Track-based Alignment using a Kalman Filter Technique

R. Frühwirth and E. Widl

Institute for High-Energy Physics, Vienna, Austria
Outline

❖ Introduction
❖ Sequential updating
❖ Implementation and computational complexity
❖ Two-track fitter
❖ Examples
❖ Summary and outlook
Introduction

- Global alignment without having to solve a large system of linear equations
- Recursive approach, based on Kalman filter equations
- Update after each track
- Update is not restricted to modules crossed by the track
- Update is limited to modules with significant correlations
- Some bookkeeping required
Introduction

- Easy to use prior information from mechanical or laser alignment
- Easy to fix the position of reference detectors
- Method suitable for alignment relative to another detector
- Still in the experimental phase
### Sequential Updating

- **Notation for alignment related objects:**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>total number of alignable detector units (modules)</td>
</tr>
<tr>
<td>$d_t$</td>
<td>vector of true alignment parameters</td>
</tr>
<tr>
<td>$d_0$</td>
<td>expansion point of alignment parameters</td>
</tr>
<tr>
<td>$d$</td>
<td>current estimate of alignment parameters</td>
</tr>
<tr>
<td>$d_i$</td>
<td>subvector of alignment parameters of detector unit $i$</td>
</tr>
<tr>
<td>$D$</td>
<td>covariance matrix of $d$</td>
</tr>
<tr>
<td>$D_{ij}$</td>
<td>submatrix of cross-correlations between detector units $i$ and $j$</td>
</tr>
<tr>
<td>$\hat{d}$</td>
<td>updated estimate of alignment parameters</td>
</tr>
<tr>
<td>$\hat{D}$</td>
<td>covariance matrix of $\hat{d}$</td>
</tr>
</tbody>
</table>
Sequential Updating

- Notation for track related objects:

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I$</td>
<td>list of modules crossed by the current track</td>
</tr>
<tr>
<td>$k$</td>
<td>size of $I$</td>
</tr>
<tr>
<td>$m$</td>
<td>observations of the current track</td>
</tr>
<tr>
<td>$V$</td>
<td>covariance matrix of $m$</td>
</tr>
<tr>
<td>$x_t$</td>
<td>true track parameters of the current track</td>
</tr>
<tr>
<td>$x_0$</td>
<td>expansion point of track model</td>
</tr>
<tr>
<td>$x$</td>
<td>predicted track parameters of the current track</td>
</tr>
<tr>
<td>$C$</td>
<td>covariance matrix of $x$</td>
</tr>
<tr>
<td>$\hat{x}$</td>
<td>updated track parameters of the current track</td>
</tr>
</tbody>
</table>
Sequential Updating

- The observations \( m \) depend on the track parameters \( x_t \) via the track model \( f \):

\[
m = f(x_t) + \varepsilon, \quad \text{cov}(\varepsilon) = V
\]

- \( \varepsilon \) contains the effects of the observation error and of multiple scattering. Energy loss is considered as deterministic and is included in the track model.

- Its variance-covariance matrix \( V \) can be assumed to be known.

- A preliminary track fit gives a provisional estimate \( \hat{x} \) of the track parameters. The actual estimates of the alignment parameters are used at this stage.
Sequential Updating

- The observations also depend on the alignment parameters $d_t$.
- The track model is extended accordingly:

$$m = f(x_t, d_t) + \varepsilon, \quad \text{cov}(\varepsilon) = V$$

- First-order Taylor expansion at expansion points $d_0$ and $x_0$:

$$m = c + Ad_t + Bx_t + \varepsilon = c + \begin{pmatrix} A & B \end{pmatrix} \begin{pmatrix} d_t \\ x_t \end{pmatrix} + \varepsilon$$

- Jacobians:

$$A = \frac{\partial m}{\partial d_t}
\bigg|_{d_0}, \quad B = \frac{\partial m}{\partial x_t}
\bigg|_{x_0}$$
Sequential Updating

- **Constant:**
  \[ c = f(x_0, d_0) - A d_0 - B x_0 \]

- **Expansion point** \( d_0 \): the nominal or the currently estimated module alignment.

- **Expansion point** \( x_0 \): result of a preliminary track fit.

- The Kalman filter requires a prediction \( x \) of the track parameters, along with its variance-covariance matrix \( C \).

- The prediction has to be stochastically independent of the observations in the track.
Sequential Updating

- First case: independent prediction exists
  - External prediction from already aligned detector
  - External information form vertex or kinematical constraint

- Update equation of the Kalman filter:

\[
\begin{pmatrix}
\hat{d} \\
\hat{x}
\end{pmatrix} = \begin{pmatrix}
d \\
x
\end{pmatrix} + K (m - c - Ad - Bx)
\]
Sequential Updating

- Gain matrix of the filter:

\[
K = \begin{pmatrix} D & 0 \\ 0 & C \end{pmatrix} \begin{pmatrix} A^T \\ B^T \end{pmatrix} \left( V + ADA^T + BCB^T \right)^{-1}
\]

\[
= \begin{pmatrix} DA^T G \\ CB^T G \end{pmatrix}
\]

