Lecture 7: Rao–Blackwellized Particle Filtering

Simo Särkkä

Match 4, 2020

Learning Outcomes

1. Summary of the Last Lecture
2. Particle Filtering: Recap
3. Conditionally Linear Gaussian Models
4. Rao–Blackwellized Particle Filter
5. Summary

Summary of the Last Lecture

- Particle filters can be used for approximate filtering in general probabilistic state-space models.
- Particle filters use weighted set of samples (particles) for approximating the filtering distributions.
- Sequential importance resampling (SIR) is the general framework and bootstrap filter is a simple special case of it.
- EKF, UKF and other Gaussian filters can be used for forming good importance distributions.
- The optimal importance distribution is the minimum variance importance distribution.

Particle Filtering: General Idea

- Given the general nonlinear, non-Gaussian state space model

$$\begin{aligned} \mathbf{x}_k &\sim p(\mathbf{x}_k \mid \mathbf{x}_{k-1}) \\ \mathbf{y}_k &\sim p(\mathbf{y}_k \mid \mathbf{x}_k) \end{aligned}$$

- Particle filters approximate the filtering distribution using a weighted set of *particles* $\{(w_k^{(i)}, \mathbf{x}_k^{(i)}) \; : \; i = 1, \ldots, N\}$ such that

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

Particle Filtering: Algorithm

Sequential Importance Resampling

- Sample $\mathbf{x}_k^{(i)}$ from the importance distribution:

$$\mathbf{x}_k^{(i)} \sim \pi(\mathbf{x}_k \mid \mathbf{x}_{0:k-1}^{(i)}, \mathbf{y}_{1:k}), \qquad i = 1, \ldots, N.$$

- Calculate the weights

$$w_k^{(i)} \propto w_{k-1}^{(i)} \frac{p(\mathbf{y}_k \mid \mathbf{x}_k^{(i)}) \, p(\mathbf{x}_k^{(i)} \mid \mathbf{x}_{k-1}^{(i)})}{\pi(\mathbf{x}_k^{(i)} \mid \mathbf{x}_{0:k-1}^{(i)}, \mathbf{y}_{1:k})}, \qquad i = 1, \ldots, N,$$

  and normalize them to sum to unity.
- If the effective number of particles is too low, perform resampling.

Particle Filtering: Some Properties

- The bootstrap filter uses the dynamic model as the importance distribution

$$\pi(\mathbf{x}_k^{(i)} \mid \mathbf{x}_{0:k-1}^{(i)}, \mathbf{y}_{1:k}) = p(\mathbf{x}_k^{(i)} \mid \mathbf{x}_{k-1}^{(i)})$$

- The optimal importance distribution is given by

$$\pi(\mathbf{x}_k^{(i)} \mid \mathbf{x}_{0:k-1}^{(i)}, \mathbf{y}_{1:k}) = p(\mathbf{x}_k^{(i)} \mid \mathbf{x}_{k-1}^{(i)}, \mathbf{y}_k)$$

- The unscented particle filter uses a Gaussian approximation to the optimal importance distribution
- Particle filters can handle any kind of model and provide a global approximation and converge to the exact solution
- Higher computational requirements than Kalman filters and difficult to implement in practice for some models

Particle Filtering: Problems

- The particle filter requires a very high number of particles to work reasonably well.
- This is called to curse of dimensionality:
  - It is difficult to get the particles into the right place in high-dimensional problems (cf. finding the needle in a haystack)
  - The number of particles generally scales exponentially with the state dimension
- In Rao–Blackwellized particle filters we sample only as small number of states as we need.
- Kalman filters are used to integrate out the linear parts of the state-space.

Hierarchical Model

- Definition of a hierarchical (RBPF) model

$$
\begin{aligned}
\mathbf{u}_k &\sim p(\mathbf{u}_k \mid \mathbf{u}_{k-1}) \\
\mathbf{x}_k &= \mathbf{f}_{k-1}(\mathbf{u}_k) + \mathbf{A}_{k-1}(\mathbf{u}_k)\mathbf{x}_{k-1} + \mathbf{q}_{k-1} \\
\mathbf{y}_k &= \mathbf{h}_k(\mathbf{u}_k) + \mathbf{H}_k(\mathbf{u}_k)\mathbf{x}_k + \mathbf{r}_k
\end{aligned}
$$

