Bayesian Analysis (2008) **3**, Number 3, pp. 555–584

Application of Girsanov Theorem to Particle Filtering of Discretely Observed Continuous-Time Non-Linear Systems

Simo Särkkä* and Tommi Sottinen†

Abstract. This article considers the application of particle filtering to continuous-discrete optimal filtering problems, where the system model is a stochastic differential equation, and noisy measurements of the system are obtained at discrete instances of time. It is shown how the Girsanov theorem can be used for evaluating the likelihood ratios needed in importance sampling. It is also shown how the methodology can be applied to a class of models, where the driving noise process is lower in the dimensionality than the state and thus the laws of the state and the noise are not absolutely continuous. Rao-Blackwellization of conditionally Gaussian models and unknown static parameter models is also considered.

Keywords: Girsanov theorem, particle filtering, continuous-discrete filtering

1 Introduction

This article considers the application of sequential importance sampling based *particle filtering* (see, e.g. Kitagawa 1996; Doucet et al. 2001) to *continuous-discrete filtering problems* (Jazwinski 1970), where the dynamic model is a stochastic differential equation of the form

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t), t)\, dt + \mathbf{L}(t)\, d\boldsymbol{\beta}(t), \tag{1}$$

where $\mathbf{x}(t) \in \mathbb{R}^n$ is the state, $\mathbf{f} : \mathbb{R}^n \times \mathbb{R}_+ \mapsto \mathbb{R}^n$ is the drift term, $\mathbf{L}(t) : \mathbb{R}^n \times \mathbb{R}_+ \mapsto \mathbb{R}^{n \times s}$ is the dispersion matrix, and $\boldsymbol{\beta}(t) \in \mathbb{R}^s$ is an s-dimensional Brownian motion with diffusion matrix $\mathbf{Q}(t)$. It is assumed that the required conditions (Karatzas and Shreve 1991; Øksendal 2003) for existence of a strong solution to the equation are satisfied. In this article, we first consider the case where the dimensionality of the state is the same as the dimensionality of the Brownian motion, that is, where $s = n$. We also extend the results to the singular case where the dimensionality of the Brownian motion is less than the dimensionality of the state, that is, where $s < n$.

The likelihood of a measurement $\mathbf{y}_k$ is modeled by a probability density, which is a function of the state at time t_k:

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

The purpose of the *Bayesian optimal continuous-discrete filter* is to compute the posterior distribution (or at least the posterior mean) of the current state $\mathbf{x}(t_k)$ given the

*Lab. of Computational Engineering, Helsinki University of Technology, Espoo, Finland, mailto:simo.sarkka@hut.fi

†School of Science and Engineering, Reykjavik University, Reykjavik, Iceland, mailto:tommi@ru.is

 DOI:10.1214/08-BA322

measurements up to the current time, that is (Jazwinski 1966, 1970),

$$p(\mathbf{x}(t_k) \,|\, \mathbf{y}_1, \ldots, \mathbf{y}_k). \tag{3}$$

These continuous-discrete filtering models are common in engineering applications, especially in the fields of navigation, communication and control (Bar-Shalom et al. 2001; Grewal et al. 2001; Stengel 1994; Van Trees 1968, 1971). In these applications, the dynamic system or a physical phenomenon can be modeled as a stochastic differential equation, which is observed at discrete instances of time with certain physical sensors. The purpose of the filtering or recursive estimation is to infer the state of the system from the observed noisy measurements.

In this article, novel measure transformation based methods for continuous-discrete sequential importance resampling
(see, e.g. Gordon et al. 1993; Kitagawa 1996; Pitt and Shephard 1999; Doucet et al. 2001; Ristic et al. 2004) are presented. Some of the methods have already been presented in (Särkkä 2006b,a), but here the methods are significantly extended. The methods are based on transformations of probability measures by the Girsanov theorem (Kallianpur 1980; Karatzas and Shreve 1991; Øksendal 2003), 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 a Brownian motion, that is, the Radon-Nikodym derivative of the measure of the stochastic process with respect to the measure of the 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 (Kallianpur 1980), but are less common in continuous-discrete filtering. The general idea of using the Girsanov theorem in importance sampling of SDEs has been presented, for example, in Kloeden and Platen (1999). Similar ideas have also been presented by several authors (Ionides 2004; Crisan and Lyons 1999; Crisan et al. 1998; Crisan 2003; Moral and Miclo 2000).

Beskos et al. (2006) consider exact Monte Carlo simulation of a restricted class of diffusion models, which are observed at discrete instances of time without any observation error. As shown in the discussion of the article, the observation errors can be included in the model. Fearnhead et al. (2008) introduce particles filters for a class of multidimensional diffusion processes, and the used Monte Carlo sampling methodology is based on the exact simulation framework of Beskos et al. (2006). The difference to the present methodology is that the methods of Fearnhead et al. (2008) are not based on time-discretization.

Durham and Gallant (2002) consider simulated maximum likelihood estimation of parameters of discretely observed stochastic differential equations, where all or some of the components are perfectly observed. The methods are based on approximating the transition densities of the processes and modeling the unobserved sample paths as latent data. Golightly and Wilkinson (2006) apply similar methodology to sequential estimation of state and parameters of stochastic differential equation models. Chib et al.

(2004) consider MCMC based simulation of diffusion driven state space models. In the article, it is also shown how the methodology can be applied to particle filtering of such models.

The advantages of the method proposed here over the previously proposed methods are:

- Unlike many measure transformation based approaches the methodology presented here is not restricted to one-dimensional or to SDE models with non-singular dispersion or diffusion matrices. The state dimensionality can be higher than the dimensionality of the driving Brownian motion, which is equivalent to the case that the dispersion/diffusion matrix is singular.
- The SDE formulation of the likelihood ratio computation allows efficient numerical solution to the problem. In particular, simulation based approaches (Kloeden and Platen 1999) can be applied. Of course, any other numerical methods for SDEs could be applied as well.
- Dispersion (and diffusion) matrices may depend on time, that is, the driving process can be time inhomogeneous.
- The observation errors can be easily modeled and the model flexibility is the same as with discrete-time particle filtering.
- Efficient importance distributions and Rao-Blackwellization can be easily used for improving the efficiency of the sampling.

2 Continuous-Discrete Sequential Importance Resampling

2.1 Filtering Models

We shall concentrate on the following four forms of dynamic models:

1. *Non-singular models*, where the dispersion matrices are invertible and thus the dimensionality of the process is the same as that of the driving Brownian motion. The advantage of these processes is that their likelihood ratios can be easily evaluated using the Girsanov theorem, but the problem is that they are too restricted models for many applications.
2. *Singular models*, where there is a non-singular type of model, which is embedded inside a deterministic differential equation model and thus the joint model is singular because the dimensionality of the process is higher than that of the driving Brownian motion. These 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 process is very similar to that for the processes of non-singular type above.

3. *Conditional Gaussian models*, where a linear stochastic differential equation is driven by a model of the non-singular or singular type above. These models can be handled such that we only sample the inner process and integrate the linear part using the Kalman filter. In this way we can form a Rao-Blackwellized estimate, where the probability density is approximated by a mixture of Gaussian distributions.

4. *Conjugate static parameter models*, where the model contains a static parameter in such conjugate form that certain marginalizations can be analytically evaluated. This results in a particle filter, where only the dynamic state is sampled and the sufficient statistics of the static parameter are evaluated at each update stage.

2.2 Non-Singular and Singular Models

Assume that the filtering model is of the form

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

where $\boldsymbol{\beta}(t)$ is a Brownian motion with positive definite diffusion matrix $\mathbf{Q}(t)$, $\mathbf{L}(t)$ is an invertible matrix for all $t \geq 0$ and the initial conditions are $\mathbf{x}(0) \sim p(\mathbf{x}(0))$. Further assume that we have constructed an importance process $\mathbf{s}(t)$, which is defined by the SDE

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

and which has a probability law that is a rough approximation to the filtering (or smoothing) distribution of the model (4), at least at the measurement times. The matrix $\mathbf{B}(t)$ is also assumed to be invertible for all $t \geq 0$. Note that at this point we do not want to restrict the matrix $\mathbf{B}(t)$ to be the same as $\mathbf{L}(t)$, because this allows usage of greater class of importance processes as shown later in this article.

