On Sequential Monte Carlo Sampling of Discretely Observed Stochastic Differential Equations

Simo Särkkä <simo.sarkka@hut.fi>

Laboratory of Computational Engineering
Helsinki University of Technology
Finland

September 15, 2006

Contents

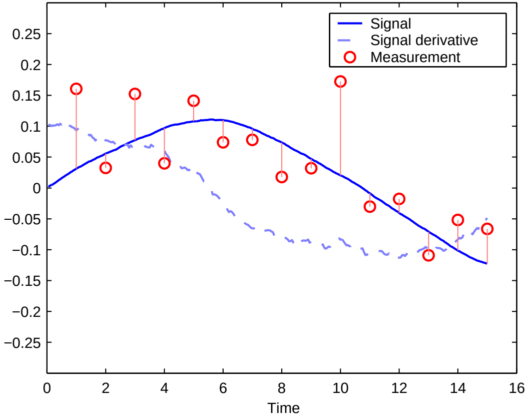
Contents

Contents

Contents

Continuous-Discrete Filtering Problem

- Estimate the unobserved continuous-time signal from noisy discrete-time measurements



Mathematical Problem Formulation

- The dynamics of state $\mathbf{x}(t)$ modeled as a stochastic differential equation (diffusion process)

$$\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\,\mathrm{d}t + \mathbf{L}\,\mathrm{d}\beta(t).$$

- Measurements $\mathbf{y}_k$ are obtained at discrete times

$$\mathbf{y}_k \sim p(\mathbf{y}_k \,|\, \mathbf{x}(t_k)).$$

- Formal solution: Compute the posterior distribution(s)

$$p(\mathbf{x}(t) \,|\, \mathbf{y}_1, \ldots, \mathbf{y}_k), \qquad t \geq t_k.$$

Mathematical Problem Formulation

- The dynamics of state $\mathbf{x}(t)$ modeled as a stochastic differential equation (diffusion process)

$$\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\,\mathrm{d}t + \mathbf{L}\,\mathrm{d}\boldsymbol{\beta}(t).$$

- Measurements $\mathbf{y}_k$ are obtained at discrete times

$$\mathbf{y}_k \sim p(\mathbf{y}_k \,|\, \mathbf{x}(t_k)).$$

- Formal solution: Compute the posterior distribution(s)

$$p(\mathbf{x}(t) \,|\, \mathbf{y}_1, \ldots, \mathbf{y}_k), \qquad t \geq t_k.$$

Mathematical Problem Formulation

- The dynamics of state $\mathbf{x}(t)$ modeled as a stochastic differential equation (diffusion process)

$$\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\,\mathrm{d}t + \mathbf{L}\,\mathrm{d}\beta(t).$$

- Measurements $\mathbf{y}_k$ are obtained at discrete times

$$\mathbf{y}_k \sim p(\mathbf{y}_k \,|\, \mathbf{x}(t_k)).$$

- Formal solution: Compute the posterior distribution(s)

$$p(\mathbf{x}(t) \,|\, \mathbf{y}_1, \ldots, \mathbf{y}_k), \qquad t \geq t_k.$$

Formal solution

Optimal filter

1. Prediction step: Solve the Kolmogorov-forward (Fokker-Planck) partial differential equation.

$$\frac{\partial p}{\partial t} = -\sum_i \frac{\partial}{\partial x_i} (f_i(\mathbf{x}, t)\, p) + \frac{1}{2} \sum_{ij} \frac{\partial^2}{\partial x_i \partial x_j} \left([\mathbf{L}\,\mathbf{Q}\,\mathbf{L}^T]_{ij}\, p \right)$$

2. Update step: Apply the Bayes' rule.

$$p(\mathbf{x}(t_k) \,|\, \mathbf{y}_{1:k}) = \frac{p(\mathbf{y}_k \,|\, \mathbf{x}(t_k))\, p(\mathbf{x}(t_k) \,|\, \mathbf{y}_{1:k-1})}{\int p(\mathbf{y}_k \,|\, \mathbf{x}(t_k))\, p(\mathbf{x}(t_k) \,|\, \mathbf{y}_{1:k-1})\, \mathrm{d}\mathbf{x}(t_k)}$$

