Lecture 7: Optimal Smoothing

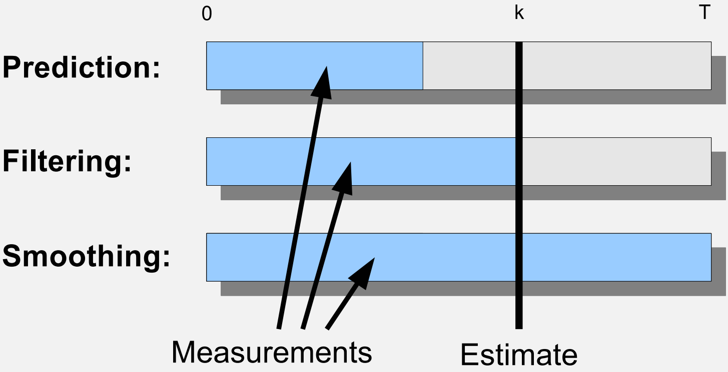
Simo Särkkä

Department of Biomedical Engineering and Computational Science
Aalto University

May 5, 2010

Contents

Filtering, Prediction and Smoothing



Types of Smoothing Problems

- Fixed-interval smoothing: estimate states on interval $[0, T]$ given measurements on the same interval.
- Fixed-point smoothing: estimate state at a fixed point of time in the past.
- Fixed-lag smoothing: estimate state at a fixed delay in the past.
- Here we shall only consider fixed-interval smoothing, the others can be quite easily derived from it.

Examples of Smoothing Problems

- Given all the radar measurements of a rocket (or missile) trajectory, what was the exact place of launch?
- Estimate the whole trajectory of a car based on GPS measurements to calibrate the inertial navigation system accurately.
- What was the history of chemical/combustion/other process given a batch of measurements from it?
- Remove noise from audio signal by using smoother to estimate the true audio signal under the noise.
- Smoothing solution also arises in EM algorithm for estimating the parameters of a state space model.

Optimal Smoothing Algorithms

- Linear Gaussian models
  - Rauch-Tung-Striebel smoother (RTSS).
  - Two-filter smoother.
- Non-linear Gaussian models
  - Extended Rauch-Tung-Striebel smoother (ERTSS).
  - Unscented Rauch-Tung-Striebel smoother (URTSS).
  - Statistically linearized Rauch-Tung-Striebel smoother (URTSS).
  - Two-filter versions of the above.
- Non-linear non-Gaussian models
  - Sequential importance resampling based smoother.
  - Rao-Blackwellized particle smoothers.
  - Grid based smoother.

Problem Formulation

- Probabilistic state space model:

$$\text{measurement model: } \mathbf{y}_k \sim p(\mathbf{y}_k \mid \mathbf{x}_k)$$
$$\text{dynamic model: } \mathbf{x}_k \sim p(\mathbf{x}_k \mid \mathbf{x}_{k-1})$$

- Assume that the filtering distributions $p(\mathbf{x}_k \mid \mathbf{y}_{1:k})$ have already beed computed for all $k = 0, \ldots, T$.
- We want recursive equations of computing the smoothing distribution for all $k < T$:

$$p(\mathbf{x}_k \mid \mathbf{y}_{1:T}).$$

- The recursion will go backwards in time, because on the last step, the filtering and smoothing distributions coincide:

$$p(\mathbf{x}_T \mid \mathbf{y}_{1:T}).$$

Derivation of Formal Smoothing Equations [1/2]

- The key: due to the Markov properties of state we have:

$$p(\mathbf{x}_k \,|\, \mathbf{x}_{k+1}, \mathbf{y}_{1:T}) = p(\mathbf{x}_k \,|\, \mathbf{x}_{k+1}, \mathbf{y}_{1:k})$$

- Thus we get:

$$\begin{aligned} p(\mathbf{x}_k \,|\, \mathbf{x}_{k+1}, \mathbf{y}_{1:T}) &= p(\mathbf{x}_k \,|\, \mathbf{x}_{k+1}, \mathbf{y}_{1:k}) \\ &= \frac{p(\mathbf{x}_k, \mathbf{x}_{k+1} \,|\, \mathbf{y}_{1:k})}{p(\mathbf{x}_{k+1} \,|\, \mathbf{y}_{1:k})} \\ &= \frac{p(\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{y}_{1:k}) \, p(\mathbf{x}_k \,|\, \mathbf{y}_{1:k})}{p(\mathbf{x}_{k+1} \,|\, \mathbf{y}_{1:k})} \\ &= \frac{p(\mathbf{x}_{k+1} \,|\, \mathbf{x}_k) \, p(\mathbf{x}_k \,|\, \mathbf{y}_{1:k})}{p(\mathbf{x}_{k+1} \,|\, \mathbf{y}_{1:k})}. \end{aligned}$$

