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

1 Problem Formulation
2 Continuous-Discrete SIR
3 Example Problem
4 Conclusion

Simo Särkkä <simo.sarkka@hut.fi>
Contents

1. Problem Formulation
2. Continuous-Discrete SIR
3. Example Problem
4. Conclusion
Continuous-Discrete Filtering Problem

- Estimate the unobserved **continuous-time signal** from noisy **discrete-time measurements**
The dynamics of state $x(t)$ modeled as a stochastic differential equation (diffusion process)

$$\frac{dx}{dt} = f(x, t) dt + L \, d\beta(t).$$

Measurements $y_k$ are obtained at discrete times

$$y_k \sim p(y_k \mid x(t_k)).$$

Formal solution: Compute the posterior distribution(s)

$$p(x(t) \mid y_1, \ldots, y_k), \quad t \geq t_k.$$
Mathematical Problem Formulation

- The dynamics of state $x(t)$ modeled as a stochastic differential equation (diffusion process)
  \[
  dx = f(x, t) \, dt + \mathbf{L} \, d\beta(t).
  \]
- Measurements $y_k$ are obtained at discrete times
  \[
  y_k \sim p(y_k \mid x(t_k)).
  \]
- Formal solution: Compute the posterior distribution(s)
  \[
  p(x(t) \mid y_1, \ldots, y_k), \quad t \geq t_k.
  \]
The dynamics of state $x(t)$ modeled as a stochastic differential equation (diffusion process)

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

Measurements $y_k$ are obtained at discrete times

$$y_k \sim p(y_k | x(t_k)).$$

Formal solution: Compute the posterior distribution(s)

$$p(x(t) | y_1, \ldots, y_k), \quad t \geq t_k.$$
Problem Formulation
Continuous-Discrete SIR
Example Problem
Conclusion

Formal solution

Optimal filter

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

\[
\frac{\partial p}{\partial t} = - \sum_i \frac{\partial}{\partial x_i} (f_i(x, t) p) + \frac{1}{2} \sum_{ij} \frac{\partial^2}{\partial x_i \partial x_j} ([LQL^T]_{ij} p)
\]

2 Update step: Apply the Bayes’ rule.

\[
p(x(t_k) | y_{1:k}) = \frac{p(y_k | x(t_k)) p(x(t_k) | y_{1:k-1})}{\int p(y_k | x(t_k)) p(x(t_k) | y_{1:k-1}) \, dx(t_k)}
\]
Formal solution

Optimal filter

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

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

2. **Update step:** Apply the Bayes’ rule.

\[
p(x(t_k) \mid y_{1:k}) = \frac{p(y_k \mid x(t_k)) p(x(t_k) \mid y_{1:k-1})}{\int p(y_k \mid x(t_k)) p(x(t_k) \mid y_{1:k-1}) \, dx(t_k)}
\]
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!
Problem Formulation
Continuous-Discrete SIR
Example Problem
Conclusion

Sequential Importance Resampling

1. Draw a random sample from the importance distribution

\[ \mathbf{x}^{(i)}(t_k) \sim q(\mathbf{x}^{(i)}(t_k) \mid \mathbf{x}^{(i)}(t_{k-1})) \]  

2. Evaluate the importance weight

\[ w_k^{(i)} \propto \frac{p(y_k \mid \mathbf{x}^{(i)}(t_k)) p(\mathbf{x}^{(i)}(t_k) \mid \mathbf{x}^{(i)}(t_{k-1}))}{q(\mathbf{x}^{(i)}(t_k) \mid \mathbf{x}^{(i)}(t_{k-1}))} \]

3. Do resampling if needed.
Sequential Importance Resampling

1. Draw a random sample from the importance distribution

\[ x^{(i)}(t_k) \sim q(x^{(i)}(t_k) \mid x^{(i)}(t_{k-1})) \] (1)

2. Evaluate the importance weight

\[ w_k^{(i)} \propto \frac{p(y_k \mid x^{(i)}(t_k)) p(x^{(i)}(t_k) \mid x^{(i)}(t_{k-1}))}{q(x^{(i)}(t_k) \mid x^{(i)}(t_{k-1}))} \]

3. Do resampling if needed.
Sequential Importance Resampling

1. Draw a random sample from the importance distribution

\[ \mathbf{x}^{(i)}(t_k) \sim q(\mathbf{x}^{(i)}(t_k) \mid \mathbf{x}^{(i)}(t_{k-1})) \] (1)

2. Evaluate the importance weight

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

3. Do resampling if needed.
The Problem of SIR Weight Evaluation

The weight evaluation of SIR is of the form

\[ w_k^{(i)} \propto \frac{p(y_k | x^{(i)}(t_k)) \cdot p(x^{(i)}(t_k) | x^{(i)}(t_{k-1}))}{q(x^{(i)}(t_k) | x^{(i)}(t_{k-1}))} \]

But \( p(x(t_k) | x(t_{k-1})) \) is the solution of an arbitrary second order partial differential equation and cannot be solved.

Actually we only need the likelihood ratio

\[ \frac{p(x(t_k) | x(t_{k-1}))}{q(x(t_k) | x(t_{k-1}))} \]

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

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

But $p(x(t_k) | x(t_{k-1}))$ is the solution of an arbitrary second order partial differential equation and cannot be solved.