  with $\mathbf{q}_{k-1} \sim \mathrm{N}(0, \mathbf{Q}_{k-1}(\mathbf{u}_k))$ and $\mathbf{r}_k \sim \mathrm{N}(0, \mathbf{R}_k(\mathbf{u}_k))$
- Transition densities:

$$
\begin{aligned}
p(\mathbf{u}_k \mid \mathbf{u}_{k-1}) &= [\text{arbitrary}] \\
p(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \mathbf{u}_k) &= \mathrm{N}(\mathbf{x}_k \mid \mathbf{f}_{k-1}(\mathbf{u}_k) + \mathbf{A}_{k-1}(\mathbf{u}_k)\mathbf{x}_{k-1}, \mathbf{Q}_{k-1}(\mathbf{u}_k))
\end{aligned}
$$

- Likelihood:

$$
p(\mathbf{y}_k \mid \mathbf{x}_k, \mathbf{u}_k) = \mathrm{N}(\mathbf{y}_k \mid \mathbf{h}_k(\mathbf{u}_k) + \mathbf{H}_k(\mathbf{u}_k)\mathbf{x}_k, \mathbf{R}_k(\mathbf{u}_k))
$$

Fully Mixing Model

- Definition of a fully mixing (RBPF) model

$$\begin{bmatrix} \mathbf{u}_k \\ \mathbf{x}_k \end{bmatrix} = \begin{bmatrix} \mathbf{g}_{k-1}(\mathbf{u}_{k-1}) \\ \mathbf{f}_{k-1}(\mathbf{u}_{k-1}) \end{bmatrix} + \begin{bmatrix} \mathbf{B}_{k-1}(\mathbf{u}_{k-1}) \\ \mathbf{A}_{k-1}(\mathbf{u}_{k-1}) \end{bmatrix} \mathbf{x}_{k-1} + \mathbf{q}_{k-1}$$

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{u}_k) + \mathbf{H}_k(\mathbf{u}_k)\mathbf{x}_k + \mathbf{r}_k$$

  with $\mathbf{q}_{k-1} \sim \mathrm{N}(0, \mathbf{Q}_{k-1}(\mathbf{u}_k))$ and $\mathbf{r}_k \sim \mathrm{N}(0, \mathbf{R}_k(\mathbf{u}_k))$
- Transition density:

$$p(\mathbf{x}_k, \mathbf{u}_k \mid \mathbf{x}_{k-1}, \mathbf{u}_{k-1})$$
$$= \mathrm{N}\left(\begin{bmatrix} \mathbf{u}_k \\ \mathbf{x}_k \end{bmatrix} \middle| \begin{bmatrix} \mathbf{g}_{k-1}(\mathbf{u}_{k-1}) \\ \mathbf{f}_{k-1}(\mathbf{u}_{k-1}) \end{bmatrix} + \begin{bmatrix} \mathbf{B}_{k-1}(\mathbf{u}_{k-1}) \\ \mathbf{A}_{k-1}(\mathbf{u}_{k-1}) \end{bmatrix} \mathbf{x}_{k-1}, \mathbf{Q}_{k-1}(\mathbf{u}_{k-1}) \right)$$

- Likelihood:

$$p(\mathbf{y}_k \mid \mathbf{x}_k, \mathbf{u}_k) = \mathrm{N}(\mathbf{y}_k \mid \mathbf{h}_k(\mathbf{u}_k) + \mathbf{H}_k(\mathbf{u}_k)\mathbf{x}_k, \mathbf{R}_k(\mathbf{u}_k))$$

Rao–Blackwellized Particle Filter: Idea

- The posterior at step k can be factorized as

$$p(\mathbf{u}_{0:k}, \mathbf{x}_k \mid \mathbf{y}_{1:k}) = p(\mathbf{x}_k \mid \mathbf{u}_{0:k}, \mathbf{y}_{1:k}) p(\mathbf{u}_{0:k} \mid \mathbf{y}_{1:k})$$

- Given $\mathbf{u}_{0:k}$, the first term has a *closed form solution* (Gaussian) and can be computed using a Kalman filter
- The second nonlinear/non-Gaussian term (marginal filtering density) is targeted using SIR
- This is the application of a variance reduction technique called Rao–Blackwellization
- This yields the posterior approximation

$$p(\mathbf{u}_k, \mathbf{x}_k \mid \mathbf{y}_{1:k}) \approx \sum_{i=1}^{N} w_k^{(i)} \, \mathrm{N}(\mathbf{x}_k \mid \mathbf{m}_k^{(i)}, \mathbf{P}_k^{(i)}) \delta(\mathbf{u}_k - \mathbf{u}_k^{(i)})$$

