Lecture 8: Bayesian optimal smoother, Rauch-Tung-Striebel smoothing

Simo Särkkä

March 6, 2019

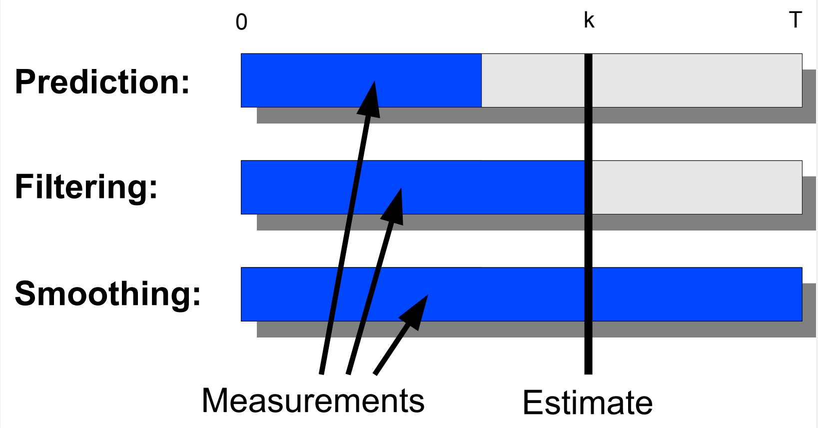
Learning Outcomes

1. Summary of the Last Lecture
2. What is Bayesian Smoothing?
3. Bayesian Smoothing Equations
4. Rauch-Tung-Striebel Smoother
5. Summary and Demonstration

Summary of the Last Lecture

- Rao–Blackwellization is a variance reduction technique that can be used to handle analytically tractable substructures
- In Rao–Blackwellized particle filters a part of the state is sampled and part is integrated in closed form with Kalman filter
- Rao–Blackwellized particle filters use a Gaussian mixture for approximating the filtering distributions
- Rao–Blackwellization may significantly reduce the number of particles required in a particle filter
- It is possible to do approximate Rao–Blackwellization by replacing the Kalman filter with a Gaussian filter

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.

Bayesian 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 (SLRTSS).
  - Gaussian Rauch-Tung-Striebel smoothers (GRTSS), cubature, Gauss-Hermite, Bayes-Hermite, Monte Carlo.
  - Two-filter versions of the above.
- Non-linear non-Gaussian models
  - Particle smoothers.
  - Rao-Blackwellized particle smoothers.
  - Grid based smoothers.

Problem Formulation

- Probabilistic state space model:

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

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

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

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

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

Derivation of Formal Smoothing Equations [1/2]

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

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

- Thus we get:

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

Derivation of Formal Smoothing Equations [2/2]

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

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

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

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

Bayesian Smoothing Equations

Bayesian Smoothing Equations

The Bayesian 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

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

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

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

RTS: Derivation Preliminaries

- Gaussian probability density

$$\mathrm{N}(\mathbf{x} \mid \mathbf{m}, \mathbf{P}) = \frac{1}{(2\,\pi)^{n/2}\,|\mathbf{P}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\,\mathbf{P}^{-1}\,(\mathbf{x}-\mathbf{m})\right),$$

- Let **x** and **y** have the Gaussian densities

$$p(\mathbf{x}) = \mathrm{N}(\mathbf{x} \mid \mathbf{m}, \mathbf{P}), \qquad p(\mathbf{y} \mid \mathbf{x}) = \mathrm{N}(\mathbf{y} \mid \mathbf{H}\,\mathbf{x}, \mathbf{R}),$$

- Then the joint and marginal distributions are

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mathbf{m} \\ \mathbf{H}\,\mathbf{m} \end{pmatrix}, \begin{pmatrix} \mathbf{P} & \mathbf{P}\,\mathbf{H}^T \\ \mathbf{H}\,\mathbf{P} & \mathbf{H}\,\mathbf{P}\,\mathbf{H}^T + \mathbf{R} \end{pmatrix} \right)$$

$$\mathbf{y} \sim \mathrm{N}(\mathbf{H}\,\mathbf{m}, \mathbf{H}\,\mathbf{P}\,\mathbf{H}^T + \mathbf{R}).$$

RTS: Derivation Preliminaries (cont.)

- If the random variables $\mathbf{x}$ and $\mathbf{y}$ have the joint Gaussian probability density

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}, \begin{pmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{C}^T & \mathbf{B} \end{pmatrix} \right),$$

- Then the marginal and conditional densities of $\mathbf{x}$ and $\mathbf{y}$ are given as follows:

$$\begin{aligned} \mathbf{x} &\sim \mathrm{N}(\mathbf{a}, \mathbf{A}) \\ \mathbf{y} &\sim \mathrm{N}(\mathbf{b}, \mathbf{B}) \\ \mathbf{x} \,|\, \mathbf{y} &\sim \mathrm{N}(\mathbf{a} + \mathbf{C}\,\mathbf{B}^{-1}\,(\mathbf{y} - \mathbf{b}), \mathbf{A} - \mathbf{C}\,\mathbf{B}^{-1}\mathbf{C}^T) \\ \mathbf{y} \,|\, \mathbf{x} &\sim \mathrm{N}(\mathbf{b} + \mathbf{C}^T\,\mathbf{A}^{-1}\,(\mathbf{x} - \mathbf{a}), \mathbf{B} - \mathbf{C}^T\,\mathbf{A}^{-1}\,\mathbf{C}). \end{aligned}$$

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 conditioning rule of Gaussian distribution we get

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

where

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

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

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

- The marginal mean and covariance are thus given as

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

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

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

Rauch-Tung-Striebel Smoother

Rauch-Tung-Striebel Smoother

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

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

- $\mathbf{m}_k$ 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 $\mathbf{P}_T^s = \mathbf{P}_T$.

Summary

- Bayesian 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.
- RTSS is fixed-interval smoother, there are also fixed-point and fixed-lag smoothers.

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 $\mathbf{Q}$:

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