Now it is possible to generate a set of importance samples from the conditioned (i.e., filtered) process $\mathbf{x}(t)$, which is conditional on the measurements $\mathbf{y}_{1:k}$ using $\mathbf{s}(t)$ as the importance process. The motivation of this is that because the process $\mathbf{s}(t)$ already is 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 $\mathbf{L}(t)$ and $\mathbf{B}(t)$ 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 $\boldsymbol{\beta}(t)$ and it is possible to compute the likelihood ratio between the target and importance processes by applying the Girsanov theorem. The explicit expression and derivation of this likelihood ratio is given in Theorem 3 of Appendix 1.

The SIR algorithm recursion starts by drawing samples $\{\mathbf{x}_0^{(i)}\}$ from the initial distribution and setting $w_0^{(i)} = 1/N$, where N is the number of Monte Carlo samples. The continuous-discrete SIR filter algorithm then proceeds as follows:

Algorithm 2.1 *(CD-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 performed as follows:

1. Simulate N realizations of the importance processes

$$\begin{aligned} d\mathbf{s}^{(i)} &= \mathbf{g}(\mathbf{s}^{(i)}, t)\, dt + \mathbf{B}(t)\, d\boldsymbol{\beta}^{(i)}(t), & \mathbf{s}^{(i)}(t_{k-1}) &= \mathbf{x}_{k-1}^{(i)} \\ d\mathbf{s}^{*(i)}(t) &= \mathbf{L}(t)\, \mathbf{B}^{-1}(t)\, d\mathbf{s}^{(i)}(t), & \mathbf{s}^{*(i)}(t_{k-1}) &= \mathbf{x}_{k-1}^{(i)}, \end{aligned}$$

   from $t = t_{k-1}$ to $t = t_k$, where $\boldsymbol{\beta}^{(i)}(t)$ are independent Brownian motions, and $i = 1, \ldots, N$.

2. At the same time, simulate the log-likelihood ratios

$$\begin{aligned} d\Lambda^{(i)} &= \left\{\mathbf{f}(\mathbf{s}^{*(i)}(t), t) - \mathbf{L}(t)\, \mathbf{B}^{-1}(t)\, \mathbf{g}(\mathbf{s}^{(i)}(t), t)\right\}^T \\ &\quad \times \mathbf{L}^{-T}(t)\, \mathbf{Q}^{-1}(t)\, d\boldsymbol{\beta}^{(i)}(t) \\ &\quad - \frac{1}{2}\left\{\mathbf{f}(\mathbf{s}^{*(i)}(t), t) - \mathbf{L}(t)\, \mathbf{B}^{-1}(t)\, \mathbf{g}(\mathbf{s}^{(i)}(t), t)\right\}^T \\ &\quad \times \left\{\mathbf{L}(t)\, \mathbf{Q}(t)\, \mathbf{L}^T(t)\right\}^{-1} \\ &\quad \times \left\{\mathbf{f}(\mathbf{s}^{*(i)}(t), t) - \mathbf{L}(t)\, \mathbf{B}^{-1}(t)\, \mathbf{g}(\mathbf{s}^{(i)}(t), t)\right\} dt, \\ \Lambda^{(i)}(t_{k-1}) &= 0, \end{aligned}$$

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

$$\begin{aligned} \mathbf{x}_k^{(i)} &= \mathbf{s}^{*(i)}(t_k) \\ Z_k^{(i)} &= \exp\left\{\Lambda^{(i)}(t_k)\right\}. \end{aligned}$$

   Note that the realizations of the Brownian motions must be the same as in the simulation of the importance processes.

3. For each i compute

$$w_k^{(i)} = w_{k-1}^{(i)}\, Z_k^{(i)}\, p(\mathbf{y}_k \,|\, \mathbf{x}_k^{(i)}),$$

   and re-normalize the weights to sum to unity.

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

Some practical points about the implementation:

- The importance process $\mathbf{s}(t)$ required by the algorithm can be obtained by using, for example, the extended Kalman filter (EKF). An example of this approach is given in Section 3.1 of this article.

- The simulation of the importance processes and likelihood ratios above can be performed using any of the well known numerical methods for simulation of stochastic differential equations (Kloeden and Platen 1999). In this article we have used the simple Euler-Maruyama method, which can be considered as a stochastic version of the Euler integration for non-stochastic differential equations.

The class (4) is actually a very restricted class of dynamic models, where it is required that the probability law of the state is absolutely continuous with respect to the law of the driving Brownian motion. These models are common in mathematical treatment of stochastic differential equations and such models can be found, for example, in mathematical finance (see, e.g., Karatzas and Shreve 1991; Øksendal 2003). However, most of the models used in navigation and telecommunications applications do not fit into this class, and for this reason the results need to be extended.

It is also possible to construct a similar SIR algorithm for more general models, where there is an absolutely continuous type of model, which is *embedded* inside a *deterministic* differential equation model. These models are typical in navigation, communication and stochastic control applications (Bar-Shalom et al. 2001; Grewal et al. 2001; Stengel 1994; Van Trees 1968, 1971), 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 process is very similar to sampling of the processes considered above.

Assume that the model is of the form

$$\begin{aligned} \frac{d\mathbf{x}_1}{dt} &= \mathbf{f}_1(\mathbf{x}_1, \mathbf{x}_2, t), \\ d\mathbf{x}_2 &= \mathbf{f}_2(\mathbf{x}_1, \mathbf{x}_2, t)\, dt + \mathbf{L}(t)\, d\boldsymbol{\beta} \\ \mathbf{y}_k &\sim p(\mathbf{y}_k \,|\, \mathbf{x}_1(t_k), \mathbf{x}_2(t_k)), \end{aligned} \tag{6}$$

where $\mathbf{f}_1(\cdot)$ and $\mathbf{f}_2(\cdot)$ are deterministic functions, $\boldsymbol{\beta}(t)$ is a Brownian motion, $\mathbf{L}(t)$ is invertible matrix and the initial conditions are $\mathbf{x}_1(0), \mathbf{x}_2(0) \sim p(\mathbf{x}_1(0), \mathbf{x}_2(0))$. Note that because the dimensionality of the Brownian motion is less than that of the joint state $(\mathbf{x}_1\ \mathbf{x}_2)^T$ it is not possible to compute the likelihood ratio between the process and the 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

$$\begin{aligned} \frac{d\mathbf{s}_1}{dt} &= \mathbf{f}_1(\mathbf{s}_1, \mathbf{s}_2, t) \\ d\mathbf{s}_2 &= \mathbf{g}_2(\mathbf{s}_1, \mathbf{s}_2, t)\, dt + \mathbf{B}(t)\, d\boldsymbol{\beta}, \end{aligned} \tag{7}$$

then the importance weights can be computed in exactly the same way as when forming importance samples of $\mathbf{x}_2(t)$ using $\mathbf{s}_2(t)$ as the importance process.