Nonlinear States

- For the marginal filtering density we get the recursion:

$$
\begin{aligned}
p(\mathbf{u}_{0:k} \mid \mathbf{y}_{1:k}) &\propto p(\mathbf{y}_k \mid \mathbf{u}_{0:k}, \mathbf{y}_{1:k-1}) p(\mathbf{u}_{0:k} \mid \mathbf{y}_{1:k-1}) \\
&= \underbrace{p(\mathbf{y}_k \mid \mathbf{u}_{0:k}, \mathbf{y}_{1:k-1})}_{\text{Marginal Likelihood}} \underbrace{p(\mathbf{u}_k \mid \mathbf{u}_{0:k-1}, \mathbf{y}_{1:k-1})}_{\text{Marginal Dynamics}} \underbrace{p(\mathbf{u}_{0:k-1} \mid \mathbf{y}_{1:k-1})}_{\text{Posterior at } k-1}
\end{aligned}
$$

- We can form the importance distribution recursively:

$$
\pi(\mathbf{u}_{0:k} \mid \mathbf{y}_{1:k}) = \pi(\mathbf{u}_k \mid \mathbf{u}_{0:k-1}, \mathbf{y}_{1:k}) \, \pi(\mathbf{u}_{0:k-1} \mid \mathbf{y}_{1:k-1})
$$

- We then get the following weight recursion:

$$
w_k^{(i)} \propto \frac{p(\mathbf{y}_k \mid \mathbf{u}_{0:k}^{(i)}, \mathbf{y}_{1:k-1}) p(\mathbf{u}_k^{(i)} \mid \mathbf{u}_{0:k-1}^{(i)}, \mathbf{y}_{1:k-1})}{\pi(\mathbf{u}_k^{(i)} \mid \mathbf{u}_{0:k-1}^{(i)}, \mathbf{y}_{1:k})} w_{k-1}^{(i)}
$$

Linear States

- Assume that at time $k-1$ we have that

$$p(\mathbf{x}_{k-1} \mid \mathbf{u}_{0:k-1}, \mathbf{y}_{1:k}) = \mathrm{N}(\mathbf{x}_{k-1} \mid \mathbf{m}_{k-1}, \mathbf{P}_{k-1})$$

- At time k, the posterior for $\mathbf{x}_k$ can be factorized as

$$\begin{aligned} p(\mathbf{x}_k \mid \mathbf{u}_{0:k}, \mathbf{y}_{1:k}) &\propto p(\mathbf{y}_k \mid \mathbf{x}_k, \mathbf{u}_{0:k}, \mathbf{y}_{1:k-1}) p(\mathbf{x}_k \mid \mathbf{u}_{0:k}, \mathbf{y}_{1:k-1}) \\ &= \underbrace{p(\mathbf{y}_k \mid \mathbf{x}_k, \mathbf{u}_k)}_{\text{Likelihood}} \underbrace{p(\mathbf{x}_k \mid \mathbf{u}_{0:k}, \mathbf{y}_{1:k-1})}_{\text{Prediction}} \end{aligned}$$

Hierarchical Model: Prediction step [1/3]

- Objective: Find $p(\mathbf{x}_k \mid \mathbf{u}_{0:k}, \mathbf{y}_{1:k-1})$, the predictive density of the linear states $\mathbf{x}_k$
- Dynamic model:

$$\begin{aligned} \mathbf{u}_k &\sim p(\mathbf{u}_k \mid \mathbf{u}_{k-1}) \\ p(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \mathbf{u}_k) &= \mathrm{N}(\mathbf{x}_k \mid \mathbf{f}_{k-1}(\mathbf{u}_k) + \mathbf{A}_{k-1}(\mathbf{u}_k)\mathbf{x}_{k-1}, \mathbf{Q}_{k-1}(\mathbf{u}_k)) \end{aligned}$$

