ON SEQUENTIAL MONTE CARLO SAMPLING OF DISCRETELY OBSERVED STOCHASTIC DIFFERENTIAL EQUATIONS

Simo Särkkä

Helsinki University of Technology
Laboratory of Computational Engineering
P.O. Box 9203, FIN-02015 HUT, Finland

ABSTRACT

This article considers the application of sequential importance resampling to optimal continuous-discrete filtering problems, where the dynamic model is a stochastic differential equation and the measurements are obtained at discrete instances of time. In this article it is shown how the Girsanov theorem from mathematical probability theory can be used for numerically evaluating the likelihood ratios needed by the sequential importance resampling. Rao-Blackwellization of continuous-discrete filtering models is also considered. The practical applicability of the proposed methods is demonstrated with a numerical simulation.

1. INTRODUCTION

In many applications, especially in navigation [1, 2], it is most natural to formulate the system dynamics and model uncertainties as stochastic differential equations (SDE) [3, 4], which are measured at discrete instances of time. The advantage of this kind of continuous-discrete filtering model formulation [5] over a discrete time model formulation is that the case of non-uniform sampling (i.e., varying sampling interval) is naturally included in the model. Non-uniform sampling arises in practice, for example, when processing data from multiple sensors that are not synchronized. This is common, for example, in multiple target tracking applications [6, 7]. The continuous-discrete formulation is also more realistic than a pure continuous time model (see, e.g., [8]), because sensor measurements are often processed with digital computer which only allows processing of discrete time measurements.

In this article, novel measure transformation based methods to continuous-discrete sequential importance resampling (i.e., particle filtering) are presented. The methods are based on transformations of probability measures by the Girsanov theorem [3, 4], which is a theorem from mathematical probability theory. The theorem can be used for computing likelihood ratios of stochastic processes. It states that the likelihood ratio of a stochastic process and Brownian motion, that is, the Radon-Nikodym derivative of the measure of the stochastic process with respect to the measure of Brownian motion, can be represented as an exponential martingale which is the solution to a certain stochastic differential equation.

Measure transformation based approaches are particularly successful in continuous time filtering [8], but have less been used in continuous-discrete filtering. The general idea of using the Girsanov theorem in importance sampling of SDEs has been presented, for example, in [9]. Article [10] presents the idea of using transformations of probability measures for computing the likelihood ratios between importance process and the true process in context of continuous-discrete filtering. However, the results of [10] only apply when the Euler integration scheme is used and when the dispersion matrix is invertible. The Girsanov theorem is also used in the interacting and branching particle systems [11], which are particle based solutions to nonlinear filtering problems also in the continuous-discrete setting. In these methods the Girsanov theorem is used for transforming the measure of the observation process.

Technical details as well as further analysis and applications of the methods presented in this article can be found in Author's doctoral dissertation [12].

1.1. Continuous-Discrete Filtering

The continuous-discrete filtering models considered here have the form

$$\begin{aligned} \mathrm{d}\mathbf{x}(t) &= \mathbf{f}(\mathbf{x}(t), t)\,\mathrm{d}t + \mathbf{L}\,\mathrm{d}\boldsymbol{\beta}(t) \\ \mathbf{y}_k &\sim p(\mathbf{y}_k \,|\, \mathbf{x}(t_k)), \end{aligned} \tag{1}$$

where $\mathbf{x}(t) \in \mathbb{R}^n$ is the state, $\mathbf{y}_k \in \mathbb{R}^m$ is the measurement, $\mathbf{f}$ is the drift function, $\mathbf{L}$ is the time-independent dispersion matrix, $\boldsymbol{\beta}(t)$ is Brownian motion with diffusion matrix $\mathbf{Q}_c(t)$ and $p(\mathbf{y}_k \,|\, \mathbf{x}(t_k))$ is the measurement likelihood model and it can be an arbitrary probability density or discrete probability distribution. The dynamic model, which is the first equation

The work was funded by the Academy of Finland Centre of Excellence in Computational Science and Engineering, and the SmartSpace1 project, coordinated by the Technical Research Centre of Finland VTT and funded by the National Technology Agency TEKES.

in the model (1), is interpreted as a Itô type *stochastic differential equation*.