The likelihood ratio expression is given in Theorem 4 of Appendix 1. The SIR algorithm is started by first drawing samples from the initial distribution and then for each measurement, the following steps are performed:

Algorithm 2.2 *(CD-SIR II).* Given the importance process $\mathbf{s}_1(t), \mathbf{s}_2(t)$, a weighted set of samples $\{\mathbf{x}_{1,k-1}^{(i)}, \mathbf{x}_{2,k-1}^{(i)}, w_{k-1}^{(i)}\}$ and the new measurement $\mathbf{y}_k$, a single step of *continuous-discrete sequential importance resampling* can be performed as follows:

1. Simulate N realizations of the importance processes

$$
\begin{aligned}
\frac{d\mathbf{s}_1^{(i)}}{dt} &= \mathbf{f}_1(\mathbf{s}_1^{(i)}, \mathbf{s}_2^{(i)}, t), & \mathbf{s}_1^{(i)}(t_{k-1}) &= \mathbf{x}_{1,k-1}^{(i)} \\
d\mathbf{s}_2^{(i)} &= \mathbf{g}_2(\mathbf{s}_1^{(i)}, \mathbf{s}_2^{(i)}, t)\, dt + \mathbf{B}(t)\, d\boldsymbol{\beta}^{(i)}(t), & \mathbf{s}_2^{(i)}(t_{k-1}) &= \mathbf{x}_{2,k-1}^{(i)} \\
\frac{d\mathbf{s}_1^{*(i)}}{dt} &= \mathbf{f}_1(\mathbf{s}_1^{*(i)}, \mathbf{s}_2^{*(i)}, t), & \mathbf{s}_1^{*(i)}(t_{k-1}) &= \mathbf{x}_{1,k-1}^{(i)} \\
d\mathbf{s}_2^{*(i)} &= \mathbf{L}(t)\, \mathbf{B}^{-1}(t)\, d\mathbf{s}_2, & \mathbf{s}_2^{*(i)}(t_{k-1}) &= \mathbf{x}_{2,k-1}^{(i)},
\end{aligned}
$$

2. Simulate the log-likelihood ratios (using the same Brownian motion realizations as above)

$$
\begin{aligned}
d\Lambda^{(i)} = &\left\{\mathbf{f}_2(\mathbf{s}_1^{*(i)}(t), \mathbf{s}_2^{*(i)}(t), t) - \mathbf{L}(t)\, \mathbf{B}^{-1}(t)\, \mathbf{g}_2(\mathbf{s}_1^{(i)}(t), \mathbf{s}_2^{(i)}(t), t)\right\}^T \\
&\times \mathbf{L}^{-T}(t)\, \mathbf{Q}^{-1}(t)\, d\boldsymbol{\beta}^{(i)}(t) \\
&- \frac{1}{2}\left\{\mathbf{f}_2(\mathbf{s}_1^{*(i)}(t), \mathbf{s}_2^{*(i)}(t), t) - \mathbf{L}(t)\, \mathbf{B}^{-1}(t)\, \mathbf{g}_2(\mathbf{s}_1^{(i)}(t), \mathbf{s}_2^{(i)}(t), t)\right\}^T \\
&\times \left\{\mathbf{L}(t)\, \mathbf{Q}(t)\, \mathbf{L}^T(t)\right\}^{-1} \\
&\times \left\{\mathbf{f}_2(\mathbf{s}_1^{*(i)}(t), \mathbf{s}_2^{*(i)}(t), t) - \mathbf{L}(t)\, \mathbf{B}^{-1}(t)\, \mathbf{g}_2(\mathbf{s}_1^{(i)}(t), \mathbf{s}_2^{(i)}(t), t)\right\} dt, \\
\Lambda^{(i)}&(t_{k-1}) = 0,
\end{aligned}
$$

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

$$
\begin{aligned}
\mathbf{x}_{1,k}^{(i)} &= \mathbf{s}_1^{*(i)}(t_k) \\
\mathbf{x}_{2,k}^{(i)} &= \mathbf{s}_2^{*(i)}(t_k) \\
Z_k^{(i)} &= \exp\left\{\Lambda^{(i)}(t_k)\right\}.
\end{aligned}
\tag{8}
$$

3. For each i compute

$$
w_k^{(i)} = w_{k-1}^{(i)}\, Z_k^{(i)}\, p(\mathbf{y}_k \,|\, \mathbf{x}_{1,k}^{(i)}, \mathbf{x}_{2,k}^{(i)}),
\tag{9}
$$

   and re-normalize the weights to sum to unity.

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

The importance process $\mathbf{s}(t)$ required by the algorithm can be obtained by using, for example, continuous-discrete EKF and then extracting the estimate of the inner process $\mathbf{s}_2(t)$ from the joint estimate.

2.3 Rao-Blackwellization of Conditional Gaussian Models

Now we shall consider dynamic models, where a *linear* stochastic differential equation is driven by a singular or non-singular model considered in the previous section. These models can be handled such that only the inner process is sampled and the linear part is integrated out using the continuous-discrete Kalman filter. Then it is possible to form a Rao-Blackwellized estimate, where the probability density is approximated by a mixture of Gaussian distributions. The measurement model is assumed to be of the same form as in previous sections, but linear with respect to the state variables corresponding to the linear part of the dynamic process.

Dynamic models with conditional Gaussian parts arise, for example, when the measurement noise correlations are modeled with state augmentation (see, e.g., Gelb 1974). Actually, in this case, the direct application of particle filter without Rao-Blackwellization would be impossible because the measurement model is formally singular. However, the Rao-Blackwellized filter can be easily applied to these models.

Assume that the dynamic model is of the form

$$\begin{aligned} d\mathbf{x}_1 &= \mathbf{F}(\mathbf{x}_2, \mathbf{x}_3, t)\, \mathbf{x}_1 \, dt + \mathbf{f}_1(\mathbf{x}_2, \mathbf{x}_3, t)\, dt + \mathbf{V}(\mathbf{x}_2, \mathbf{x}_3, t)\, d\boldsymbol{\eta} \\ \frac{d\mathbf{x}_2}{dt} &= \mathbf{f}_2(\mathbf{x}_2, \mathbf{x}_3, t) \\ d\mathbf{x}_3 &= \mathbf{f}_3(\mathbf{x}_2, \mathbf{x}_3, t)\, dt + \mathbf{L}(t)\, d\boldsymbol{\beta}, \end{aligned} \tag{10}$$

where $\boldsymbol{\eta}$ and $\boldsymbol{\beta}$ are independent Brownian motions with diffusion matrices $\mathbf{Q}_\eta(t)$ and $\mathbf{Q}_\beta(t)$, respectively. Also assume that the initial conditions are given as:

$$\begin{aligned} \mathbf{x}_1(0) &\sim \mathcal{N}(\mathbf{m}_0, \mathbf{P}_0) \\ \mathbf{x}_2(0), \mathbf{x}_3(0) &\sim p(\mathbf{x}_2(0), \mathbf{x}_3(0)), \end{aligned} \tag{11}$$

and the initial conditions of $\mathbf{x}_1(0)$ are independent from those of $\mathbf{x}_2(0)$ and $\mathbf{x}_3(0)$.

In this case an importance process can be formed as

$$\begin{aligned} d\mathbf{s}_1 &= \mathbf{F}(\mathbf{s}_2, \mathbf{s}_3, t)\, \mathbf{s}_1 \, dt + \mathbf{f}_1(\mathbf{s}_2, \mathbf{s}_3, t)\, dt + \mathbf{V}(\mathbf{s}_2, \mathbf{s}_3, t)\, d\boldsymbol{\eta}, \\ \frac{d\mathbf{s}_2}{dt} &= \mathbf{f}_2(\mathbf{s}_2, \mathbf{s}_3, t) \\ d\mathbf{s}_3 &= \mathbf{g}_3(\mathbf{s}_2, \mathbf{s}_3, t)\, dt + \mathbf{B}(t)\, d\boldsymbol{\beta}, \end{aligned} \tag{12}$$

with the same initial conditions. In both the original and importance processes, conditional on the filtration of the second Brownian motion $\mathcal{F}_t = \sigma(\boldsymbol{\beta}(s), 0 \le s \le t)$ and to initial conditions, the law of the first equation is determined by the mean and covariance of the Gaussian process, which is driven by the process $\boldsymbol{\eta}(t)$. Thus, conditional on $\mathbf{x}_2$ and $\mathbf{x}_3$ the process $\mathbf{x}_1(t)$ is Gaussian for all t. The same applies to the importance process.