- Prediction of linear states:

$$\begin{aligned} p(\mathbf{x}_k \mid \mathbf{u}_{0:k}, \mathbf{y}_{1:k-1}) &= \int p(\mathbf{x}_k, \mathbf{x}_{k-1} \mid \mathbf{u}_{0:k}, \mathbf{y}_{1:k-1}) \mathrm{d}\mathbf{x}_{k-1} \\ &\propto \int p(\mathbf{x}_k, \mathbf{u}_k \mid \mathbf{x}_{k-1}, \mathbf{u}_{0:k-1}, \mathbf{y}_{1:k-1}) p(\mathbf{x}_{k-1} \mid \mathbf{u}_{0:k-1}, \mathbf{y}_{1:k-1}) \mathrm{d}\mathbf{x}_{k-1} \\ &= \int p(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \mathbf{u}_k) p(\mathbf{u}_k \mid \mathbf{u}_{k-1}) p(\mathbf{x}_{k-1} \mid \mathbf{u}_{0:k-1}, \mathbf{y}_{1:k-1}) \mathrm{d}\mathbf{x}_{k-1} \\ &\propto \int p(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \mathbf{u}_k) p(\mathbf{x}_{k-1} \mid \mathbf{u}_{0:k-1}, \mathbf{y}_{1:k-1}) \mathrm{d}\mathbf{x}_{k-1} \end{aligned}$$

Hierarchical Model: Prediction step [2/3]

- Prediction of the linear states $\mathbf{x}_k$:

$$
\begin{aligned}
& p(\mathbf{x}_k \mid \mathbf{u}_{0:k}, \mathbf{y}_{1:k-1}) \\
& \quad \propto \int p(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \mathbf{u}_k) p(\mathbf{x}_{k-1} \mid \mathbf{u}_{0:k-1}, \mathbf{y}_{1:k-1}) \mathrm{d}\mathbf{x}_{k-1} \\
& \quad = \mathrm{N}(\mathbf{x}_k \mid \mathbf{m}_k^-, \mathbf{P}_k^-)
\end{aligned}
$$

where

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

Hierarchical Model: Prediction step [3/3]

RBPF Prediction step: Hierarchical Model

For each particle $\mathbf{u}_k^{(i)}$ ($i = 1, \ldots, N$):

- Sample $\mathbf{u}_k^{(i)}$
$$\mathbf{u}_k^{(i)} \sim \pi(\mathbf{u}_k \mid \mathbf{u}_{0:k-1}^{(i)}, \mathbf{y}_{1:k}).$$
- Predict the means $\mathbf{m}_k^{-(i)}$ and covariances $\mathbf{P}_k^{-(i)}$
$$\begin{aligned}
\mathbf{m}_k^{-(i)} &= \mathbf{f}_{k-1}(\mathbf{u}_k^{(i)}) + \mathbf{A}_{k-1}(\mathbf{u}_k^{(i)}) \mathbf{m}_{k-1}^{(i)}, \\
\mathbf{P}_k^{-(i)} &= \mathbf{A}_{k-1}(\mathbf{u}_k^{(i)}) \mathbf{P}_{k-1}^{(i)} \mathbf{A}_{k-1}(\mathbf{u}_k^{(i)})^\mathsf{T} + \mathbf{Q}_{k-1}(\mathbf{u}_k^{(i)})
\end{aligned}$$

Mixing Model: Prediction step [1/3]

- Objective: Find $p(\mathbf{x}_k \mid \mathbf{u}_{0:k}, \mathbf{y}_{1:k-1})$
- Dynamic model:

$$
\begin{aligned}
& p(\mathbf{x}_k, \mathbf{u}_k \mid \mathbf{x}_{k-1}, \mathbf{u}_{k-1}) \\
& = \mathrm{N}\left(\begin{bmatrix} \mathbf{u}_k \\ \mathbf{x}_k \end{bmatrix} \,\middle|\, \begin{bmatrix} \mathbf{g}_{k-1}(\mathbf{u}_{k-1}) \\ \mathbf{f}_{k-1}(\mathbf{u}_{k-1}) \end{bmatrix} + \begin{bmatrix} \mathbf{B}_{k-1}(\mathbf{u}_{k-1}) \\ \mathbf{A}_{k-1}(\mathbf{u}_{k-1}) \end{bmatrix} \mathbf{x}_{k-1}, \mathbf{Q}_{k-1}(\mathbf{u}_{k-1}) \right)
\end{aligned}
$$