Formal solution

Optimal filter

1. Prediction step: Solve the Kolmogorov-forward (Fokker-Planck) partial differential equation.

$$\frac{\partial p}{\partial t} = -\sum_i \frac{\partial}{\partial x_i} (f_i(\mathbf{x}, t)\, p) + \frac{1}{2} \sum_{ij} \frac{\partial^2}{\partial x_i \partial x_j} \left([\mathbf{L}\,\mathbf{Q}\,\mathbf{L}^T]_{ij}\, p \right)$$

2. Update step: Apply the Bayes' rule.

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

Numerical Approximations

- Gaussian models and approximations, extended Kalman filters and unscented Kalman filters.
- FEM and finite difference based approximations to the Kolmogorov forward PDE.
- Bootstrap filter: Simulates trajectories from the SDE.
- Sequential importance resampling: *Not applicable!*

Numerical Approximations

- Gaussian models and approximations, extended Kalman filters and unscented Kalman filters.
- FEM and finite difference based approximations to the Kolmogorov forward PDE.
- Bootstrap filter: Simulates trajectories from the SDE.
- Sequential importance resampling: *Not applicable!*

Numerical Approximations

- Gaussian models and approximations, extended Kalman filters and unscented Kalman filters.
- FEM and finite difference based approximations to the Kolmogorov forward PDE.
- Bootstrap filter: Simulates trajectories from the SDE.
- Sequential importance resampling: *Not applicable!*

Numerical Approximations

- Gaussian models and approximations, extended Kalman filters and unscented Kalman filters.
- FEM and finite difference based approximations to the Kolmogorov forward PDE.
- Bootstrap filter: Simulates trajectories from the SDE.
- Sequential importance resampling: *Not applicable!*

Sequential Importance Resampling

Sequential Importance Resampling

1. Draw a random sample from the importance distribution

$$\mathbf{x}^{(i)}(t_k) \sim q(\mathbf{x}^{(i)}(t_k) \,|\, \mathbf{x}^{(i)}(t_{k-1})) \tag{1}$$

2. Evaluate the importance weight

$$w_k^{(i)} \propto \frac{p(\mathbf{y}_k \,|\, \mathbf{x}^{(i)}(t_k))\, p(\mathbf{x}^{(i)}(t_k) \,|\, \mathbf{x}^{(i)}(t_{k-1}))}{q(\mathbf{x}^{(i)}(t_k) \,|\, \mathbf{x}^{(i)}(t_{k-1}))}$$

3. Do resampling if needed.

Sequential Importance Resampling

Sequential Importance Resampling

1. Draw a random sample from the importance distribution

$$\mathbf{x}^{(i)}(t_k) \sim q(\mathbf{x}^{(i)}(t_k) \,|\, \mathbf{x}^{(i)}(t_{k-1})) \tag{1}$$

2. Evaluate the importance weight

$$w_k^{(i)} \propto \frac{p(\mathbf{y}_k \,|\, \mathbf{x}^{(i)}(t_k))\, p(\mathbf{x}^{(i)}(t_k) \,|\, \mathbf{x}^{(i)}(t_{k-1}))}{q(\mathbf{x}^{(i)}(t_k) \,|\, \mathbf{x}^{(i)}(t_{k-1}))}$$

3. Do resampling if needed.

Sequential Importance Resampling

Sequential Importance Resampling

1. Draw a random sample from the importance distribution

$$\mathbf{x}^{(i)}(t_k) \sim q(\mathbf{x}^{(i)}(t_k) \,|\, \mathbf{x}^{(i)}(t_{k-1})) \tag{1}$$

2. Evaluate the importance weight

$$w_k^{(i)} \propto \frac{p(\mathbf{y}_k \,|\, \mathbf{x}^{(i)}(t_k))\, p(\mathbf{x}^{(i)}(t_k) \,|\, \mathbf{x}^{(i)}(t_{k-1}))}{q(\mathbf{x}^{(i)}(t_k) \,|\, \mathbf{x}^{(i)}(t_{k-1}))}$$