Derivation of Formal Smoothing Equations [2/2]

- Assuming that the smoothing distribution of the next step $p(\mathbf{x}_{k+1} \,|\, \mathbf{y}_{1:T})$ is available, we get

$$
\begin{aligned}
p(\mathbf{x}_k, \mathbf{x}_{k+1} \,|\, \mathbf{y}_{1:T}) &= p(\mathbf{x}_k \,|\, \mathbf{x}_{k+1}, \mathbf{y}_{1:T}) \, p(\mathbf{x}_{k+1} \,|\, \mathbf{y}_{1:T}) \\
&= p(\mathbf{x}_k \,|\, \mathbf{x}_{k+1}, \mathbf{y}_{1:k}) \, p(\mathbf{x}_{k+1} \,|\, \mathbf{y}_{1:T}) \\
&= \frac{p(\mathbf{x}_{k+1} \,|\, \mathbf{x}_k) \, p(\mathbf{x}_k \,|\, \mathbf{y}_{1:k}) \, p(\mathbf{x}_{k+1} \,|\mathbf{y}_{1:T})}{p(\mathbf{x}_{k+1} \,|\, \mathbf{y}_{1:k})}
\end{aligned}
$$

- Integrating over $\mathbf{x}_{k+1}$ gives

$$
p(\mathbf{x}_k \,|\, \mathbf{y}_{1:T}) = p(\mathbf{x}_k \,|\, \mathbf{y}_{1:k}) \int \left[\frac{p(\mathbf{x}_{k+1} \,|\, \mathbf{x}_k) \, p(\mathbf{x}_{k+1} \,|\, \mathbf{y}_{1:T})}{p(\mathbf{x}_{k+1} \,|\, \mathbf{y}_{1:k})} \right] \mathrm{d}\mathbf{x}_{k+1}
$$

Bayesian Optimal Smoothing Equations

The Bayesian optimal smoothing equations consist of prediction step and backward update step:

$$p(\mathbf{x}_{k+1} \mid \mathbf{y}_{1:k}) = \int p(\mathbf{x}_{k+1} \mid \mathbf{x}_k) \, p(\mathbf{x}_k \mid \mathbf{y}_{1:k}) \, \mathrm{d}\mathbf{x}_k$$

$$p(\mathbf{x}_k \mid \mathbf{y}_{1:T}) = p(\mathbf{x}_k \mid \mathbf{y}_{1:k}) \int \left[\frac{p(\mathbf{x}_{k+1} \mid \mathbf{x}_k) \, p(\mathbf{x}_{k+1} \mid \mathbf{y}_{1:T})}{p(\mathbf{x}_{k+1} \mid \mathbf{y}_{1:k})} \right] \mathrm{d}\mathbf{x}_{k+1}$$

The recursion is started from the filtering (and smoothing) distribution of the last time step $p(\mathbf{x}_T \mid \mathbf{y}_{1:T})$.

Linear-Gaussian Smoothing Problem

- Gaussian driven linear model, i.e., Gauss-Markov model:

$$\mathbf{x}_k = \mathbf{A}_{k-1}\,\mathbf{x}_{k-1} + \mathbf{q}_{k-1}$$
$$\mathbf{y}_k = \mathbf{H}_k\,\mathbf{x}_k + \mathbf{r}_k,$$

- In probabilistic terms the model is

$$p(\mathbf{x}_k \,|\, \mathbf{x}_{k-1}) = \mathrm{N}(\mathbf{x}_k \,|\, \mathbf{A}_{k-1}\,\mathbf{x}_{k-1}, \mathbf{Q}_{k-1})$$
$$p(\mathbf{y}_k \,|\, \mathbf{x}_k) = \mathrm{N}(\mathbf{y}_k \,|\, \mathbf{H}_k\,\mathbf{x}_k, \mathbf{R}_k).$$