- Update decomposes:

\[
\hat{d} = d + DA^T G (m - c - Ad - Bx)
\]

\[
\hat{x} = x + CB^T G (m - c - Ad - Bx)
\]
Sequential Updating

- Case two: no independent prediction exists
- The prediction $x_0$ gets zero weight in order not to bias the estimation
- Multiply $C$ by a scale factor $\alpha$ and let $\alpha$ tend to infinity:

$$G = \lim_{\alpha \to \infty} \left( V + ADA^T + \alpha BC B^T \right)^{-1}$$

$$= V_D^{-1} - V_D^{-1} B (B^T V_D^{-1} B)^{-1} B^T V_D^{-1}$$

with

$$V_D = V + ADA^T$$
Sequential Updating

Because of $GB = 0$ the update equation of the alignment parameters can be simplified to

$$\hat{d} = d + DA^T G (m - c - Ad)$$

The update of the covariance matrix can be calculated by linear error propagation:

$$\hat{D} = (I - DA^T GA) D (I - A^T GAD) + DA^T GV GAD$$

Both terms on the right hand side are positive definite, so the left hand side is guaranteed to be positive definite as well.
Sequential Updating

- The iteration needs starting values.
- Mechanical and laser alignment can be used for the starting values.
- Reference units can be fixed by giving them very small initial errors.
Implementation and computational complexity

- The current track crosses $k$ detector units.
- The dimension $n = 2^k$ of $m$ is small, in the order of 30 for the CMS Inner Tracker.
- The matrix $B$ is of size $n \times 5$ and is therefore small.
- The matrix $A$ is a row of $N$ blocks $A_i$ of size $n \times m$, where $m$ is the number of alignment parameters per detector unit (usually 6).
- For each track, only $k$ out of these $N$ blocks are different from zero:

$$A = \begin{pmatrix} 0 & \cdots & 0 & A_{i_1} & 0 & \cdots & 0 & A_{i_2} & 0 & \cdots & 0 & A_{i_k} & 0 & \cdots & 0 \end{pmatrix}$$
Implementation and computational complexity

- The only large matrix in the update formulas is the product $DA^T$.
- It is a column of $N$ blocks each of which has size $m \times n$.
- Complete computation of $DA^T$ would lead to an algorithm that scales with $N^2$.
- This is too slow for practical purposes.
Implementation and computational complexity

- Two alternatives:
  - Algorithm A: Compute only the blocks of the modules in the current track, neglecting all correlations.
  - Algorithm B: Compute the blocks of the modules having significant correlations with the modules in the current track.

- Algorithm A gives an unbiased estimate, but is suboptimal because of the missing correlations.

- Algorithm B is nearly optimal, but it has to be guaranteed that $\hat{D}$ is positive definite all the time. This problem is being studied, but there is not yet a foolproof solution.
Implementation and computational complexity

- Tradeoff between speed and precision.

- In order to keep track of the necessary updates, a list $L_i$ is attached to each detector unit $i$, containing the detector units that have significant correlations with $i$.

- This list may contain only $i$ itself in the beginning and grows as more tracks are processed.

- If there is prior knowledge about correlations, for instance because of mechanical constraints, it can be incorporated in the list and in the initial covariance matrix.
Implementation and computational complexity

- Update of the alignment parameters

1. Update the list $L_i$ for every $i \in I$ (see below).
2. Form the list $L$ of all detector units that are correlated with the ones crossed by the current track: $L = \bigcup_{i \in I} L_i$. The size of $L$ should be much smaller than $N$.
3. For all $j \in L$ compute: $(DA^T)_j = \sum_{i \in I} D_{ji} A^T_i$. Each block $D_{ji}$ is of size $m \times m$.
4. Compute: $ADA^T = \sum_{i \in I} A_i(DA^T)_i$.
5. Compute: $V_D = V + ADA^T$ and $G$. All matrices involved are of size $n \times n$.
6. Compute: $m' = G(m - c - \sum_{i \in I} A_i d_i)$.
7. For all $j \in L$ compute: $\hat{d}_j = d_j + (DA^T)_j m'$.
Implementation and computational complexity

- Update of the covariance matrix
  
  For all $j, l \in L$ compute: 
  $$\hat{D}_{jl} = D_{jl} + (DA^T)_j(GV_DG - 2G)[(DA^T)_l]^T$$

- The computational complexity of the parameter update is of the order $|L| \cdot |I|$.

- The computational complexity of the update of the covariance matrix is of the order $|L|^2$.

- Restricting the size of the lists $L_i$ is of crucial importance.
Implementation and computational complexity

- Current proposal for building the lists \( L_i \) is based on the concept of a distance between two modules \( i \) and \( j \).

- Define the following relation:

\[
 i \sim j \iff i \text{ and } j \text{ have been crossed by the same track}
\]

- The relation “\( \sim \)” is symmetric, but not transitive.