3. Do resampling if needed.

The Problem of SIR Weight Evaluation

- The weight evaluation of SIR is of the form

$$w_k^{(i)} \propto \frac{p(\mathbf{y}_k \,|\, \mathbf{x}^{(i)}(t_k))\, p(\mathbf{x}^{(i)}(t_k) \,|\, \mathbf{x}^{(i)}(t_{k-1}))}{q(\mathbf{x}^{(i)}(t_k) \,|\, \mathbf{x}^{(i)}(t_{k-1}))}$$

- But $p(\mathbf{x}(t_k) \,|\, \mathbf{x}(t_{k-1}))$ is the solution of an arbitrary second order partial differential equation and cannot be solved.
- Actually we only need the likelihood ratio

$$\frac{p(\mathbf{x}(t_k) \,|\, \mathbf{x}(t_{k-1}))}{q(\mathbf{x}(t_k) \,|\, \mathbf{x}(t_{k-1}))}$$

- This can be computed with the Girsanov theorem without solving the PDE.

The Problem of SIR Weight Evaluation

- The weight evaluation of SIR is of the form

$$w_k^{(i)} \propto \frac{p(\mathbf{y}_k \,|\, \mathbf{x}^{(i)}(t_k))\, p(\mathbf{x}^{(i)}(t_k) \,|\, \mathbf{x}^{(i)}(t_{k-1}))}{q(\mathbf{x}^{(i)}(t_k) \,|\, \mathbf{x}^{(i)}(t_{k-1}))}$$

- But $p(\mathbf{x}(t_k) \,|\, \mathbf{x}(t_{k-1}))$ is the solution of an arbitrary second order partial differential equation and cannot be solved.
- Actually we only need the likelihood ratio

$$\frac{p(\mathbf{x}(t_k) \,|\, \mathbf{x}(t_{k-1}))}{q(\mathbf{x}(t_k) \,|\, \mathbf{x}(t_{k-1}))}$$

- This can be computed with the Girsanov theorem without solving the PDE.

The Problem of SIR Weight Evaluation

- The weight evaluation of SIR is of the form

$$w_k^{(i)} \propto \frac{p(\mathbf{y}_k \mid \mathbf{x}^{(i)}(t_k)) \, p(\mathbf{x}^{(i)}(t_k) \mid \mathbf{x}^{(i)}(t_{k-1}))}{q(\mathbf{x}^{(i)}(t_k) \mid \mathbf{x}^{(i)}(t_{k-1}))}$$

- But $p(\mathbf{x}(t_k) \mid \mathbf{x}(t_{k-1}))$ is the solution of an arbitrary second order partial differential equation and cannot be solved.
- Actually we only need the likelihood ratio

$$\frac{p(\mathbf{x}(t_k) \mid \mathbf{x}(t_{k-1}))}{q(\mathbf{x}(t_k) \mid \mathbf{x}(t_{k-1}))}$$

- This can be computed with the Girsanov theorem without solving the PDE.

The Problem of SIR Weight Evaluation

- The weight evaluation of SIR is of the form

$$w_k^{(i)} \propto \frac{p(\mathbf{y}_k \mid \mathbf{x}^{(i)}(t_k))\, p(\mathbf{x}^{(i)}(t_k) \mid \mathbf{x}^{(i)}(t_{k-1}))}{q(\mathbf{x}^{(i)}(t_k) \mid \mathbf{x}^{(i)}(t_{k-1}))}$$

- But $p(\mathbf{x}(t_k) \mid \mathbf{x}(t_{k-1}))$ is the solution of an arbitrary second order partial differential equation and cannot be solved.
- Actually we only need the likelihood ratio

$$\frac{p(\mathbf{x}(t_k) \mid \mathbf{x}(t_{k-1}))}{q(\mathbf{x}(t_k) \mid \mathbf{x}(t_{k-1}))}$$

- This can be computed with the Girsanov theorem without solving the PDE.