The purpose of (Bayesian) continuous-discrete filtering is to recursively compute the posterior distribution

$$p(\mathbf{x}(t_k) \,|\, \mathbf{y}_1, \ldots, \mathbf{y}_k), \tag{2}$$

where t_k is the time of measurement $\mathbf{y}_k$. By using optimal prediction the corresponding distribution can also be computed for all time instances before the next measurement $t \in [t_k, t_{k+1})$.

1.2. Sequential Importance Resampling

Sequential importance resampling (SIR) [13, 14] is a generalization of the *particle filtering* framework for the estimation of generic state space models of the form

$$\begin{aligned} \mathbf{x}_k &\sim p(\mathbf{x}_k \mid \mathbf{x}_{k-1}) \\ \mathbf{y}_k &\sim p(\mathbf{y}_k \mid \mathbf{x}_k), \end{aligned} \tag{3}$$

where $\mathbf{x}_k \in \mathbb{R}^n$ is the state at time instance t_k and $\mathbf{y}_k \in \mathbb{R}^m$ is the measurement. The state and measurements may contain both discrete and continuous components. The dynamic model $p(\mathbf{x}_k \mid \mathbf{x}_{k-1})$ can also be the transition density of a continuous time Markov process as is the case in this article. That is, in continuous-discrete-time point of view this model can be interpreted such that the notation $\mathbf{x}_k \triangleq \mathbf{x}(t_k)$ is used, where the time dependence of state is not explicitly stated.

2. CONTINUOUS-DISCRETE PARTICLE FILTERING

2.1. Sequential Importance Resampling of Absolutely Continuous SDEs

Assume that the filtering model is of the form

$$\begin{aligned} \mathrm{d}\mathbf{x} &= \mathbf{f}(\mathbf{x}, t)\,\mathrm{d}t + \mathbf{L}\,\mathrm{d}\boldsymbol{\beta} \\ \mathbf{y}_k &\sim p(\mathbf{y}_k \,|\, \mathbf{x}(t_k)), \end{aligned} \tag{4}$$

where $\mathbf{L}$ is an invertible matrix. Further assume that there exists importance process $\mathbf{s}(t)$, which is defined by the SDE

$$\mathrm{d}\mathbf{s} = \mathbf{g}(\mathbf{s}, t)\,\mathrm{d}t + \mathbf{B}\,\mathrm{d}\boldsymbol{\beta}, \tag{5}$$

and which has the law that is a rough approximation to the filtering (or smoothing) result of the model (4), at least at the measurement times. The matrix $\mathbf{B}$ is also assumed to be invertible.

Now it is possible to generate a set of importance samples from the conditioned (i.e., filtered) process $\mathbf{x}(t)$, which is conditional to the measurements $\mathbf{y}_{1:k}$ using $\mathbf{s}(t)$ as the importance process. The motivation of this is that because the process $\mathbf{s}(t)$ is already an approximation to the optimal result, using it as the importance process is likely to produce a less degenerate particle set and thus more accurate presentation of the filtering distribution.

Because the matrices $\mathbf{L}$ and $\mathbf{B}$ are invertible, the probability measures of $\mathbf{x}$ and $\mathbf{s}$ are absolutely continuous with respect to the probability measure of the driving Brownian motion $\boldsymbol{\beta}$ and it is possible to compute likelihood ratio between the target and importance processes by applying the Girsanov theorem. The continuous-discrete SIR filter for the model can be now constructed as follows [12]:

Algorithm 2.1 (Continuous-discrete SIR I)**.** *Given the importance process $\mathbf{s}(t)$, a weighted set of samples $\{\mathbf{x}_{k-1}^{(i)}, w_{k-1}^{(i)}\}$ and the new measurement $\mathbf{y}_k$, a single step of continuous-discrete sequential importance resampling can be now performed as follows:*

1. *Draw N Brownian motions $\{\boldsymbol{\beta}^{(i)}(t), t_{k-1} \le t \le t_k, i = 1, \ldots, N\}$ and simulate the corresponding importance processes*

$$\mathrm{d}\mathbf{s}^{(i)} = \mathbf{g}(\mathbf{s}^{(i)}, t)\,\mathrm{d}t + \mathbf{B}\,\mathrm{d}\boldsymbol{\beta}^{(i)}, \qquad \mathbf{s}^{(i)}(t_{k-1}) = \mathbf{x}_{k-1}^{(i)} \tag{6}$$

from $t = t_{k-1}$ to $t = t_k$, and compute

$$\mathbf{s}^{*(i)}(t) = \mathbf{x}_{k-1}^{(i)} + \mathbf{L}\,\mathbf{B}^{-1}\,(\mathbf{s}^{(i)}(t) - \mathbf{x}_{k-1}^{(i)}), \tag{7}$$