- First note that:

$$
\begin{aligned}
& p(\mathbf{x}_k, \mathbf{u}_k \mid \mathbf{u}_{0:k-1}, \mathbf{y}_{1:k-1}) \\
& = \int p(\mathbf{x}_k, \mathbf{u}_k, \mathbf{x}_{k-1} \mid \mathbf{u}_{0:k-1}, \mathbf{y}_{1:k-1}) \mathrm{d}\mathbf{x}_{k-1} \\
& = \int p(\mathbf{x}_k, \mathbf{u}_k \mid \mathbf{x}_{k-1}, \mathbf{u}_{k-1}) p(\mathbf{x}_{k-1} \mid \mathbf{u}_{0:k-1}, \mathbf{y}_{1:k-1}) \mathrm{d}\mathbf{x}_{k-1} \\
& = \mathrm{N}\left(\begin{bmatrix} \mathbf{u}_k \\ \mathbf{x}_k \end{bmatrix} \,\middle|\, \begin{bmatrix} \mathbf{g} \\ \mathbf{f} \end{bmatrix} + \begin{bmatrix} \mathbf{B} \\ \mathbf{A} \end{bmatrix} \mathbf{m}_{k-1}, \begin{bmatrix} \mathbf{B} \\ \mathbf{A} \end{bmatrix} \mathbf{P}_{k-1} \begin{bmatrix} \mathbf{B} \\ \mathbf{A} \end{bmatrix}^\mathsf{T} + \begin{bmatrix} \mathbf{Q}^{\mathbf{u}} & \mathbf{Q}^{\mathbf{ux}} \\ \mathbf{Q}^{\mathbf{xu}} & \mathbf{Q}^{\mathbf{x}} \end{bmatrix} \right)
\end{aligned}
$$

Mixing Model: Prediction step [2/3]

- Conditioning on $\mathbf{u}_k$ yields the prediction of the linear states

$$p(\mathbf{x}_k \mid \mathbf{u}_{0:k}, \mathbf{y}_{1:k-1}) = \mathrm{N}(\mathbf{x}_k \mid \mathbf{m}_k^-, \mathbf{P}_k^-)$$

with

$$\begin{aligned}
\mathbf{M}_k &= \mathbf{B}_{k-1}(\mathbf{u}_{k-1}) \mathbf{P}_{k-1} \mathbf{B}_{k-1}(\mathbf{u}_{k-1})^\mathsf{T} + \mathbf{Q}^{\mathbf{u}}_{k-1}(\mathbf{u}_{k-1}) \\
\mathbf{L}_k &= \left(\mathbf{A}_{k-1}(\mathbf{u}_{k-1}) \mathbf{P}_{k-1} \mathbf{B}_{k-1}(\mathbf{u}_{k-1})^\mathsf{T} + \mathbf{Q}^{\mathbf{xu}}_{k-1}(\mathbf{u}_{k-1}) \right) \mathbf{M}_k^{-1} \\
\mathbf{m}_k^- &= \mathbf{f}_{k-1}(\mathbf{u}_{k-1}) + \mathbf{A}_{k-1}(\mathbf{u}_{k-1}) \mathbf{m}_{k-1} \\
&\qquad + \mathbf{L}_k \left(\mathbf{u}_k - \mathbf{g}_{k-1}(\mathbf{u}_{k-1}) - \mathbf{B}_{k-1}(\mathbf{u}_{k-1}) \mathbf{m}_{k-1} \right) \\
\mathbf{P}_k^- &= \mathbf{A}_{k-1}(\mathbf{u}_{k-1}) \mathbf{P}_{k-1} \mathbf{A}_{k-1}(\mathbf{u}_{k-1})^\mathsf{T} + \mathbf{Q}^{\mathbf{x}}_{k-1}(\mathbf{u}_{k-1}) - \mathbf{L}_k \mathbf{M}_k \mathbf{L}_k^\mathsf{T}
\end{aligned}$$

- This can be seen as a measurement update for the linear states using the nonlinear states

RBPF Prediction step: Mixing Model

For each particle $\mathbf{u}_k^{(i)}$ $(i = 1, \ldots, N)$:

- Sample $\mathbf{u}_k^{(i)}$:

$$\mathbf{u}_k^{(i)} \sim \pi(\mathbf{u}_k \mid \mathbf{u}_{0:k-1}^{(i)}, \mathbf{y}_{1:k}).$$