Girsanov Theorem

- Let $\theta(t)$ be a stochastic process, which is driven by ("adapted to") a Brownian motion $\beta(t)$.
- The likelihood ratio between $\theta(t)$ and $\beta(t)$ is:

$$\frac{\mathrm{d}P_\theta}{\mathrm{d}P_\beta} = \exp\left(\int_0^t \theta^T(t)\,\mathrm{d}\beta(t) - \frac{1}{2}\int_0^t ||\theta(t)||^2\,\mathrm{d}t\right).$$

- The likelihood ratio can be exactly computed by above stochastic integral.
- Efficient simulation based numerical solutions possible.

Girsanov Theorem

- Let $\theta(t)$ be a stochastic process, which is driven by ("adapted to") a Brownian motion $\beta(t)$.
- The likelihood ratio between $\theta(t)$ and $\beta(t)$ is:

$$\frac{\mathrm{d}P_\theta}{\mathrm{d}P_\beta} = \exp\left(\int_0^t \theta^T(t)\,\mathrm{d}\beta(t) - \frac{1}{2}\int_0^t ||\theta(t)||^2\,\mathrm{d}t \right).$$

- The likelihood ratio can be exactly computed by above stochastic integral.
- Efficient simulation based numerical solutions possible.

Girsanov Theorem

- Let $\theta(t)$ be a stochastic process, which is driven by ("adapted to") a Brownian motion $\beta(t)$.
- The likelihood ratio between $\theta(t)$ and $\beta(t)$ is:

$$\frac{\mathrm{d}P_\theta}{\mathrm{d}P_\beta} = \exp\left(\int_0^t \theta^T(t)\,\mathrm{d}\beta(t) - \frac{1}{2}\int_0^t ||\theta(t)||^2\,\mathrm{d}t \right).$$

- The likelihood ratio can be exactly computed by above stochastic integral.
- Efficient simulation based numerical solutions possible.

Girsanov Theorem

- Let $\theta(t)$ be a stochastic process, which is driven by ("adapted to") a Brownian motion $\beta(t)$.
- The likelihood ratio between $\theta(t)$ and $\beta(t)$ is:

$$\frac{\mathrm{d}P_{\theta}}{\mathrm{d}P_{\beta}} = \exp\left(\int_0^t \theta^T(t)\,\mathrm{d}\beta(t) - \frac{1}{2}\int_0^t ||\theta(t)||^2\,\mathrm{d}t\right).$$

- The likelihood ratio can be exactly computed by above stochastic integral.
- Efficient simulation based numerical solutions possible.

Evaluating the Likelihood Ratio

- With Girsanov theorem, we can derive expression for likelihood ratio of two SDEs:

$$\begin{aligned} \mathrm{d}\mathbf{x} &= \mathbf{f}(\mathbf{x}, t)\,\mathrm{d}t + \mathbf{L}\,\mathrm{d}\beta \\ \mathrm{d}\mathbf{s} &= \mathbf{g}(\mathbf{s}, t)\,\mathrm{d}t + \mathbf{B}\,\mathrm{d}\beta. \end{aligned}$$

- Process $\mathbf{s}(t)$ can be the importance process for estimated process $\mathbf{x}(t)$.
- It is a *stochastic integral*: Well known numerical methods for SDEs can be used.
- It is a *Monte Carlo solution*: Solution converges to the exact solution.

Evaluating the Likelihood Ratio

- With Girsanov theorem, we can derive expression for likelihood ratio of two SDEs:

$$\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\,\mathrm{d}t + \mathbf{L}\,\mathrm{d}\beta$$
$$\mathrm{d}\mathbf{s} = \mathbf{g}(\mathbf{s}, t)\,\mathrm{d}t + \mathbf{B}\,\mathrm{d}\beta.$$

- Process $\mathbf{s}(t)$ can be the importance process for estimated process $\mathbf{x}(t)$.
- It is a *stochastic integral*: Well known numerical methods for SDEs can be used.
- It is a *Monte Carlo solution*: Solution converges to the exact solution.