and set

$$\mathbf{x}_k^{(i)} = \mathbf{s}^{*(i)}(t_k). \tag{8}$$

2. *For each i compute*

$$\begin{aligned} w_k^{(i)} = w_{k-1}^{(i)} \, \exp\Bigg(& \int_{t_{k-1}}^{t_k} \Big[\mathbf{L}^{-1}\,\mathbf{f}(\mathbf{s}^{*(i)}(t), t) \\ & \quad - \mathbf{B}^{-1}\,\mathbf{g}(\mathbf{s}^{(i)}(t), t)\Big]^T \mathrm{d}\boldsymbol{\beta}^{(i)} \\ & + \int_{t_{k-1}}^{t_k} \Big[\mathbf{L}^{-1}\,\mathbf{f}(\mathbf{s}^{*(i)}(t), t)\Big]^T \\ & \quad \times \Big[\mathbf{B}^{-1}\,\mathbf{g}(\mathbf{s}^{(i)}(t), t)\Big]\,\mathrm{d}t \\ & - \frac{1}{2}\int_{t_{k-1}}^{t_k} \Big(||\mathbf{L}^{-1}\,\mathbf{f}(\mathbf{s}^{*(i)}(t), t)||^2 \\ & \quad + ||\mathbf{B}^{-1}\,\mathbf{g}(\mathbf{s}^{(i)}(t), t)||^2\Big)\,\mathrm{d}t\Bigg) \\ & \times p(\mathbf{y}_k \,|\, \mathbf{x}_k^{(i)}). \end{aligned} \tag{9}$$

and re-normalize the weights to sum to unity.

3. *If the effective number of particles is too low, perform resampling.*

2.2. Sequential Importance Resampling for More General SDEs

It is also possible to construct a similar SIR algorithm for models, where there is an absolutely continuous type of model, which is *embedded* inside a *deterministic* differential equation model. This kind of models are typical in navigation and stochastic control applications, where the deterministic part is typically a plain integral operator. Because the outer operator is deterministic, the likelihood ratios of processes are determined by the inner stochastic processes alone and thus importance sampling of this kind of process is very similar to sampling of the processes considered above.

Assume that the model is of the form

$$\begin{aligned} \frac{\mathrm{d}\mathbf{x}_1}{\mathrm{d}t} &= \mathbf{f}_1(\mathbf{x}_1, \mathbf{x}_2, t), & \mathbf{x}_1(0) &= \mathbf{x}_{1,0} \\ \mathrm{d}\mathbf{x}_2 &= \mathbf{f}_2(\mathbf{x}_1, \mathbf{x}_2, t)\,\mathrm{d}t + \mathbf{L}\,\mathrm{d}\boldsymbol{\beta}, & \mathbf{x}_2(0) &= \mathbf{x}_{2,0}, \end{aligned} \quad (10)$$

where $\mathbf{f}_1(\cdot)$ and $\mathbf{f}_2(\cdot)$ are deterministic functions, $\boldsymbol{\beta}(t)$ is a Brownian motion and $\mathbf{L}$ is invertible matrix. Note that because the dimensionality of Brownian motion is less than of the joint state $(\mathbf{x}_1\ \mathbf{x}_2)^T$ it is not possible to compute the likelihood ratio between the process and Brownian motion by the Girsanov theorem directly.

However, it turns out that if the importance process for $(\mathbf{x}_1\ \mathbf{x}_2)^T$ is formed as follows

$$\begin{aligned} \frac{\mathrm{d}\mathbf{s}_1}{\mathrm{d}t} &= \mathbf{f}_1(\mathbf{s}_1, \mathbf{s}_2, t), & \mathbf{s}_1(0) &= \mathbf{x}_{1,0} \\ \mathrm{d}\mathbf{s}_2 &= \mathbf{g}_2(\mathbf{s}_1, \mathbf{s}_2, t)\,\mathrm{d}t + \mathbf{B}\,\mathrm{d}\boldsymbol{\beta}, & \mathbf{s}_2(0) &= \mathbf{x}_{2,0}, \end{aligned} \quad (11)$$

then the importance weights can be computed in exactly the same way as when forming importance sample of $\mathbf{x}_2(t)$ using $\mathbf{s}_2(t)$ as the importance process. The resulting SIR algorithm is very similar to Algorithm 2.1 [12].

