
</tr>
<tr>
<td>3</td>
<td>0.94</td>
<td>0.95</td>
</tr>
<tr>
<td>4</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>6</td>
<td>0.96</td>
<td>0.97</td>
</tr>
<tr>
<td>8</td>
<td>0.96</td>
<td>0.97</td>
</tr>
</tbody>
</table>

### Table

<table>
<thead>
<tr>
<th>Nbr Comp.</th>
<th>Average $\chi^2$</th>
<th>Res. [µm]</th>
<th>Pull</th>
<th>Average $\chi^2$</th>
<th>Res. [µm]</th>
<th>Pull</th>
</tr>
</thead>
<tbody>
<tr>
<td>No limitation</td>
<td>0.99</td>
<td>84.8</td>
<td>0.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kullback-Leibler Distance</td>
<td></td>
<td></td>
<td></td>
<td>Mahalanobis Distance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.91</td>
<td>90.5</td>
<td>0.88</td>
<td>0.93</td>
<td>92.2</td>
<td>0.84</td>
</tr>
<tr>
<td>3</td>
<td>0.94</td>
<td>90.5</td>
<td>0.88</td>
<td>0.95</td>
<td>89.7</td>
<td>0.85</td>
</tr>
<tr>
<td>4</td>
<td>0.95</td>
<td>85.4</td>
<td>0.89</td>
<td>0.95</td>
<td>84.9</td>
<td>0.89</td>
</tr>
<tr>
<td>6</td>
<td>0.96</td>
<td>84.9</td>
<td>0.89</td>
<td>0.97</td>
<td>84.6</td>
<td>0.89</td>
</tr>
<tr>
<td>8</td>
<td>0.96</td>
<td>83.9</td>
<td>0.9</td>
<td>0.97</td>
<td>83.9</td>
<td>0.9</td>
</tr>
</tbody>
</table>

### Diagrams

- **Relative efficiency**
- **50% coverage**
- **90% coverage**
Component limitation

4 component Gaussians mixture:

- 1\textsuperscript{st} (narrow) comp.: 50\% rel. weight ($\sigma_1$)
- 2\textsuperscript{nd} comp.: 30\% rel. weight ($\sigma_2=5*\sigma_1$)
- 3\textsuperscript{rd} comp.: 10\% rel. weight ($\sigma_3=10*\sigma_1$)
- 4\textsuperscript{th} comp.: 10\% rel. weight ($\sigma_4=15*\sigma_1$)

With 4 tracks: up to 256 components

For the Kalman filter, the collapsed state of the track has been used
Component limitation

Kalman Filter

GSF - No limitation of the number of components

\( x \) Residuals (\( \mu \text{m} \))

\( x \) Pull

\( P(\chi^2) \)

Thomas Speer

ACAT03 - 4th December 2003 - p. 19
Component limitation

GSF - Limit of 2 components (using Kullback-Leibler Distance)

GSF - Limit of 4 components (using Kullback-Leibler Distance)
Component limitation

<table>
<thead>
<tr>
<th>Nbr Comp.</th>
<th>Average $\chi^2$</th>
<th>Res. [µm]</th>
<th>Pull</th>
<th>Average $\chi^2$</th>
<th>Res. [µm]</th>
<th>Pull</th>
</tr>
</thead>
<tbody>
<tr>
<td>No limitation</td>
<td>0.99</td>
<td>178</td>
<td>0.85</td>
<td>Mahalanobis Distance</td>
<td>0.93</td>
<td>217</td>
</tr>
<tr>
<td>Kullback-Leibler Distance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.91</td>
<td>217</td>
<td>0.81</td>
<td>0.93</td>
<td>217</td>
<td>0.81</td>
</tr>
<tr>
<td>4</td>
<td>0.94</td>
<td>197</td>
<td>0.83</td>
<td>0.95</td>
<td>191</td>
<td>0.82</td>
</tr>
<tr>
<td>6</td>
<td>0.95</td>
<td>197</td>
<td>0.83</td>
<td>0.95</td>
<td>188</td>
<td>0.82</td>
</tr>
<tr>
<td>8</td>
<td>0.96</td>
<td>191</td>
<td>0.84</td>
<td>0.97</td>
<td>184</td>
<td>0.82</td>
</tr>
<tr>
<td>12</td>
<td>0.96</td>
<td>188</td>
<td>0.84</td>
<td>0.97</td>
<td>181</td>
<td>0.83</td>
</tr>
<tr>
<td>18</td>
<td>0.99</td>
<td>179</td>
<td>0.85</td>
<td>0.99</td>
<td>178</td>
<td>0.85</td>
</tr>
</tbody>
</table>

Relative efficiency

50% coverage

90% coverage

Thomas Speer

ACAT03 - 4th December 2003 - p. 21
Conclusion

- A Gaussian-sum Filter for vertex reconstruction has been implemented in the CMS reconstruction software.
- Shows an improvement of the resolution and error estimate of the fitted vertex and of the $\chi^2$ of the fit with respect of the Kalman Filter when the track parameters residuals have non-Gaussian tails.
- For electrons reconstructed with the GSF:
  - Allows to use the full mixture, and not only the single collapsed state.
- Shows little sensitivity to the number of components kept during fit.
- A small number of components can be kept without degrading the fit too much.