Now it is possible to integrate out the Gaussian parts of both processes. This procedure results in the following marginalized equations for the original process:

$$\begin{aligned}\frac{d\mathbf{m}_x(t)}{dt} &= \mathbf{F}(\mathbf{x}_2, \mathbf{x}_3, t)\,\mathbf{m}_x(t) + \mathbf{f}_1(\mathbf{x}_2, \mathbf{x}_3, t)\\ \frac{d\mathbf{P}_x(t)}{dt} &= \mathbf{F}(\mathbf{x}_2, \mathbf{x}_3, t)\,\mathbf{P}_x(t) + \mathbf{P}_x(t)\,\mathbf{F}^T(\mathbf{x}_2, \mathbf{x}_3, t)\\ &\quad + \mathbf{V}(\mathbf{x}_2, \mathbf{x}_3, t)\,\mathbf{Q}_\eta(t)\,\mathbf{V}^T(\mathbf{x}_2, \mathbf{x}_3, t)\\ \frac{d\mathbf{x}_2}{dt} &= \mathbf{f}_2(\mathbf{x}_2, \mathbf{x}_3, t)\\ d\mathbf{x}_3 &= \mathbf{f}_3(\mathbf{x}_2, \mathbf{x}_3, t)\,dt + \mathbf{L}(t)\,d\boldsymbol{\beta},\end{aligned} \tag{13}$$

where $\mathbf{m}_x(t)$ and $\mathbf{P}_x(t)$ are the mean and covariance of the Gaussian process. For the importance process we get similarly:

$$\begin{aligned}\frac{d\mathbf{m}_s(t)}{dt} &= \mathbf{F}(\mathbf{s}_2, \mathbf{s}_3, t)\,\mathbf{m}_s(t) + \mathbf{f}_1(\mathbf{s}_2, \mathbf{s}_3, t)\\ \frac{d\mathbf{P}_s(t)}{dt} &= \mathbf{F}(\mathbf{s}_2, \mathbf{s}_3, t)\,\mathbf{P}_s(t) + \mathbf{P}_s(t)\,\mathbf{F}^T(\mathbf{s}_2, \mathbf{s}_3, t)\\ &\quad + \mathbf{V}(\mathbf{s}_2, \mathbf{s}_3, t)\,\mathbf{Q}_\eta(t)\,\mathbf{V}^T(\mathbf{s}_2, \mathbf{s}_3, t)\\ \frac{d\mathbf{s}_2}{dt} &= \mathbf{f}_2(\mathbf{s}_2, \mathbf{s}_3, t)\\ d\mathbf{s}_3 &= \mathbf{g}_3(\mathbf{s}_2, \mathbf{s}_3, t)\,dt + \mathbf{B}(t)\,d\boldsymbol{\beta},\end{aligned} \tag{14}$$

The models (13) and (14) have now the form, where the Algorithm 2.2 can be used. The importance sampling now results in the set of weighted samples

$$\{w^{(i)}, \mathbf{m}^{(i)}, \mathbf{P}^{(i)}, \mathbf{x}_2^{(i)}, \mathbf{x}_3^{(i)}\}, \tag{15}$$

such that the probability density of the state $\mathbf{x}(t) = (\mathbf{x}_1(t), \mathbf{x}_2(t), \mathbf{x}_3(t))$ is approximately given as

$$\begin{aligned}&p(\mathbf{x}_1(t), \mathbf{x}_2(t), \mathbf{x}_3(t))\\ &\approx \sum_i w^{(i)}\,\mathcal{N}(\mathbf{x}_1(t)\,|\,\mathbf{m}^{(i)}, \mathbf{P}^{(i)})\,\delta(\mathbf{x}_2(t) - \mathbf{x}_2^{(i)})\,\delta(\mathbf{x}_3(t) - \mathbf{x}_3^{(i)}).\end{aligned} \tag{16}$$

where $\delta(\cdot)$ is the Dirac delta function. If the measurement model is of the form

$$p(\mathbf{y}_k\,|\,\mathbf{x}(t_k)) = \mathcal{N}\left(\mathbf{y}_k\,|\,\mathbf{H}_k\left(\mathbf{x}_2(t_k), \mathbf{x}_3(t_k)\right)\,\mathbf{x}_1(t_k), \mathbf{R}_k\left(\mathbf{x}_2(t_k), \mathbf{x}_3(t_k)\right)\right), \tag{17}$$

then conditional on $\mathbf{x}_2(t_k), \mathbf{x}_3(t_k)$ also the measurement model is linear Gaussian and the Kalman filter update equations can be applied. The resulting algorithm is the following:

Algorithm 2.3 *(CDRB-SIR I).* Given the importance process, a set of importance samples $\{\mathbf{x}_{2,k-1}^{(i)}, \mathbf{x}_{3,k-1}^{(i)}, \mathbf{m}_{k-1}^{(i)}, \mathbf{P}_{k-1}^{(i)}, w_{k-1}^{(i)} \;:\; i = 1, \ldots, N\}$ and the measurement $\mathbf{y}_k$, a single step of *conditional Gaussian continuous-discrete Rao-Blackwellized SIR* is the following:

1. Simulate N realizations of the importance process

$$
\begin{aligned}
\frac{d\mathbf{m}_s^{(i)}}{dt} &= \mathbf{F}(\mathbf{s}_2^{(i)}(t), \mathbf{s}_3^{(i)}, t)\, \mathbf{m}_s^{(i)}(t) + \mathbf{f}_1(\mathbf{s}_2^{(i)}, \mathbf{s}_3^{(i)}, t) \\
\frac{d\mathbf{P}_s^{(i)}}{dt} &= \mathbf{F}(\mathbf{s}_2^{(i)}, \mathbf{s}_3^{(i)}, t)\, \mathbf{P}_s^{(i)}(t) + \mathbf{P}_s^{(i)}(t)\, \mathbf{F}^T(\mathbf{s}_2^{(i)}, \mathbf{s}_3^{(i)}, t) \\
&\quad + \mathbf{V}(\mathbf{s}_2^{(i)}, \mathbf{s}_3^{(i)}, t)\, \mathbf{Q}_\eta(t)\, \mathbf{V}^T(\mathbf{s}_2^{(i)}, \mathbf{s}_3^{(i)}, t) \\
\frac{d\mathbf{s}_2^{(i)}}{dt} &= \mathbf{f}_2(\mathbf{s}_2^{(i)}, \mathbf{s}_3^{(i)}, t) \\
d\mathbf{s}_3^{(i)} &= \mathbf{g}_3(\mathbf{s}_2^{(i)}, \mathbf{s}_3^{(i)}, t)\, dt + \mathbf{B}(t)\, d\boldsymbol{\beta}^{(i)},
\end{aligned} \tag{18}
$$

with initial conditions

$$
\begin{aligned}
\mathbf{m}_s^{(i)}(t_{k-1}) &= \mathbf{m}_{k-1}^{(i)} \\
\mathbf{P}_s^{(i)}(t_{k-1}) &= \mathbf{P}_{k-1}^{(i)} \\
\mathbf{s}_2^{(i)}(t_{k-1}) &= \mathbf{x}_{2,k-1}^{(i)} \\
\mathbf{s}_3^{(i)}(t_{k-1}) &= \mathbf{x}_{3,k-1}^{(i)},
\end{aligned} \tag{19}
$$

2. Simulate the scaled importance process

$$
\begin{aligned}
\frac{d\mathbf{m}_s^{*(i)}}{dt} &= \mathbf{F}(\mathbf{s}_2^{*(i)}(t), \mathbf{s}_3^{*(i)}, t)\, \mathbf{m}_s^{*(i)}(t) + \mathbf{f}_1(\mathbf{s}_2^{*(i)}, \mathbf{s}_3^{*(i)}, t) \\
\frac{d\mathbf{P}_s^{*(i)}}{dt} &= \mathbf{F}(\mathbf{s}_2^{*(i)}, \mathbf{s}_3^{*(i)}, t)\, \mathbf{P}_s^{*(i)}(t) + \mathbf{P}_s^{*(i)}(t)\, \mathbf{F}^T(\mathbf{s}_2^{*(i)}, \mathbf{s}_3^{*(i)}, t) \\
&\quad + \mathbf{V}(\mathbf{s}_2^{*(i)}, \mathbf{s}_3^{*(i)}, t)\, \mathbf{Q}_\eta(t)\, \mathbf{V}^T(\mathbf{s}_2^{*(i)}, \mathbf{s}_3^{*(i)}, t) \\
\frac{d\mathbf{s}_2^{*(i)}}{dt} &= \mathbf{f}_2(\mathbf{s}_2^{*(i)}, \mathbf{s}_3^{*(i)}, t) \\
d\mathbf{s}_3^{*(i)} &= \mathbf{L}(t)\, \mathbf{B}^{-1}(t)\, d\mathbf{s}_3,
\end{aligned} \tag{20}
$$

with the same initial conditions from $t = t_{k-1}$ to $t = t_k$ and set

$$
\begin{aligned}
\mathbf{m}_k^{-(i)} &= \mathbf{m}_s^{*(i)}(t_k) \\
\mathbf{P}_k^{-(i)} &= \mathbf{P}_s^{*(i)}(t_k) \\
\mathbf{x}_{2,k}^{(i)} &= \mathbf{s}_2^{*(i)}(t_k) \\
\mathbf{x}_{3,k}^{(i)} &= \mathbf{s}_3^{*(i)}(t_k).
\end{aligned} \tag{21}
$$