Rao-Blackwellized Sequential Importance Resampling

If the dynamic model has the form

$$\begin{aligned} \mathrm{d}\mathbf{x}_1 &= \mathbf{F}(\mathbf{x}_2, \mathbf{x}_3, t)\,\mathbf{x}_1\,\mathrm{d}t + \mathbf{f}_1(\mathbf{x}_2, \mathbf{x}_3, t)\,\mathrm{d}t \\ &\quad + \mathbf{V}(\mathbf{x}_2, \mathbf{x}_3, t)\,\mathrm{d}\boldsymbol{\eta} \\ \frac{\mathrm{d}\mathbf{x}_2}{\mathrm{d}t} &= \mathbf{f}_2(\mathbf{x}_2, \mathbf{x}_3, t) \\ \mathrm{d}\mathbf{x}_3 &= \mathbf{f}_3(\mathbf{x}_2, \mathbf{x}_3, t)\,\mathrm{d}t + \mathbf{L}\,\mathrm{d}\boldsymbol{\beta}, \end{aligned} \quad (12)$$

where $\boldsymbol{\eta}$ and $\boldsymbol{\beta}$ are independent Brownian motions, and if the measurement model is suitably conditionally Gaussian, it is possible to integrate linear part of the filtering equations analytically Kalman filter and the other part by Monte Carlo sampling [12].

Analogously to the discrete time case presented in [15], the procedure of Rao-Blackwellization can often be applied to models with unknown static parameters. If the posterior distribution of the unknown static parameters $\boldsymbol{\theta}$ depends only on a suitable set of sufficient statistics $\mathbf{T}_k = \mathbf{T}_k(\mathbf{x}_{1:k}, \mathbf{y}_{1:k})$, the parameter can be marginalized out analytically and only the state needs to be sampled.

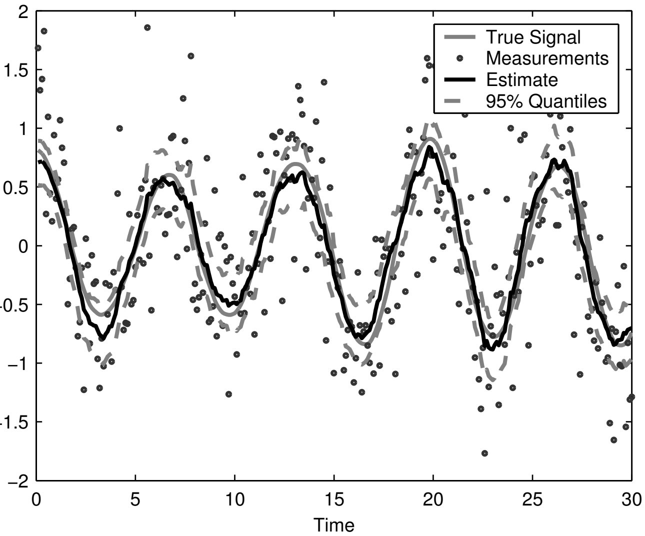

Fig. 1. The result of applying continuous-discrete particle filter with UKF proposal to a simulated noisy pendulum data.

3. ILLUSTRATIVE EXAMPLE

3.1. Simple Pendulum with Noise

The stochastic differential equation for the angular position of a simple pendulum [16], which is distorted by random white noise accelerations $w(t)$ with spectral density q can be written as

$$\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} + a^2\,\sin(x) = w(t), \quad (13)$$

where a is the angular velocity of the (linearized) pendulum.

Assume that the state of the pendulum is measured once per unit time and the measurements are corrupted by Gaussian measurement noise with an unknown variance σ^2. A suitable model in this case is

$$\begin{aligned} y_k &\sim \mathrm{N}(x_1(t_k), \sigma^2) \\ \sigma^2 &\sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2). \end{aligned} \quad (14)$$

The variance σ^2 is now an unknown static variable, where the procedure of Rao-Blackwellization can be applied. Figure 1 shows the result of applying the continuous-discrete particle filter with UKF proposal and 1000 particles to a simulated data. The data was generated from the noisy pendulum model with process noise spectral density $q = 0.01$, angular velocity $a = 1$ and the sampling step size was $\Delta t = 0.1$. The estimate can be seen to be quite close to the true signal.