- Kalman filter can be used for computing all the Gaussian filtering distributions:

$$p(\mathbf{x}_k \,|\, \mathbf{y}_{1:k}) = \mathrm{N}(\mathbf{x}_k \,|\, \mathbf{m}_k, \mathbf{P}_k).$$

Derivation of Rauch-Tung-Striebel Smoother [1/4]

- By the Gaussian distribution computation rules we get

$$\begin{aligned} p(\mathbf{x}_k, \mathbf{x}_{k+1} \mid \mathbf{y}_{1:k}) &= p(\mathbf{x}_{k+1} \mid \mathbf{x}_k)\, p(\mathbf{x}_k \mid \mathbf{y}_{1:k}) \\ &= \mathrm{N}(\mathbf{x}_{k+1} \mid \mathbf{A}_k\, \mathbf{x}_k, \mathbf{Q}_k)\; \mathrm{N}(\mathbf{x}_k \mid \mathbf{m}_k, \mathbf{P}_k) \\ &= \mathrm{N}\left(\begin{bmatrix} \mathbf{x}_k \\ \mathbf{x}_{k+1} \end{bmatrix} \,\middle|\, \mathbf{m}_1, \mathbf{P}_1 \right), \end{aligned}$$

where

$$\mathbf{m}_1 = \begin{pmatrix} \mathbf{m}_k \\ \mathbf{A}_k\, \mathbf{m}_k \end{pmatrix}, \qquad \mathbf{P}_1 = \begin{pmatrix} \mathbf{P}_k & \mathbf{P}_k\, \mathbf{A}_k^T \\ \mathbf{A}_k\, \mathbf{P}_k & \mathbf{A}_k\, \mathbf{P}_k\, \mathbf{A}_k^T + \mathbf{Q}_k \end{pmatrix}.$$

Derivation of Rauch-Tung-Striebel Smoother [2/4]

- By conditiong rule of Gaussian distribution we get

$$\begin{aligned} p(\mathbf{x}_k \mid \mathbf{x}_{k+1}, \mathbf{y}_{1:T}) &= p(\mathbf{x}_k \mid \mathbf{x}_{k+1}, \mathbf{y}_{1:k}) \\ &= \mathrm{N}(\mathbf{x}_k \mid \mathbf{m}_2, \mathbf{P}_2), \end{aligned}$$

where

$$\begin{aligned} \mathbf{C}_k &= \mathbf{P}_k \, \mathbf{A}_k^T \, (\mathbf{A}_k \, \mathbf{P}_k \, \mathbf{A}_k^T + \mathbf{Q}_k)^{-1} \\ \mathbf{m}_2 &= \mathbf{m}_k + \mathbf{C}_k \, (\mathbf{x}_{k+1} - \mathbf{A}_k \, \mathbf{m}_k) \\ \mathbf{P}_2 &= \mathbf{P}_k - \mathbf{C}_k \, (\mathbf{A}_k \, \mathbf{P}_k \, \mathbf{A}_k^T + \mathbf{Q}_k) \, \mathbf{C}_k^T. \end{aligned}$$

Derivation of Rauch-Tung-Striebel Smoother [3/4]

- The joint distribution of $\mathbf{x}_k$ and $\mathbf{x}_{k+1}$ given all the data is

$$
\begin{aligned}
p(\mathbf{x}_{k+1}, \mathbf{x}_k \mid \mathbf{y}_{1:T}) &= p(\mathbf{x}_k \mid \mathbf{x}_{k+1}, \mathbf{y}_{1:T})\, p(\mathbf{x}_{k+1} \mid \mathbf{y}_{1:T}) \\
&= \mathrm{N}(\mathbf{x}_k \mid \mathbf{m}_2, \mathbf{P}_2)\ \mathrm{N}(\mathbf{x}_{k+1} \mid \mathbf{m}^s_{k+1}, \mathbf{P}^s_{k+1}) \\
&= \mathrm{N}\left(\begin{bmatrix} \mathbf{x}_{k+1} \\ \mathbf{x}_k \end{bmatrix} \,\middle|\, \mathbf{m}_3, \mathbf{P}_3 \right)
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{m}_3 &= \begin{pmatrix} \mathbf{m}^s_{k+1} \\ \mathbf{m}_k + \mathbf{C}_k\, (\mathbf{m}^s_{k+1} - \mathbf{A}_k\, \mathbf{m}_k) \end{pmatrix} \\
\mathbf{P}_3 &= \begin{pmatrix} \mathbf{P}^s_{k+1} & \mathbf{P}^s_{k+1}\, \mathbf{C}_k^T \\ \mathbf{C}_k\, \mathbf{P}^s_{k+1} & \mathbf{C}_k\, \mathbf{P}^s_{k+1}\, \mathbf{C}_k^T + \mathbf{P}_2 \end{pmatrix}.
\end{aligned}
$$