- Define the distance \( d(i, j) \) between detector units \( i \) and \( j \) by:

1. \( d(i, i) = 0 \)
2. If \( i \neq j \) and \( i \sim i_1 \sim i_2 \sim \cdots \sim i_n \sim j \) is the shortest chain connecting \( i \) to \( j \), the distance is \( d(i, j) = n + 1 \).
Implementation and computational complexity

- $i \sim j \iff d(i, j) = 1$

- $d$ is a proper metrics:

  1. $d(i, j) = 0$ if and only of $i = j$
  2. $d(i, j) = d(j, i)$
  3. $d(i, j) \leq d(i, k) + d(k, j)$ for all $k$

- The list $L_i$ contains all modules $k$ with $d(k, i) \leq d_{\text{max}}$. 
Implementation and computational complexity

- Alternative approaches conceivable:
  - Let the lists grow until the correlations have stabilized. Then, drop all correlations below an upper limit. Dynamic, adapts to the track sample used.
  - Determine “optimal” correlation structure from simulated data. Static, has to be done separately for every potential track sample (cosmics, beam halo, interactions).

- Detailed studies required.
Two-track fitter

- Use vertex- and mass-constrained track pairs to improve momentum resolution

- Examples: $Z \rightarrow \mu^+\mu^-$, $J/\psi \rightarrow \mu^+\mu^-$

- The five track parameters are replaced by nine decay parameters:
  - the position $v$ of the decay vertex
  - the momentum $p$ of the mother particle in the lab frame
  - the two decay angles $(\theta, \varphi)$ in the rest frame of the mother particle
  - the mass $m$ of the mother particle
Two-track fitter

The mass is constrained by adding a virtual observation of the mass (theoretical value plus width).

The Jacobians w.r.t. the decay parameters are obtained by the chain rule:

\[
\frac{\partial m_i}{\partial (v, p, \theta, \varphi)} = \frac{\partial m_i}{\partial x_i} \cdot \frac{\partial x_i}{\partial (v, p_i)} \cdot \frac{\partial (v, p_i)}{\partial (p, \theta, \varphi)}
\]

Otherwise the formalism remains unchanged.
Examples

- Two setups, subsets of CMS Tracker
  - “Small wheel”: 3 Pixel layers (180 modules), 4 TIB layers (156 modules)
  - “Large wheel”: 3 Pixel layers (180 modules), 4 TIB layers (156 modules), 6 TOB layers (344 modules)

- Alignment of TIB and TOB relative to Pixel barrel
Examples

- Small wheel
- Pixels are fixed, 156 TIB modules are aligned
- Misalignment:
  \[ \sigma(\Delta u) = 100 \, \mu m, \sigma(\Delta v) = 100 \, \mu m, \sigma(\Delta \gamma) = 5 \, \text{mrad} \]
- 25000 muon pairs from \( Z \rightarrow \mu^+ \mu^- \) (50000 tracks)
  - No correlations
  - Full correlations
Examples

Small wheel with 4 TIB layers
Examples

Residuals after alignment

R. Frühwirth, E. Widl

LHC Alignment Workshop
Examples

Standardized residuals after alignment

R. Frühwirth, E. Widl  LHC Alignment Workshop  29
Examples

Evolution of residuals
Examples

no correlations - 3 estimated parameters

full correlations - 3 estimated parameters

no correlations - 1 estimated parameter

full correlations - 1 estimated parameter

Evolution of residuals (1 or 3 alignment parameters)
Examples

- Large wheel
- pixels are fixed, 500 TIB/TOB modules are aligned
- Misalignment:
  \[ \sigma(\Delta u) = 100 \mu m, \sigma(\Delta v) = 100 \mu m \]
- 25000 muon pairs from \( Z \rightarrow \mu^+ \mu^- \) (50000 tracks)
  - No correlations
  - Full correlations
  - Correlations up to \( d_{\text{max}} = 5 \)
  - Two-track fitter with correlations up to \( d_{\text{max}} = 5 \)
Examples

Residuals after alignment

R. Frühwirth, E. Widl
LHC Alignment Workshop
Examples

Standardized residuals after alignment
Examples

Evolution of residuals
Examples

Evolution of correlations

size of correlations

updates
- 1.000
- 5.000
- 10.000
- 15.000
- 20.000

R. Frühwirth, E. Widl

LHC Alignment Workshop
Examples

- Approximate total processing times
  - No correlations: 1.5 h
  - Full correlations: 39.5 h
  - Correlations up to $d_{\text{max}} = 5$: 7.5 h
  - Two-track fitter with correlations up to $d_{\text{max}} = 5$: 16 h

- Can be speeded up considerably by using pixel tracks as prediction (no update of pixel modules)
Summary and Outlook

- Kalman filter for sequential estimation of alignment constants
- Successful test on small-scale setups

Advantages

✧ No solution of large systems of equations
✧ Depth of correlations can be tailored to setup
✧ Errors of estimated alignment constants are always available
✧ Can be used for stopping criterion
Summary and Outlook

- Disadvantages
  - Larger computational expense per track
  - More bookkeeping required

- Outlook
  - Extend to full set of angles and shifts
  - Study alternative approaches to correlation lists
  - Speed optimization
  - Large-scale examples
Acknowledgements

- We thank Wolfgang Adam (HEPHY Vienna) for technical support.
- We thank CERN for financial support to E. Widl.