3. Simulate the log-likelihood ratios (again, using the same Brownian motion real-

izations as in importance process)

$$
\begin{aligned}
d\Lambda^{(i)} = & \left\{ \mathbf{f}_3(\mathbf{s}_2^{*(i)}(t), \mathbf{s}_3^{*(i)}(t), t) - \mathbf{L}(t)\, \mathbf{B}^{-1}(t)\, \mathbf{g}_3(\mathbf{s}_2^{(i)}(t), \mathbf{s}_3^{(i)}(t), t) \right\}^T \\
& \times \mathbf{L}^{-T}(t)\, \mathbf{Q}_\beta^{-1}(t)\, d\boldsymbol{\beta}^{(i)}(t) \\
& - \frac{1}{2} \left\{ \mathbf{f}_3(\mathbf{s}_2^{*(i)}(t), \mathbf{s}_3^{*(i)}(t), t) - \mathbf{L}(t)\, \mathbf{B}^{-1}(t)\, \mathbf{g}_3(\mathbf{s}_2^{(i)}(t), \mathbf{s}_3^{(i)}(t), t) \right\}^T \\
& \times \left\{ \mathbf{L}(t)\, \mathbf{Q}_\beta(t)\, \mathbf{L}^T(t) \right\}^{-1} \\
& \times \left\{ \mathbf{f}_3(\mathbf{s}_2^{*(i)}(t), \mathbf{s}_3^{*(i)}(t), t) - \mathbf{L}(t)\, \mathbf{B}^{-1}(t)\, \mathbf{g}_3(\mathbf{s}_2^{(i)}(t), \mathbf{s}_3^{(i)}(t), t) \right\} dt, \\
\Lambda^{(i)}(t_{k-1}) = & \; 0,
\end{aligned}
\tag{22}
$$

and set

$$
Z_k^{(i)} = \exp\left\{ \Lambda^{(i)}(t_k) \right\} \tag{23}
$$

4. For each i perform the Kalman filter update

$$
\begin{aligned}
\boldsymbol{\mu}_k^{(i)} &= \mathbf{H}_k(\mathbf{x}_{2,k}^{(i)}, \mathbf{x}_{3,k}^{(i)})\, \mathbf{m}_k^{-(i)} \\
\mathbf{S}_k^{(i)} &= \mathbf{H}_k(\mathbf{x}_{2,k}^{(i)}, \mathbf{x}_{3,k}^{(i)})\, \mathbf{P}_k^{-(i)}\, \mathbf{H}_k^T(\mathbf{x}_{2,k}^{(i)}, \mathbf{x}_{3,k}^{(i)}) + \mathbf{R}_k(\mathbf{x}_{2,k}^{(i)}, \mathbf{x}_{3,k}^{(i)}) \\
\mathbf{K}_k^{(i)} &= \mathbf{P}_k^{-(i)}\, \mathbf{H}_k^T(\mathbf{x}_{2,k}^{(i)}, \mathbf{x}_{3,k}^{(i)})\, \{\mathbf{S}_k^{(i)}\}^{-1} \\
\mathbf{m}_k^{(i)} &= \mathbf{m}_k^{-(i)} + \mathbf{K}_k^{(i)}\, (\mathbf{y}_k - \boldsymbol{\mu}_k^{(i)}) \\
\mathbf{P}_k^{(i)} &= \mathbf{P}_k^{-(i)} - \mathbf{K}_k^{(i)}\, \mathbf{S}_k^{(i)}\, \{\mathbf{K}_k^{(i)}\}^T,
\end{aligned}
\tag{24}
$$

compute the importance weight

$$
w_k^{(i)} = w_{k-1}^{(i)}\, Z_k^{(i)}\, \mathcal{N}(\mathbf{y}_k \,|\, \boldsymbol{\mu}_k^{(i)}, \mathbf{S}_k^{(i)}), \tag{25}
$$

and re-normalize the weights to sum to unity.

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

The importance process can be formed, for example, by computing a joint Gaussian approximation by EKF and then extracting only the estimates corresponding to the innermost process. Note that the Rao-Blackwellization procedure can often be performed approximately, even when the model is not completely Gaussian. The Kalman filter steps can be replaced with the corresponding steps of EKF, when the model is slightly non-linear. This approach has been successfully applied in the context of multiple target tracking by Särkkä et al. (2007).

2.4 Rao-Blackwellization of Models with Static Parameters

Analogously to the discrete time case presented in Storvik (2002), 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 $\mathcal{T}_k = \mathcal{T}_k(\mathbf{x}(t_1), \ldots, \mathbf{x}(t_k), \mathbf{y}_{1:k})$, the parameter can be marginalized out analytically and only the state needs to be sampled.

These models arise, for example, when the measurement noise variance or some other other parameters of the measurement model are unknown. Two models of this kind are presented in Section 3.1.

Assume that the model is of the form

$$\begin{aligned} d\mathbf{x} &= \mathbf{f}(\mathbf{x}, t, \boldsymbol{\theta})\, dt + \mathbf{L}(t, \boldsymbol{\theta})\, d\boldsymbol{\beta} \\ \mathbf{y}_k &\sim p(\mathbf{y}_k \,|\, \mathbf{x}(t_k), \boldsymbol{\theta}), \end{aligned} \tag{26}$$

where $\boldsymbol{\theta}$ is an unknown static parameter. Also assume that $\mathbf{f}(\cdot)$ and $\mathbf{L}(\cdot)$ are of such form that the model is either non-singular or a singular model of the type considered in Section 2.2.

Now assume that the prior distribution of $\boldsymbol{\theta}$ has some finite dimensional sufficient statistics $\mathcal{T}_0$:

$$p(\boldsymbol{\theta}) = p(\boldsymbol{\theta} \,|\, \mathcal{T}_0). \tag{27}$$

Also assume that conditional posterior distribution of $\boldsymbol{\theta}$ has sufficient statistics $\mathcal{T}_k = \mathcal{T}_k(\mathbf{x}(t_1), \ldots, \mathbf{x}(t_k), \mathbf{y}_{1:k})$ of the same dimensionality as $\mathcal{T}_0$

$$p(\boldsymbol{\theta} \,|\, \mathbf{x}(t_1), \ldots, \mathbf{x}(t_k), \mathbf{y}_{1:k}) = p(\boldsymbol{\theta} \,|\, \mathcal{T}_k), \tag{28}$$

such that there exists an algorithm $\Phi(\cdot)$ that can be used for efficiently performing the update

$$\mathcal{T}_k = \Phi\left(\mathcal{T}_{k-1}, \mathbf{x}(t_k), \mathbf{y}_k\right). \tag{29}$$

Further assume that the marginal likelihood

$$p(\mathbf{y}_k \,|\, \mathbf{x}(t_k), \mathcal{T}_{k-1}) = \int p(\mathbf{y}_k \,|\, \mathbf{x}(t_k), \boldsymbol{\theta})\, p(\boldsymbol{\theta} \,|\, \mathcal{T}_{k-1})\, d\boldsymbol{\theta}, \tag{30}$$

can be efficiently evaluated. The above conditions are met, for example, when for fixed $\mathbf{x}(t_k)$ the distribution $p(\boldsymbol{\theta} \,|\, \mathcal{T}_{k-1})$ is conjugate to the likelihood $p(\mathbf{y}_k \,|\, \mathbf{x}(t_k), \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$.