Derivation of Rauch-Tung-Striebel Smoother [4/4]

- The marginal mean and covariance are thus given as

$$\mathbf{m}_k^s = \mathbf{m}_k + \mathbf{C}_k \, (\mathbf{m}_{k+1}^s - \mathbf{A}_k \, \mathbf{m}_k)$$
$$\mathbf{P}_k^s = \mathbf{P}_k + \mathbf{C}_k \, (\mathbf{P}_{k+1}^s - \mathbf{A}_k \, \mathbf{P}_k \, \mathbf{A}_k^T - \mathbf{Q}_k) \, \mathbf{C}_k^T.$$

- The smoothing distribution is then Gaussian with the above mean and covariance:

$$p(\mathbf{x}_k \mid \mathbf{y}_{1:T}) = \mathrm{N}(\mathbf{x}_k \mid \mathbf{m}_k^s, \mathbf{P}_k^s),$$

Rauch-Tung-Striebel Smoother

Rauch-Tung-Striebel Smoother

Backward recursion equations for the smoothed means $\mathbf{m}_k^s$ and covariances $\mathbf{P}_k^s$:

$$
\begin{aligned}
\mathbf{m}_{k+1}^- &= \mathbf{A}_k \, \mathbf{m}_k \\
\mathbf{P}_{k+1}^- &= \mathbf{A}_k \, \mathbf{P}_k \, \mathbf{A}_k^T + \mathbf{Q}_k \\
\mathbf{C}_k &= \mathbf{P}_k \, \mathbf{A}_k^T \, [\mathbf{P}_{k+1}^-]^{-1} \\
\mathbf{m}_k^s &= \mathbf{m}_k + \mathbf{C}_k \, [\mathbf{m}_{k+1}^s - \mathbf{m}_{k+1}^-] \\
\mathbf{P}_k^s &= \mathbf{P}_k + \mathbf{C}_k \, [\mathbf{P}_{k+1}^s - \mathbf{P}_{k+1}^-] \, \mathbf{C}_k^T,
\end{aligned}
$$

- $\mathbf{m}_k$ and $\mathbf{P}_k$ are the mean and covariance computed by the Kalman filter.
- The recursion is started from the last time step T, with $\mathbf{m}_T^s = \mathbf{m}_T$ and $\mathbf{P}_T^s = \mathbf{P}_T$.

RTS Smoother: Car Tracking Example

The dynamic model of the car tracking model from the first & third lectures was:

$$\begin{pmatrix} x_k \\ y_k \\ \dot{x}_k \\ \dot{y}_k \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\mathbf{A}} \begin{pmatrix} x_{k-1} \\ y_{k-1} \\ \dot{x}_{k-1} \\ \dot{y}_{k-1} \end{pmatrix} + \mathbf{q}_{k-1}$$

where $\mathbf{q}_k$ is zero mean with a covariance matrix $\mathbf{Q}$.

$$\mathbf{Q} = \begin{pmatrix} q_1^c \, \Delta t^3/3 & 0 & q_1^c \, \Delta t^2/2 & 0 \\ 0 & q_2^c \, \Delta t^3/3 & 0 & q_2^c \, \Delta t^2/2 \\ q_1^c \, \Delta t^2/2 & 0 & q_1^c \, \Delta t & 0 \\ 0 & q_2^c \, \Delta t^2/2 & 0 & q_2^c \, \Delta t \end{pmatrix}$$

Non-Linear Smoothing Problem

- Non-linear Gaussian state space model:

$$\begin{aligned}\mathbf{x}_k &= \mathbf{f}(\mathbf{x}_{k-1}) + \mathbf{q}_{k-1} \\ \mathbf{y}_k &= \mathbf{h}(\mathbf{x}_k) + \mathbf{r}_k,\end{aligned}$$