- Predict the means $\mathbf{m}_k^{-(i)}$ and covariances $\mathbf{P}_k^{-(i)}$:

$$\begin{aligned}
\mathbf{M}_k^{(i)} &= \mathbf{B}_{k-1}^{(i)} \mathbf{P}_{k-1}^{(i)} (\mathbf{B}_{k-1}^{(i)})^{\mathsf{T}} + \mathbf{Q}_{k-1}^{\mathbf{u}(i)} \\
\mathbf{L}_k^{(i)} &= \left(\mathbf{A}_{k-1}^{(i)} \mathbf{P}_{k-1}^{(i)} (\mathbf{B}_{k-1}^{(i)})^{\mathsf{T}} + \mathbf{Q}_{k-1}^{\mathbf{xu}(i)} \right) (\mathbf{M}_k^{(i)})^{-1} \\
\mathbf{m}_k^{-(i)} &= \mathbf{f}_{k-1}^{(i)} + \mathbf{A}_{k-1}^{(i)} \mathbf{m}_{k-1}^{(i)} + \mathbf{L}_k^{(i)} \left(\mathbf{u}_k^{(i)} - \mathbf{g}_{k-1}^{(i)} - \mathbf{B}_{k-1}^{(i)} \mathbf{m}_{k-1}^{(i)} \right) \\
\mathbf{P}_k^{-(i)} &= \mathbf{A}_{k-1}^{(i)} \mathbf{P}_{k-1}^{(i)} (\mathbf{A}_{k-1}^{(i)})^{\mathsf{T}} + \mathbf{Q}_{k-1}^{\mathbf{x}(i)} - \mathbf{L}_k^{(i)} \mathbf{M}_k^{(i)} (\mathbf{L}_k^{(i)})^{\mathsf{T}}
\end{aligned}$$

Measurement Update [1/2]

- Likelihood model (both models):

$$p(\mathbf{y}_k \mid \mathbf{x}_k, \mathbf{u}_k) = \mathrm{N}(\mathbf{y}_k \mid \mathbf{h}_k(\mathbf{u}_k) + \mathbf{H}_k(\mathbf{u}_k)\mathbf{x}_k, \mathbf{R}_k(\mathbf{u}_k))$$

- Measurement update for the linear states $\mathbf{x}_k$:

$$\begin{aligned} & p(\mathbf{x}_k \mid \mathbf{u}_{0:k}, \mathbf{y}_{1:k}) \propto p(\mathbf{y}_k \mid \mathbf{x}_k, \mathbf{u}_k) p(\mathbf{x}_k \mid \mathbf{u}_{0:k}, \mathbf{y}_{1:k-1}) \\ & = \mathrm{N}(\mathbf{y}_k \mid \mathbf{h}_k(\mathbf{u}_k) + \mathbf{H}_k(\mathbf{u}_k)\mathbf{x}_k, \mathbf{R}_k(\mathbf{u}_k)) \, \mathrm{N}(\mathbf{x}_k \mid \mathbf{m}_k^-, \mathbf{P}_k^-) \\ & \propto \mathrm{N}(\mathbf{x}_k \mid \mathbf{m}_k, \mathbf{P}_k) \end{aligned}$$

where

$$\begin{aligned} \mathbf{y}_k^- &= \mathbf{h}_k(\mathbf{u}_k) + \mathbf{H}_k(\mathbf{u}_k)\mathbf{m}_k^-, \\ \mathbf{S}_k &= \mathbf{H}_k(\mathbf{u}_k)\mathbf{P}_k^-\mathbf{H}_k(\mathbf{u}_k)^\mathsf{T} + \mathbf{R}_k(\mathbf{u}_k), \\ \mathbf{K}_k &= \mathbf{P}_k^-\mathbf{H}_k(\mathbf{u}_k)^\mathsf{T}\mathbf{S}_k^{-1}, \\ \mathbf{m}_k &= \mathbf{m}_k^- + \mathbf{K}_k(\mathbf{y}_k - \mathbf{y}_k^-), \\ \mathbf{P}_k &= \mathbf{P}_k^- - \mathbf{K}_k\mathbf{S}_k\mathbf{K}_k^\mathsf{T} \end{aligned}$$

RBPF Measurement Update

For each particle $\mathbf{u}_k^{(i)}$ $(i = 1, \ldots, N)$:

- Update the means $\mathbf{m}_k^{(i)}$ and covariances $\mathbf{P}_k^{(i)}$:

$$
\begin{aligned}
\mathbf{y}_k^{-(i)} &= \mathbf{h}_k^{(i)} + \mathbf{H}_k^{(i)} \mathbf{m}_k^{-(i)}, \\
\mathbf{S}_k^{(i)} &= \mathbf{H}_k^{(i)} \mathbf{P}_k^{-(i)} (\mathbf{H}_k^{(i)})^\mathsf{T} + \mathbf{R}_k^{(i)}, \\
\mathbf{K}_k^{(i)} &= \mathbf{P}_k^{-(i)} (\mathbf{H}_k^{(i)})^\mathsf{T} (\mathbf{S}_k^{(i)})^{-1}, \\
\mathbf{m}_k^{(i)} &= \mathbf{m}_k^{-(i)} + \mathbf{K}_k^{(i)} (\mathbf{y}_k - \mathbf{y}_k^{-(i)}), \\
\mathbf{P}_k^{(i)} &= \mathbf{P}_k^{-(i)} - \mathbf{K}_k^{(i)} \mathbf{S}_k^{(i)} (\mathbf{K}_k^{(i)})^\mathsf{T}
\end{aligned}
$$

Importance Weights: Marginal Dynamics

- Recall that:

$$w_k^{(i)} \propto \frac{p(\mathbf{y}_k \mid \mathbf{u}_{0:k}^{(i)}, \mathbf{y}_{1:k-1}) p(\mathbf{u}_k^{(i)} \mid \mathbf{u}_{0:k-1}^{(i)}, \mathbf{y}_{1:k-1})}{\pi(\mathbf{u}_k^{(i)} \mid \mathbf{u}_{0:k-1}^{(i)}, \mathbf{y}_{1:k})} w_{k-1}^{(i)}$$

- Objective: Find $p(\mathbf{u}_k \mid \mathbf{u}_{0:k-1}, \mathbf{y}_{1:k-1})$ (marginal dynamics) and the $p(\mathbf{y}_k \mid \mathbf{u}_{0:k}, \mathbf{y}_{1:k-1})$ (marginal likelihood)
- Hierarchical model: Given $\mathbf{u}_{k-1}$, the marginal dynamics are independent of $\mathbf{u}_{0:k-2}$ and $\mathbf{y}_{1:k-1}$ (independent of $\mathbf{x}_{k-1}$), hence

$$p(\mathbf{u}_k \mid \mathbf{u}_{0:k-1}, \mathbf{y}_{1:k-1}) = p(\mathbf{u}_k \mid \mathbf{u}_{k-1})$$

- Mixing model: Marginalizing $p(\mathbf{x}_k, \mathbf{u}_k \mid \mathbf{u}_{0:k-1}, \mathbf{y}_{1:k-1})$ (see above) with respect to $\mathbf{x}_k$ yields

$$p(\mathbf{u}_k \mid \mathbf{u}_{0:k-1}, \mathbf{y}_{1:k-1}) = \mathrm{N}(\mathbf{u}_k \mid \mathbf{g} + \mathbf{B}\mathbf{m}_{k-1}, \mathbf{B}\mathbf{P}_{k-1}\mathbf{B}^{\mathsf{T}} + \mathbf{Q}^{\mathbf{u}})$$

Importance Weights: Marginal Likelihood

- For both models, the likelihood is

$$p(\mathbf{y}_k \mid \mathbf{x}_k, \mathbf{u}_k) = \mathrm{N}(\mathbf{y}_k \mid \mathbf{h}_k(\mathbf{u}_k) + \mathbf{H}_k(\mathbf{u}_k)\mathbf{x}_k, \mathbf{R}_k(\mathbf{u}_k))$$

- Marginal likelihood

$$\begin{aligned}
&p(\mathbf{y}_k \mid \mathbf{u}_{0:k}, \mathbf{y}_{1:k-1}) \\
&= \int p(\mathbf{y}_k \mid \mathbf{x}_k, \mathbf{u}_k) p(\mathbf{x}_k \mid \mathbf{u}_{0:k}, \mathbf{y}_{1:k-1}) \mathrm{d}\mathbf{x}_k \\
&= \int \mathrm{N}(\mathbf{y}_k \mid \mathbf{h}_k(\mathbf{u}_k) + \mathbf{H}_k(\mathbf{u}_k)\mathbf{x}_k, \mathbf{R}_k(\mathbf{u}_k)) \, \mathrm{N}(\mathbf{x}_k \mid \mathbf{m}_k^-, \mathbf{P}_k^-) \mathrm{d}\mathbf{x}_k \\
&= \mathrm{N}(\mathbf{y}_k \mid \mathbf{y}_k^-, \mathbf{S}_k)
\end{aligned}$$

with $\mathbf{y}_k^-$ and $\mathbf{S}_k$ as in the measurement update

Rao–Blackwellized Particle Filter: Algorithm

Rao–Blackwellized Particle Filter

For each particle $\mathbf{u}_k^{(i)}$ ($i = 1, \ldots, N$):