Evaluating the Likelihood Ratio

- With Girsanov theorem, we can derive expression for likelihood ratio of two SDEs:

$$
\begin{aligned}
\mathrm{d}\mathbf{x} &= \mathbf{f}(\mathbf{x}, t)\,\mathrm{d}t + \mathbf{L}\,\mathrm{d}\boldsymbol{\beta} \\
\mathrm{d}\mathbf{s} &= \mathbf{g}(\mathbf{s}, t)\,\mathrm{d}t + \mathbf{B}\,\mathrm{d}\boldsymbol{\beta}.
\end{aligned}
$$

- Process $\mathbf{s}(t)$ can be the importance process for estimated process $\mathbf{x}(t)$.
- It is a *stochastic integral*: Well known numerical methods for SDEs can be used.
- It is a *Monte Carlo solution*: Solution converges to the exact solution.

Evaluating the Likelihood Ratio

- With Girsanov theorem, we can derive expression for likelihood ratio of two SDEs:

$$\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\,\mathrm{d}t + \mathbf{L}\,\mathrm{d}\beta$$
$$\mathrm{d}\mathbf{s} = \mathbf{g}(\mathbf{s}, t)\,\mathrm{d}t + \mathbf{B}\,\mathrm{d}\beta.$$

- Process $\mathbf{s}(t)$ can be the importance process for estimated process $\mathbf{x}(t)$.
- It is a *stochastic integral*: Well known numerical methods for SDEs can be used.
- It is a *Monte Carlo solution*: Solution converges to the exact solution.

Continuous-Discrete SIR Algorithm

- Similar to SIR, but the importance samples and importance weights are computed by simulating a set of SDEs numerically.
- The importance processes can be obtained by continuous-discrete EKF or UKF.
- Conditionally Gaussian parts can be integrated out - Rao-Blackwellized - analytically.
- Static parameters, when in suitable conjugate form, can also be integrated out.
- Could be extended to Lévy process driven stochastic differential equations.

Continuous-Discrete SIR Algorithm

- Similar to SIR, but the importance samples and importance weights are computed by simulating a set of SDEs numerically.
- The importance processes can be obtained by continuous-discrete EKF or UKF.
- Conditionally Gaussian parts can be integrated out - Rao-Blackwellized - analytically.
- Static parameters, when in suitable conjugate form, can also be integrated out.
- Could be extended to Lévy process driven stochastic differential equations.

Continuous-Discrete SIR Algorithm

- Similar to SIR, but the importance samples and importance weights are computed by simulating a set of SDEs numerically.
- The importance processes can be obtained by continuous-discrete EKF or UKF.
- Conditionally Gaussian parts can be integrated out - Rao-Blackwellized - analytically.
- Static parameters, when in suitable conjugate form, can also be integrated out.
- Could be extended to Lévy process driven stochastic differential equations.

Continuous-Discrete SIR Algorithm

- Similar to SIR, but the importance samples and importance weights are computed by simulating a set of SDEs numerically.
- The importance processes can be obtained by continuous-discrete EKF or UKF.
- Conditionally Gaussian parts can be integrated out - Rao-Blackwellized - analytically.
- Static parameters, when in suitable conjugate form, can also be integrated out.
- Could be extended to Lévy process driven stochastic differential equations.

Continuous-Discrete SIR Algorithm

- Similar to SIR, but the importance samples and importance weights are computed by simulating a set of SDEs numerically.
- The importance processes can be obtained by continuous-discrete EKF or UKF.
- Conditionally Gaussian parts can be integrated out - Rao-Blackwellized - analytically.
- Static parameters, when in suitable conjugate form, can also be integrated out.
- Could be extended to Lévy process driven stochastic differential equations.

Noisy Simple Pendulum Problem

- Model of noisy simple pendulum:

$$\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} + a^2 \sin(x) = w(t).$$

- In Brownian motion notation:

$$\begin{aligned} \frac{\mathrm{d}x_1}{\mathrm{d}t} &= x_2 \\ \mathrm{d}x_2 &= -a^2 \sin(x_1)\,\mathrm{d}t + \mathrm{d}\beta, \end{aligned}$$

- Measurements:

$$\begin{aligned} y_k &\sim \mathrm{N}(x_1(t_k), \sigma^2) \\ \sigma^2 &\sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2), \end{aligned}$$