- We want to compute Gaussian approximations to the smoothing distributions:

$$p(\mathbf{x}_k \mid \mathbf{y}_{1:T}) \approx \mathrm{N}(\mathbf{x}_k \mid \mathbf{m}_k^s, \mathbf{P}_k^s).$$

Extended Rauch-Tung-Striebel Smoother Derivation

- The approximate joint distribution of $\mathbf{x}_k$ and $\mathbf{x}_{k+1}$ is

$$p(\mathbf{x}_k, \mathbf{x}_{k+1} \mid \mathbf{y}_{1:k}) = \mathrm{N}\left(\begin{bmatrix} \mathbf{x}_k \\ \mathbf{x}_{k+1} \end{bmatrix} \,\Big|\, \mathbf{m}_1, \mathbf{P}_1 \right),$$

  where

$$\begin{aligned} \mathbf{m}_1 &= \begin{pmatrix} \mathbf{m}_k \\ \mathbf{f}(\mathbf{m}_k) \end{pmatrix} \\ \mathbf{P}_1 &= \begin{pmatrix} \mathbf{P}_k & \mathbf{P}_k \, \mathbf{F}_x^T(\mathbf{m}_k) \\ \mathbf{F}_x(\mathbf{m}_k) \, \mathbf{P}_k & \mathbf{F}_x(\mathbf{m}_k) \, \mathbf{P}_k \, \mathbf{F}_x^T(\mathbf{m}_k) + \mathbf{Q}_k \end{pmatrix}. \end{aligned}$$

- The rest of the derivation is analogous to the linear RTS smoother.

Extended Rauch-Tung-Striebel Smoother

Extended Rauch-Tung-Striebel Smoother

The equations for the extended RTS smoother are

$$\begin{aligned}
\mathbf{m}_{k+1}^{-} &= \mathbf{f}(\mathbf{m}_k) \\
\mathbf{P}_{k+1}^{-} &= \mathbf{F}_{\mathbf{x}}(\mathbf{m}_k)\,\mathbf{P}_k\,\mathbf{F}_{\mathbf{x}}^T(\mathbf{m}_k) + \mathbf{Q}_k \\
\mathbf{C}_k &= \mathbf{P}_k\,\mathbf{F}_{\mathbf{x}}^T(\mathbf{m}_k)\,[\mathbf{P}_{k+1}^{-}]^{-1} \\
\mathbf{m}_k^s &= \mathbf{m}_k + \mathbf{C}_k\,[\mathbf{m}_{k+1}^s - \mathbf{m}_{k+1}^{-}] \\
\mathbf{P}_k^s &= \mathbf{P}_k + \mathbf{C}_k\,[\mathbf{P}_{k+1}^s - \mathbf{P}_{k+1}^{-}]\,\mathbf{C}_k^T,
\end{aligned}$$

where the matrix $\mathbf{F}_{\mathbf{x}}(\mathbf{m}_k)$ is the Jacobian matrix of $\mathbf{f}(\mathbf{x})$ evaluated at $\mathbf{m}_k$.

Statistically Linearized Rauch-Tung-Striebel Smoother Derivation

- With statistical linearization we get the approximation

$$p(\mathbf{x}_k, \mathbf{x}_{k+1} \mid \mathbf{y}_{1:k}) = \mathrm{N}\left(\begin{bmatrix} \mathbf{x}_k \\ \mathbf{x}_{k+1} \end{bmatrix} \,\middle|\, \mathbf{m}_1, \mathbf{P}_1 \right),$$

where

$$\mathbf{m}_1 = \begin{pmatrix} \mathbf{m}_k \\ \mathrm{E}[\mathbf{f}(\mathbf{x}_k)] \end{pmatrix}$$

$$\mathbf{P}_1 = \begin{pmatrix} \mathbf{P}_k & \mathrm{E}[\mathbf{f}(\mathbf{x}_k)\,\delta\mathbf{x}_k^T]^T \\ \mathrm{E}[\mathbf{f}(\mathbf{x}_k)\,\delta\mathbf{x}_k^T] & \mathrm{E}[\mathbf{f}(\mathbf{x}_k)\,\delta\mathbf{x}_k^T]\,\mathbf{P}_k^{-1}\,\mathrm{E}[\mathbf{f}(\mathbf{x}_k)\,\delta\mathbf{x}_k^T]^T + \mathbf{Q}_k \end{pmatrix}.$$

- The expectations are taken with respect to filtering distribution of $\mathbf{x}_k$.
- The derivation proceeds as with linear RTS smoother.