Actually we only need the likelihood ratio

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

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

\[ w_k^{(i)} \propto \frac{p(y_k | x^{(i)}(t_k)) p(x^{(i)}(t_k) | x^{(i)}(t_{k-1}))}{q(x^{(i)}(t_k) | x^{(i)}(t_{k-1}))} \]

- But \( p(x(t_k) | x(t_{k-1})) \) is the solution of an arbitrary second order partial differential equation and cannot be solved.

- Actually we only need the likelihood ratio

\[ \frac{p(x(t_k) | x(t_{k-1}))}{q(x(t_k) | x(t_{k-1}))} \]

- 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

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

But $p(x(t_k) | x(t_{k-1}))$ is the solution of an arbitrary second order partial differential equation and cannot be solved.

Actually we only need the likelihood ratio

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

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:
  \[
  \frac{dP_\theta}{dP_\beta} = \exp \left( \int_0^t \theta^T(t) \, d\beta(t) - \frac{1}{2} \int_0^t \|\theta(t)\|^2 \, dt \right).
  \]
- 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:

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

- 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:

$$\frac{dP_{\theta}}{dP_{\beta}} = \exp \left( \int_{0}^{t} \theta^T(t) \, d\beta(t) - \frac{1}{2} \int_{0}^{t} \|\theta(t)\|^2 \, dt \right).$$

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

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

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:
  \[
  \begin{align*}
  d\mathbf{x} &= f(\mathbf{x}, t) \, dt + L \, d\beta \\
  d\mathbf{s} &= g(\mathbf{s}, t) \, dt + B \, d\beta.
  \end{align*}
  \]
- Process \( s(t) \) can be the importance process for estimated process \( x(t) \).
- It is a \textit{stochastic integral}: Well known numerical methods for SDEs can be used.
- It is a \textit{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:

\[
\begin{align*}
\mathrm{d}x &= f(x, t) \, \mathrm{d}t + L \, \mathrm{d}\beta \\
\mathrm{d}s &= g(s, t) \, \mathrm{d}t + B \, \mathrm{d}\beta.
\end{align*}
\]

- Process \( s(t) \) can be the importance process for estimated process \( x(t) \).

- 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.

Simo Särkkä <simo.sarkka@hut.fi>
Evaluating the Likelihood Ratio

- With Girsanov theorem, we can derive expression for likelihood ratio of two SDEs:
  
  \[ \text{d} x = f(x, t) \, \text{d}t + L \, \text{d}\beta \]
  
  \[ \text{d} s = g(s, t) \, \text{d}t + B \, \text{d}\beta. \]

- Process \( s(t) \) can be the importance process for estimated process \( x(t) \).

- 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:

\[
\begin{align*}
    \mathrm{d}x &= f(x, t) \, \mathrm{d}t + L \, \mathrm{d}\beta \\
    \mathrm{d}s &= g(s, t) \, \mathrm{d}t + B \, \mathrm{d}\beta.
\end{align*}
\]

- Process \( s(t) \) can be the importance process for estimated process \( x(t) \).

- 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.

Simo Särkkä <simo.sarkka@hut.fi>
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.

Simo Särkkä <simo.sarkka@hut.fi>
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:

\[
\frac{d^2 x}{dt^2} + a^2 \sin(x) = w(t).
\]

- In Brownian motion notation:

\[
\frac{dx_1}{dt} = x_2 \\
\frac{dx_2}{dt} = -a^2 \sin(x_1) dt + d\beta,
\]

- Measurements:

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

Simo Särkkä <simo.sarkka@hut.fi>
Noisy Simple Pendulum Problem

- Model of noisy simple pendulum:

\[
\frac{d^2 x}{dt^2} + a^2 \sin(x) = w(t).
\]

- In Brownian motion notation:

\[
\frac{dx_1}{dt} = x_2 \\
\frac{dx_2}{dt} = -a^2 \sin(x_1) dt + d\beta,
\]

- Measurements:

\[
y_k \sim N(x_1(t_k), \sigma^2) \\
\sigma^2 \sim \text{Inv-}\chi^2(\nu_0, \sigma^0)
\]
Noisy Simple Pendulum Problem

- Model of noisy simple pendulum:
  \[
  \frac{d^2 x}{dt^2} + a^2 \sin(x) = w(t).
  \]

- In Brownian motion notation:
  \[
  \begin{align*}
  \frac{dx_1}{dt} &= x_2 \\
  dx_2 &= -a^2 \sin(x_1) \, dt + d\beta,
  \end{align*}
  \]

- Measurements:
  \[
  \begin{align*}
  y_k &\sim N(x_1(t_k), \sigma^2) \\
  \sigma^2 &\sim \text{Inv-} \chi^2(\nu_0, \sigma^2_0),
  \end{align*}
  \]
Evolution of signal estimate (left) and variance estimate (right):
The discrete-time **Sequential importance resampling (SIR)** is not applicable to **continuous-discrete filtering problems**, because $p(x(t_k) | x(t_{k-1}))$ 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.
The discrete-time **Sequential importance resampling** (SIR) is not applicable to **continuous-discrete filtering** problems, because $p(x(t_k) | x(t_{k-1}))$ 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.
The discrete-time Sequential importance resampling (SIR) is not applicable to continuous-discrete filtering problems, because \( p(x(t_k) \mid x(t_{k-1})) \) 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.
The discrete-time **Sequential importance resampling** (SIR) is not applicable to continuous-discrete filtering problems, because $p(x(t_k) \mid x(t_{k-1}))$ 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.