The resulting algorithm is now the following:

Algorithm 2.4 *(CDRB-SIR II).* Given the importance process, a weighted set of samples $\{\mathbf{x}_{k-1}^{(i)}, \mathcal{T}_{k-1}^{(i)}, w_{k-1}^{(i)}\}$ and the new measurement $\mathbf{y}_k$, a single step of *continuous-discrete Rao-Blackwellized SIR with static parameters* can be performed as follows:

1. Simulate the importance process, scaled importance process and log-likelihood ratio as in Algorithm 2.1 or 2.2. This results in the sample set $\{\mathbf{x}_k^{(i)}\}$ and likelihood ratios $\{Z_k^{(i)}\}$.
2. For each i compute new sufficient statistics

$$\mathcal{T}_k^{(i)} = \Phi\left(\mathcal{T}_{k-1}^{(i)}, \mathbf{x}_k^{(i)}, \mathbf{y}_k\right), \tag{31}$$

evaluate the importance weights as

$$w_k^{(i)} = w_{k-1}^{(i)} \, Z_k^{(i)} \, p(\mathbf{y}_k \,|\, \mathbf{x}_k^{(i)}, \mathcal{T}_{k-1}^{(i)}), \tag{32}$$

and re-normalize the weights to sum to unity.

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

Actually, the sufficient statistics could be functionals of the whole state trajectory, in which case they could be simulated together with the state.

3 Numerical Simulations

In this section the continuous-discrete sequential importance sampling is applied to estimation of a partially measured simple pendulum which is distorted by a random noise term and to estimation of the spread of an infectious disease. Several other applications and more details on the presented applications can be found in the doctoral dissertation of Särkkä (2006b).

3.1 Simple Pendulum with Noise

The stochastic differential equation for the angular position of a simple pendulum (Alonso and Finn 1980), which is distorted by random white noise accelerations $w(t)$ with spectral density q can be written as

$$\frac{d^2 x}{dt^2} + a^2 \, \sin(x) = w(t), \tag{33}$$

where a is the angular velocity of the (linearized) pendulum. If we define the state as $\mathbf{x} = (x \;\, dx/dt)^T$ and change to state space form and to the integral equation notation in terms of Brownian motion, the model can be written as

$$\begin{aligned} \frac{dx_1}{dt} &= x_2 \\ dx_2 &= -a^2 \, \sin(x_1) \, dt + d\beta, \end{aligned} \tag{34}$$

where $\beta(t)$ has the diffusion coefficient q, which is a model of the form (6). 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

$$\begin{aligned} y_k &\sim \mathcal{N}(x_1(t_k), \sigma^2) \\ \sigma^2 &\sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2), \end{aligned} \tag{35}$$

This is now a model with an unknown static parameter as discussed in Section 2.4.

The importance process can now be formed by the continuous-discrete extended Kalman filter (EKF) (see, e.g., Jazwinski 1970; Gelb 1974) and the result is a two-dimensional Gaussian approximation for the joint distribution of the state $\mathbf{x}(t_k) = (x_1(t_k)\ x_2(t_k))^T$. Forming this approximation requires that the variance σ^2 is assumed to be known, but fortunately a very rough approximation based on the estimated σ_k^2 is enough in practice. Therefore the EKF based approximation can be constructed as follows:

1. Assume that the posterior distribution of a particle $\mathbf{x}^{(i)}(t)$ is approximately Gaussian
$$\mathbf{x}^{(i)}(t) \,|\, \mathbf{y}_{1:k-1} \sim \mathcal{N}(\mathbf{m}(t), \mathbf{P}(t)). \tag{36}$$
Note that immediately after a measurement, a single sampled particle actually has a Dirac delta distribution, which also is a (degenerate) Gaussian distribution.

2. By forming a first order Taylor series expansion of the right hand side of the equation (34) we get that after a sufficiently small time interval δt the state mean and covariance can be approximated as
$$\begin{aligned}\mathbf{m}(t+\delta t) &= \mathbf{m}(t) + \mathbf{f}(\mathbf{m}(t))\,\delta t \\ \mathbf{P}(t+\delta t) &= \mathbf{P}(t) + \left[\mathbf{F}(\mathbf{m}(t))\,\mathbf{P}(t) + \mathbf{P}(t)\,\mathbf{F}^T(\mathbf{m}(t)) + \mathbf{Q}\right]\,\delta t,\end{aligned} \tag{37}$$
where $\mathbf{f}(\mathbf{x}) = (x_2\ \ -a^2 \sin(x_1))^T$, $\mathbf{F}(\mathbf{x})$ is the Jacobian matrix of $\mathbf{f}(\mathbf{x})$ and $\mathbf{Q} = \mathrm{diag}(0\ q)$.

3. We may now form Gaussian approximation to the state at time $t + \delta t$ with the mean and covariance above. If we continue this process recursively and take the limit $\delta t \to 0$, we may then approximate the process as Gaussian process with mean and covariance
$$\begin{aligned}\frac{d\mathbf{m}(t)}{dt} &= \mathbf{f}(\mathbf{m}(t)) \\ \frac{d\mathbf{P}(t)}{dt} &= \mathbf{F}(\mathbf{m}(t))\,\mathbf{P}(t) + \mathbf{P}(t)\,\mathbf{F}^T(\mathbf{m}(t)) + \mathbf{Q},\end{aligned} \tag{38}$$

The above result states that between the measurements we can approximate the mean and covariance of the process (34) by integrating the deterministic differential equations (38). The result is a Gaussian process, that is, a Gaussian approximation to the state process at any instance of time.

The importance process can be now constructed as follows. For each particle i do the following:

1. Solve the approximate predicted mean and covariance at time t_k from the differential equations (38) by starting from initial conditions $\mathbf{m}(t_{k-1}) = \mathbf{x}_{k-1}^{(i)}$, $\mathbf{P}(t_{k-1}) = \mathbf{0}$.

2. Assuming that σ^2 is known, the approximate joint distribution of the state and measurement is Gaussian and thus we can compute the posterior distribution of the state in closed form.

If the resulting approximate marginal posterior mean and covariance of $x_2(t_k)$ are $m_{2,k}$ and $P_{22,k}$, then a suitable importance process is (assuming that sampling interval is Δt)

$$\begin{aligned} \frac{ds_1}{dt} &= s_2 \\ ds_2 &= \left(\frac{m_{2,k} - x_{2,k-1}}{\Delta t} \right) dt + \sqrt{\frac{P_{22,k}}{q\,\Delta t}}\, d\beta, \end{aligned} \tag{39}$$

with initial conditions

$$\begin{aligned} s_1(t_{k-1}) &= x_{1,k-1} \\ s_2(t_{k-1}) &= x_{2,k-1}. \end{aligned} \tag{40}$$

The equations for the scaled importance process can be now written as

$$\begin{aligned} \frac{ds_1^*}{dt} &= s_2^* \\ ds_2^* &= \left(\sqrt{\frac{q}{P_{22,k}\,\Delta t}} \right) (m_{2,k} - x_{2,k-1})\; dt + d\beta, \end{aligned} \tag{41}$$

with initial conditions

$$\begin{aligned} s_1^*(t_{k-1}) &= x_{1,k-1} \\ s_2^*(t_{k-1}) &= x_{2,k-1}. \end{aligned} \tag{42}$$

The full state of the algorithm at time step $k-1$ consists of the set of particles

$$\{w_{k-1}^{(i)}, x_{1,k-1}^{(i)}, x_{2,k-1}^{(i)}, \nu_{k-1}^{(i)}, \sigma_{k-1}^{2,(i)}\} \tag{43}$$

where $w_{k-1}^{(i)}$ is the importance weight, $x_{1,k-1}^{(i)}, x_{2,k-1}^{(i)}$ is the state of the pendulum, and $\nu_{k-1}^{(i)}, \sigma_{k-1}^{2,(i)}$ are the sufficient statistics of the variance parameter.