Statistically Linearized Rauch-Tung-Striebel Smoother

The equations for the statistically linearized RTS smoother are

$$
\begin{aligned}
\mathbf{m}_{k+1}^- &= \mathrm{E}[\mathbf{f}(\mathbf{x}_k)] \\
\mathbf{P}_{k+1}^- &= \mathrm{E}[\mathbf{f}(\mathbf{x}_k)\, \delta\mathbf{x}_k^T]\, \mathbf{P}_k^{-1}\; \mathrm{E}[\mathbf{f}(\mathbf{x}_k)\, \delta\mathbf{x}_k^T]^T + \mathbf{Q}_k \\
\mathbf{C}_k &= \mathrm{E}[\mathbf{f}(\mathbf{x}_k)\, \delta\mathbf{x}_k^T]^T [\mathbf{P}_{k+1}^-]^{-1} \\
\mathbf{m}_k^s &= \mathbf{m}_k + \mathbf{C}_k\, [\mathbf{m}_{k+1}^s - \mathbf{m}_{k+1}^-] \\
\mathbf{P}_k^s &= \mathbf{P}_k + \mathbf{C}_k\, [\mathbf{P}_{k+1}^s - \mathbf{P}_{k+1}^-]\, \mathbf{C}_k^T,
\end{aligned}
$$

where the expectations are taken with respect to the filtering distribution $\mathbf{x}_k \sim \mathrm{N}(\mathbf{m}_k, \mathbf{P}_k)$.

Unscented Rauch-Tung-Striebel Smoother [1/2]

Unscented Rauch-Tung-Striebel Smoother

1. Form the matrix of sigma points:

$$\mathbf{X}_k = \begin{bmatrix} \mathbf{m}_k & \cdots & \mathbf{m}_k \end{bmatrix} + \sqrt{n+\lambda} \begin{bmatrix} \mathbf{0} & \sqrt{\mathbf{P}_k} & -\sqrt{\mathbf{P}_k} \end{bmatrix}.$$

2. Propagate the sigma points through the dynamic model:

$$\hat{\mathbf{X}}_{k+1,i} = \mathbf{f}(\mathbf{X}_{k,i}), \quad i = 1 \ldots 2n+1.$$

3. Compute the following:

$$\begin{aligned}
\mathbf{m}_{k+1}^- &= \sum_i W_{i-1}^{(m)} \, \hat{\mathbf{X}}_{k+1,i} \\
\mathbf{P}_{k+1}^- &= \sum_i W_{i-1}^{(c)} \, (\hat{\mathbf{X}}_{k+1,i} - \mathbf{m}_{k+1}^-) \, (\hat{\mathbf{X}}_{k+1,i} - \mathbf{m}_{k+1}^-)^T + \mathbf{Q}_k \\
\mathbf{D}_{k+1} &= \sum_i W_{i-1}^{(c)} \, (\mathbf{X}_{k,i} - \mathbf{m}_k) \, (\hat{\mathbf{X}}_{k+1,i} - \mathbf{m}_{k+1}^-)^T.
\end{aligned}$$

Unscented Rauch-Tung-Striebel Smoother (cont.)

3. Compute the smoother gain $\mathbf{C}_k$, the smoothed mean $\mathbf{m}_k^s$ and the covariance $\mathbf{P}_k^s$ as follows:

$$
\begin{aligned}
\mathbf{C}_k &= \mathbf{D}_{k+1} \, [\mathbf{P}_{k+1}^-]^{-1} \\
\mathbf{m}_k^s &= \mathbf{m}_k + \mathbf{C}_k \, (\mathbf{m}_{k+1}^s - \mathbf{m}_{k+1}^-) \\
\mathbf{P}_k^s &= \mathbf{P}_k + \mathbf{C}_k \, (\mathbf{P}_{k+1}^s - \mathbf{P}_{k+1}^-) \, \mathbf{C}_k^T.
\end{aligned}
$$

Particle Smoothing [1/2]

- The smoothing solution can be obtained from SIR by storing the whole state histories into the particles.
- Special care is needed on the resampling step.
- The smoothed distribution approximation is then of the form

$$p(\mathbf{x}_k \mid \mathbf{y}_{1:T}) \approx \sum_{i=1}^{N} w_T^{(i)} \delta(\mathbf{x}_k - \mathbf{x}_k^{(i)}).$$

  where $\mathbf{x}_k^{(i)}$ is the kth component in $\mathbf{x}_{1:T}^{(i)}$.
- Unfortunately, the approximation is often quite degenerate.
- Specialized algorithms for particle smoothing exists.