- Sample $\mathbf{u}_k^{(i)}$:

$$\mathbf{u}_k^{(i)} \sim \pi(\mathbf{u}_k \mid \mathbf{u}_{0:k-1}^{(i)}, \mathbf{y}_{1:k}).$$

- Predict the means $\mathbf{m}_k^{-(i)}$ and covariances $\mathbf{P}_k^{-(i)}$
- Update the means $\mathbf{m}_k^{(i)}$ and covariances $\mathbf{P}_k^{(i)}$
- Calculate and normalize the weights $w_k^{(i)}$

$$w_k^{(i)} \propto w_{k-1}^{(i)} \frac{p(\mathbf{y}_k \mid \mathbf{u}_{0:k}^{(i)}, \mathbf{y}_{1:k-1}) p(\mathbf{u}_k^{(i)} \mid \mathbf{u}_{0:k-1}^{(i)}, \mathbf{y}_{1:k-1})}{\pi(\mathbf{u}_k^{(i)} \mid \mathbf{u}_{0:k-1}^{(i)}, \mathbf{y}_{1:k})}$$

- If the effective number of particles is too low, *resample*.

Rao–Blackwellized Particle Filter: Properties [1/2]

- The optimal importance distribution is given by

$$p(\mathbf{u}_k \mid \mathbf{y}_{1:k}, \mathbf{u}^{(i)}_{0:k-1}) \propto p(\mathbf{y}_k \mid \mathbf{u}_k, \mathbf{u}^{(i)}_{0:k-1})\, p(\mathbf{u}_k \mid \mathbf{u}^{(i)}_{0:k-1}, \mathbf{y}_{1:k-1})$$

- The Rao–Blackwellized Bootstrap Particle Filter samples from the marginal dynamics, that is, from

$$\pi(\mathbf{u}_k \mid \mathbf{u}_{0:k-1}, \mathbf{y}_{1:k}) = p(\mathbf{u}_k \mid \mathbf{u}_{0:k-1}, \mathbf{y}_{1:k-1})$$

- During resampling, the means $\mathbf{m}^{(i)}_k$ and covariances $\mathbf{P}^{(i)}_k$ must be resampled too
- Special cases of the models may simplify the updates for the linear states $\mathbf{x}_k$ (e.g. only one covariance matrix $\mathbf{P}_k$ for all particles)

Rao–Blackwellized Particle Filter: Properties [2/2]

- The Rao–Blackwellized particle filter produces a set of weighted samples $\{w_k^{(i)}, \mathbf{u}_k^{(i)}, \mathbf{m}_k^{(i)}, \mathbf{P}_k^{(i)} \; : \; i = 1, \ldots, N\}$
- The expectation of a function $\mathbf{g}(\cdot)$ can be approximated as

$$\mathrm{E}[\mathbf{g}(\mathbf{x}_k, \mathbf{u}_k) \mid \mathbf{y}_{1:k}] \approx \sum_{i=1}^{N} w_k^{(i)} \int \mathbf{g}(\mathbf{x}_k, \mathbf{u}_k^{(i)}) \; \mathrm{N}(\mathbf{x}_k \mid \mathbf{m}_k^{(i)}, \mathbf{P}_k^{(i)}) \, \mathrm{d}\mathbf{x}_k.$$

- Approximation of the filtering distribution is

$$p(\mathbf{x}_k, \mathbf{u}_k \mid \mathbf{y}_{1:k}) \approx \sum_{i=1}^{N} w_k^{(i)} \, \delta(\mathbf{u}_k - \mathbf{u}_k^{(i)}) \; \mathrm{N}(\mathbf{x}_k \mid \mathbf{m}_k^{(i)}, \mathbf{P}_k^{(i)}).$$

Summary

- 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