Figure 1 shows the result of applying the continuous-discrete particle filter with EKF proposal and 1000 particles to 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.

In the simulation, the true measurement variance was $\sigma^2 = 0.25$. The prior distribution used for the unknown variance parameter was $\sigma^2 \sim \text{Inv-}\chi^2(2, 0.2)$.

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 concentrates on the neighborhood of the true value.

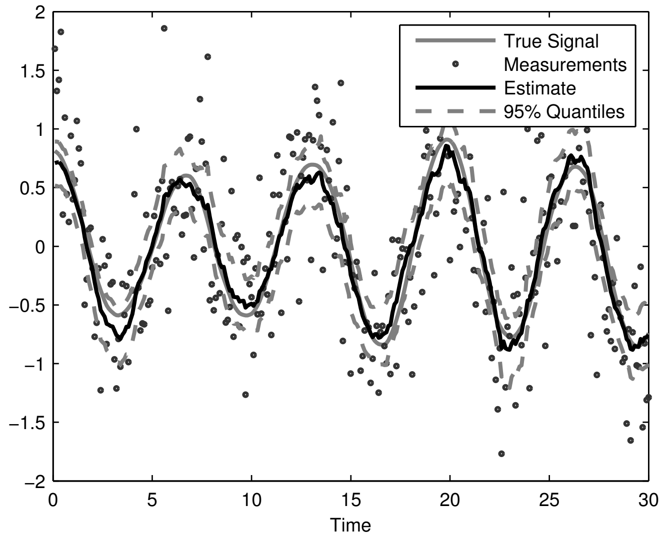

Figure 1: The result of applying a continuous-discrete particle filter with EKF proposal to simulated noisy pendulum data.

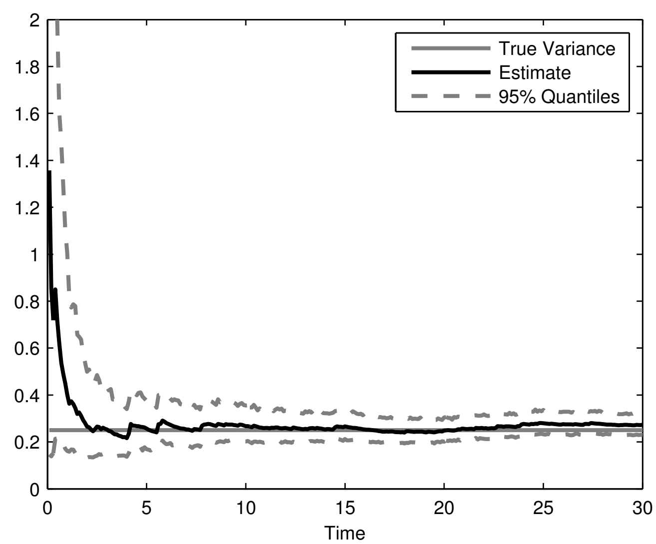

Figure 2: The evolution of the (posterior) variance distribution in the noisy pendulum problem.

3.2 Spread of Infectious Diseases

The classic model for the dynamics of infectious diseases is the SIR model (The model is called the SIR model, because the variables $X(t)$, $Y(t)$, and $Z(t)$ denote the susceptible, infective and removed compartments and for this reason are often denoted as $S(t)$, $I(t)$, and $R(t)$, respectively) (Kermack and McKendrick 1927; Anderson and May 1991; Murray 1993; Hethcote 2000), which is valid for sufficiently large N:

$$dX/dt = -b\,Y\,X/N, \qquad X(0) = X_0, \tag{44}$$

$$dY/dt = b\,Y\,X/N - g\,Y, \qquad Y(0) = Y_0, \tag{45}$$

$$dZ/dt = g\,Y, \qquad Z(0) = Z_0, \tag{46}$$

where $X(t)$ is the number of susceptibles at time t, $X_0 \geq 0$ is the initial number of susceptibles, $Y(t)$ is the number of infectives who are capable of transmitting the infection, $Y_0 \geq 0$ is the initial number of infectives, $Z(t)$ is the number of recovered or dead individuals which cannot be infected anymore, $Z_0 \geq 0$ is the initial number of individuals in this class, $N = X(t) + Y(t) + Z(t)$ is the (constant) total number of individuals, b is the contact rate which determines the rate of individuals moving from susceptible class to infectious class, and g is the waiting time parameter such that $1/g$ is the average length of the infectious period.

If we model the contact number $\sigma = b/g$ as exponential of a Brownian motion, then the stochastic equations for the proportions of individuals in each class can be written as (Särkkä 2006b):

$$\begin{aligned} dx/dt &= -g\,\exp(\lambda)\,y\,x \\ dy/dt &= g\,\exp(\lambda)\,y\,x - g\,y \\ d\lambda &= q^{1/2}\,d\beta, \end{aligned} \tag{47}$$

where $\beta(t)$ is a standard Brownian motion and $\lambda = \ln \sigma$.

A suitable initial distribution for $x(0)$ and $y(0)$ is

$$y(0) \sim \mathrm{Beta}(\alpha_y, \beta_y), \tag{48}$$

$$x(0) = 1 - y(0), \tag{49}$$

where $\beta_y \gg \alpha_y$. The initial conditions $z(0)$ can be assumed to be zero without loss of generality.

In the classical SIR model the values $X(t)$, $Y(t)$ and $Z(t)$ are not restricted to integer values, and as such they cannot be interpreted as counts. A sensible stochastic interpretation of these values is that they are the average numbers of individuals in each class and the actual numbers of individuals are Poisson distributed with these means.

Assume that the number of dead individuals is recorded. Then the number of the dead individuals d_k in time period $[t_{k-1}, t_k]$ has the distribution

$$p(d_k \mid \{x(\tau), y(\tau) : 0 \leq \tau \leq t_k\}, N) = \mathrm{Poisson}(d_k \mid N\,\theta_k), \tag{50}$$

where

$$\theta_k = x(t_{k-1}) - x(t_k) + y(t_{k-1}) - y(t_k). \tag{51}$$

The population size N is unknown and it can be modeled as having a Gamma prior distribution

$$p(N) = \text{Gamma}(N \mid \alpha_0, \beta_0), \tag{52}$$

with some suitably chosen α_0 and β_0. As shown in (Särkkä 2006b) this model is now of such a form that it is possible to integrate out the population size N from the equations and the Algorithm 2.4 can be applied.

The continuous-discrete SIR filter was applied to the classical Bombay plague data presented in (Kermack and McKendrick 1927). An EKF based Gaussian process approximation was used as the importance process (see, Särkkä 2006b, for details) and 10000 particles were used. The prior distribution for the proportion of initial infectives was Beta$(1, 100)$. The population size prior was Gamma$(10, 0.001)$. The waiting time parameter was assumed to be $g = 1$. The prior distribution for $\lambda(0)$ was $\mathcal{N}(\ln(5), 4)$. The diffusion coefficient of the Brownian motion was $q = 0.001$.

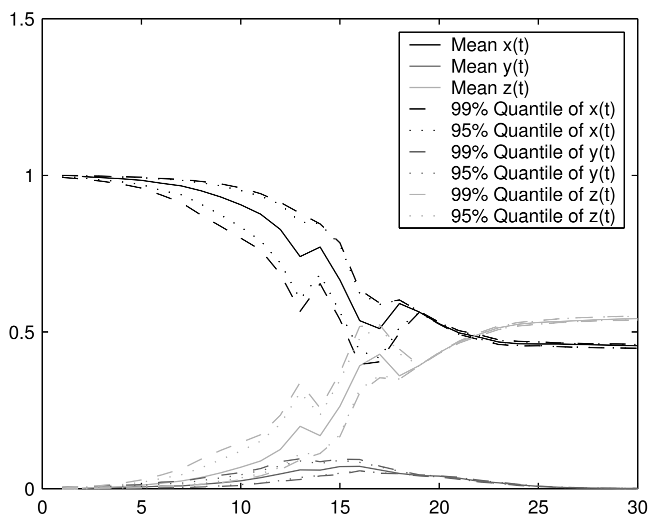

Figure 3: Filtered estimates of values of $x(t)$, $y(t)$, and $z(t)$ from the Bombay data.