Particle Smoothing [2/2]

- Recall the Rao-Blackwellized particle filtering model:

$$
\begin{aligned}
\mathbf{s}_k &\sim p(\mathbf{s}_k \,|\, \mathbf{s}_{k-1}) \\
\mathbf{x}_k &= \mathbf{A}(\mathbf{s}_{k-1})\, \mathbf{x}_{k-1} + \mathbf{q}_k, \qquad \mathbf{q}_k \sim \mathrm{N}(\mathbf{0}, \mathbf{Q}) \\
\mathbf{y}_k &= \mathbf{H}(\mathbf{s}_k)\, \mathbf{x}_k + \mathbf{r}_k, \qquad \mathbf{r}_k \sim \mathrm{N}(\mathbf{0}, \mathbf{R})
\end{aligned}
$$

- The principle of Rao-Blackwellized particle smoothing is the following:
  1. During filtering store the whole sampled state and Kalman filter histories to the particles.
  2. At the smoothing step, apply Rauch-Tung-Striebel smoothers to each of the Kalman filter histories in the particles.
- The smoothing distribution approximation will then be of the form

$$
p(\mathbf{x}_k, \mathbf{s}_k \,|\, \mathbf{y}_{1:T}) \approx \sum_{i=1}^{N} w_T^{(i)}\, \delta(\mathbf{s}_k - \mathbf{s}_k^{(i)})\, \mathrm{N}(\mathbf{x}_k \,|\, \mathbf{m}_k^{s,(i)}, \mathbf{P}_k^{s,(i)}).
$$

Summary

- Optimal smoothing is used for computing estimates of state trajectories given the measurements on the whole trajectory.
- Rauch-Tung-Striebel (RTS) smoother is the closed form smoother for linear Gaussian models.
- Extended, statistically linearized and unscented RTS smoothers are the approximate nonlinear smoothers corresponding to EKF, SLF and UKF.
- Particle smoothing can be done by storing the whole state histories in SIR algorithm.
- Rao-Blackwellized particle smoother is a combination of particle smoothing and RTS smoothing.

- Pendulum model:

$$\begin{pmatrix} x_k^1 \\ x_k^2 \end{pmatrix} = \underbrace{\begin{pmatrix} x_{k-1}^1 + x_{k-1}^2 \, \Delta t \\ x_{k-1}^2 - g \, \sin(x_{k-1}^1) \, \Delta t \end{pmatrix}}_{\mathbf{f}(\mathbf{x_{k-1}})} + \begin{pmatrix} 0 \\ q_{k-1} \end{pmatrix}$$

$$y_k = \underbrace{\sin(x_k^1)}_{\mathbf{h}(\mathbf{x_k})} + r_k,$$

- The required Jacobian matrix for ERTSS:

$$\mathbf{F}_x(\mathbf{x}) = \begin{pmatrix} 1 & \Delta t \\ -g \, \cos(x^1) \, \Delta t & 1 \end{pmatrix}$$

- The required expected value for SLRTSS is

$$\mathrm{E}[\mathbf{f}(\mathbf{x})] = \begin{pmatrix} m_1 + m_2 \, \Delta t \\ m_2 - g \, \sin(m_1) \, \exp(-P_{11}/2) \, \Delta t \end{pmatrix}$$

- And the cross term:

$$\mathrm{E}[\mathbf{f}(\mathbf{x}) \, (\mathbf{x} - \mathbf{m})^T] = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix},$$

where

$$\begin{aligned} c_{11} &= P_{11} + \Delta t \, P_{12} \\ c_{12} &= P_{12} + \Delta t \, P_{22} \\ c_{21} &= P_{12} - g \, \Delta t \, \cos(m_1) \, P_{11} \, \exp(-P_{11}/2) \\ c_{22} &= P_{22} - g \, \Delta t \, \cos(m_1) \, P_{12} \, \exp(-P_{11}/2) \end{aligned}$$