Noisy Simple Pendulum Problem

- Model of noisy simple pendulum:

$$\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} + a^2 \sin(x) = w(t).$$

- In Brownian motion notation:

$$\frac{\mathrm{d}x_1}{\mathrm{d}t} = x_2$$
$$\mathrm{d}x_2 = -a^2 \sin(x_1)\,\mathrm{d}t + \mathrm{d}\beta,$$

- Measurements:

$$y_k \sim \mathrm{N}(x_1(t_k), \sigma^2)$$
$$\sigma^2 \sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2),$$

Noisy Simple Pendulum Problem

- Model of noisy simple pendulum:

$$\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} + a^2 \sin(x) = w(t).$$

- In Brownian motion notation:

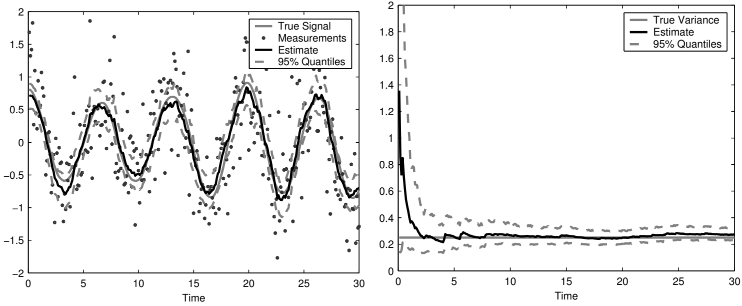
$$\begin{aligned} \frac{\mathrm{d}x_1}{\mathrm{d}t} &= x_2 \\ \mathrm{d}x_2 &= -a^2 \sin(x_1)\,\mathrm{d}t + \mathrm{d}\beta, \end{aligned}$$

- Measurements:

$$\begin{aligned} y_k &\sim \mathrm{N}(x_1(t_k), \sigma^2) \\ \sigma^2 &\sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2), \end{aligned}$$

Simulation Results

Evolution of signal estimate (left) and variance estimate (right):



Conclusion

- The discrete-time Sequential importance resampling (SIR) is not applicable to continuous-discrete filtering problems, because $p(\mathbf{x}(t_k) \mid \mathbf{x}(t_{k-1}))$ cannot be computed.
- By using the Girsanov theorem a stochastic integral formula for the importance weights can be derived.
- The weight evaluation and importance process simulation can be done with numerical methods for SDEs.
- The same efficiency improvement strategies (EKF proposal, Rao-Blackwellization, etc.) can be applied as in the discrete-time case.

Conclusion

- The discrete-time Sequential importance resampling (SIR) is not applicable to continuous-discrete filtering problems, because $p(\mathbf{x}(t_k) \,|\, \mathbf{x}(t_{k-1}))$ cannot be computed.
- By using the Girsanov theorem a stochastic integral formula for the importance weights can be derived.
- The weight evaluation and importance process simulation can be done with numerical methods for SDEs.
- The same efficiency improvement strategies (EKF proposal, Rao-Blackwellization, etc.) can be applied as in the discrete-time case.

Conclusion

- The discrete-time Sequential importance resampling (SIR) is not applicable to continuous-discrete filtering problems, because $p(\mathbf{x}(t_k) \mid \mathbf{x}(t_{k-1}))$ cannot be computed.
- By using the Girsanov theorem a stochastic integral formula for the importance weights can be derived.
- The weight evaluation and importance process simulation can be done with numerical methods for SDEs.
- The same efficiency improvement strategies (EKF proposal, Rao-Blackwellization, etc.) can be applied as in the discrete-time case.

Conclusion

- The discrete-time Sequential importance resampling (SIR) is not applicable to continuous-discrete filtering problems, because $p(\mathbf{x}(t_k) \mid \mathbf{x}(t_{k-1}))$ cannot be computed.
- By using the Girsanov theorem a stochastic integral formula for the importance weights can be derived.
- The weight evaluation and importance process simulation can be done with numerical methods for SDEs.
- The same efficiency improvement strategies (EKF proposal, Rao-Blackwellization, etc.) can be applied as in the discrete-time case.