The evolution of the posterior distribution of the variance parameter is shown in the Figure 2. In the beginning the uncertainty about the variance is higher, but the distribution quickly converges to the area of the true value.

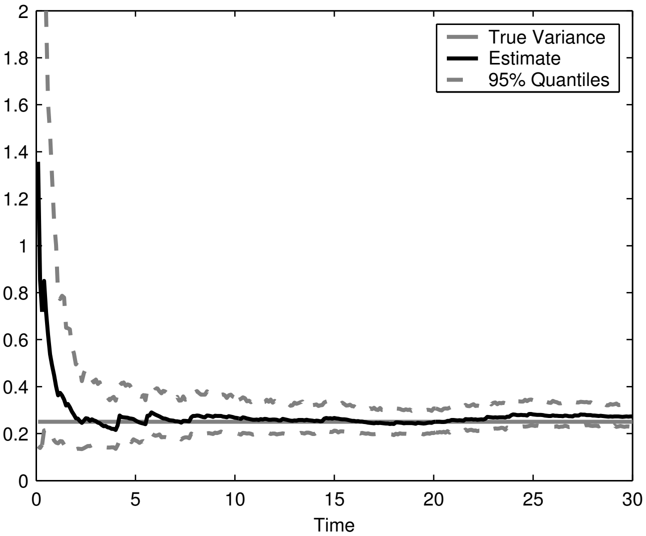

Fig. 2. The evolution of variance distribution in the noisy pendulum problem.

4. CONCLUSION

In this article, a new class of sequential Monte Carlo methods for continuous-discrete optimal filtering has been presented. These methods are based on transformations of probability measures by the Girsanov theorem. The new methods are applicable to a general class of models, in particular, they can be applied to many models with singular dispersion matrices, unlike many previously proposed measure transformation based sampling methods. The new methods have been illustrated in a simulated problem, where both the implementation details of the algorithms and the simulation results have been reported.

5. ACKNOWLEDGMENT

The author would like to thank Aki Vehtari and Jouko Lampinen for helpful discussions and comments on the manuscript.

6. REFERENCES

[1] Mohinder S. Grewal, Lawrence R. Weill, and Angus P. Andrews, *Global Positioning Systems, Inertial Navigation and Integration*, Wiley Interscience, 2001.

[2] Yaakov Bar-Shalom, Xiao-Rong Li, and Thiagalingam Kirubarajan, *Estimation with Applications to Tracking and Navigation*, Wiley Interscience, 2001.

[3] Ioannis Karatzas and Steven E. Shreve, *Brownian Motion and Stochastic Calculus*, Springer, 1991.

[4] Bernt Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, Springer, 6 edition, 2003.

[5] Andrew H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, 1970.

[6] Samuel Blackman and Robert Popoli, *Design and Analysis of Modern Tracking Systems*, Artech House Radar Library, 1999.

[7] Lawrence D. Stone, Carl A. Barlow, and Thomas L. Corwin, *Bayesian Multiple Target Tracking*, Artech House, Boston, London, 1999.

[8] Gopinath Kallianpur, *Stochastic Filtering Theory*, Springer-Verlag, 1980.

[9] Peter E. Kloeden and Eckhard Platen, *Numerical Solution to Stochastic Differential Equations*, Springer, 1999.

[10] E. L. Ionides, “Inference and filtering for partially observed diffusion processes via sequential Monte Carlo,” Tech. Rep., University of Michigan Statistics Department Technical Report #405, 2004.

[11] P. Del Moral and L. Miclo, “Branching and interacting particle systems. Approximations of Feynman-Kac formulae with applications to non-linear filtering,” in *Séminaire de Probabilités XXXIV*. Springer-Verlag, 2000.

[12] Simo Särkkä, *Recursive Bayesian Inference on Stochastic Differential Equations*, Doctoral dissertation, Helsinki University of Technology, 2006.

[13] Arnaud Doucet, Nando de Freitas, and Neil Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, Springer, 2001.

[14] Branko Ristic, Sanjeev Arulampalam, and Neil Gordon, *Beyond the Kalman Filter*, Artech House, 2004.

[15] Geir Storvik, “Particle filters in state space models with the presence of unknown static parameters,” *IEEE Transactions on Signal Processing*, vol. 50(2), pp. 281–289, 2002.

[16] Marcelo Alonso and Edward J. Finn, *Fundamental University Physics, Volume I: Mechanics and Thermodynamics*, Addison-Wesley, 2nd edition, 1980.