The final filtered estimates of the histories of $x(t)$, $y(t)$, and $z(t)$ are shown in Figure 3. These estimates are filtered estimates, that is, they are conditional on the previously observed measurements only. That is, the estimate for week t is the estimate that could be actually computed on week t without any knowledge of the future observations. The estimates look like what would be expected: the proportion of susceptibles $x(t)$ decreases in time and the number of infectives $y(t)$ increases up to a maximum and then decreases to zero. However, these estimated values are not very useful themselves. The reason for this is that, for example, the value x_∞ which is the remaining value of susceptibles in the end depends on the choice of g and other prior parameters. That is, these estimated values are not absolute in the sense that their values depend heavily on the prior assumptions.

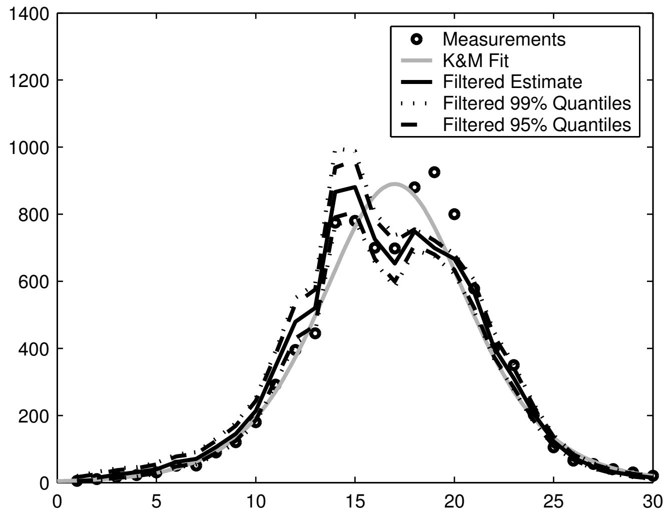

Figure 4: Filtered estimate of dZ/dt from the Bombay data. The estimate of (Kermack and McKendrick 1927) is also shown for comparison.

A much more informative quantity is the value dZ/dt, whose filtered estimate is shown in Figure 4. The classical estimate presented in (Kermack and McKendrick 1927) is also shown. The SIR filter estimate can be seen to differ a bit from the classical estimate, but still both the estimates look like what would be expected. Note that the classical estimate is based on all measurements, whereas the filtered estimate is based on observations made up to that time only. That is, the filter estimate could be actually computed in week t, but the classical estimate could not.

The filtered estimates of values $\sigma(t)$ are shown in Figure 5. The values can be seen to vary a bit on time, but the estimated expected value remains within the range $[1.4, 1.8]$ all the time. As can be seen from the figure, according to the data, the value of $\sigma(t)$ is not constant. This is not surprising, because the spatial and other unknown effects are not accounted for at all in the classical SIR model and these effects typically affect the number of contacts.

A very useful indicator value is $\sigma(t)\, x(t)$, whose filtered estimate is shown in Figure 6. In the deterministic SIR model with constant σ this indicator defines the asymptotic behavior of the epidemic (see, e.g., Hethcote 2000): If $\sigma x(t) \leq 1$ then the number of infectives will decrease to zero as $t \to \infty$. If $\sigma x(t) > 1$ then the number of infectives will first increase up to a maximum and then decrease to zero. As can be seen from the Figure 6 the filtered estimate of the indicator value goes below 1 just after the maximum somewhere between weeks 15–16, which can be seen in Figure 4. That is, the estimated value of $\sigma(t)\, x(t)$ could be used as an indicator, which tells if the epidemic is over or

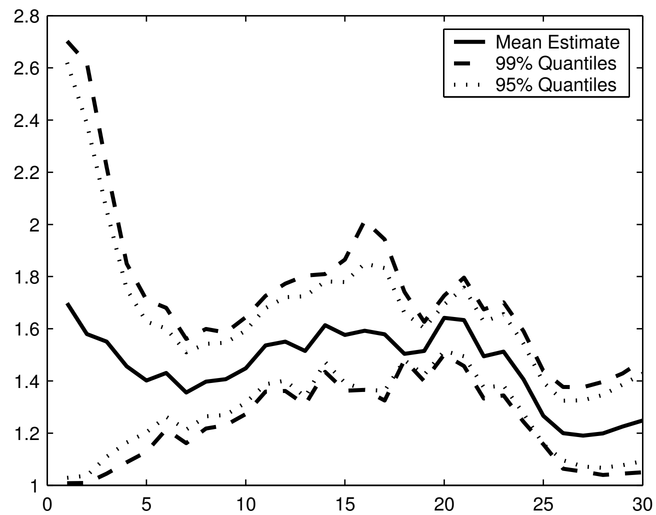

Figure 5: Bombay plague: Filtered estimate of values $\sigma(t)$.

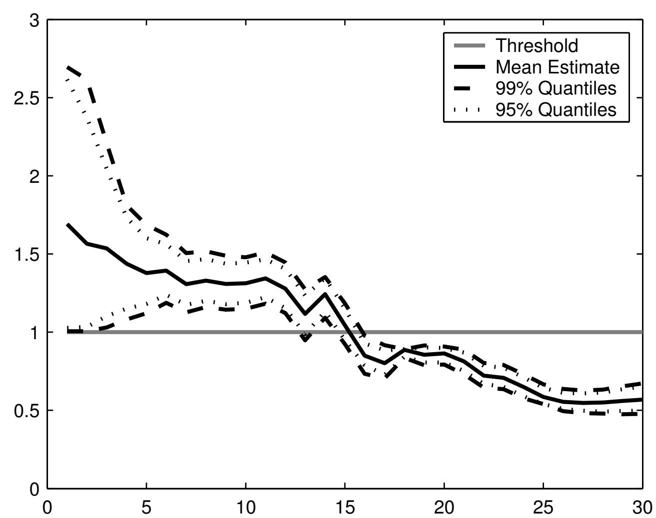

Figure 6: Bombay plague: Filtered estimate of values $x(t)\,\sigma(t)$.

not.

Using the particles it is also possible to predict ahead to the future and estimate the time when the maximum of the epidemic will be reached. The estimate computed from the filtering result is shown in the Figure 7. Again, the estimates are filtered estimates and the estimate for week t could be actually computed in week t, because

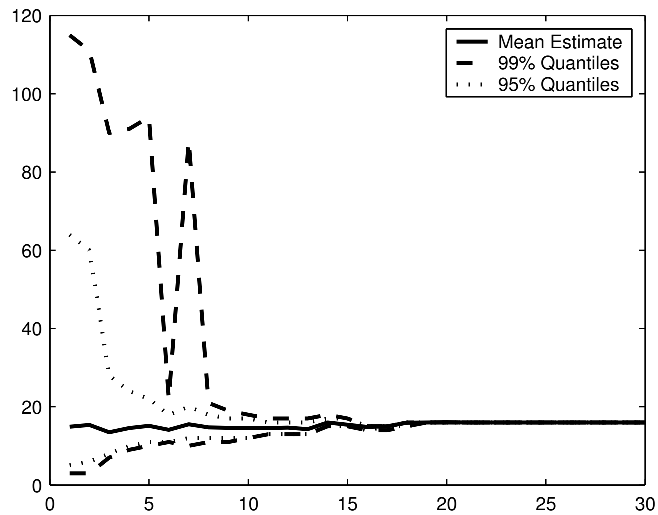

Figure 7: Bombay plague: Estimate of the time of the maximum of epidemic.

it depends only on the counts observed up to that time. The filtered estimate can be seen to quickly converge to the values near the correct maximum on weeks 15–16. It is interesting to see that the prediction is quite accurate already in week 10, which is far before reaching the actual maximum. If this kind of prediction had been done in, for example, week 10 of the disease, it would have predicted the time of actual epidemic maximum quite accurately. After the maximum has been reached, the estimate quickly converges to a constant value, which according to the Figure 4 is likely to be near the true maximum.

A very useful estimate is also the expected total number of deaths caused by the epidemic. This can be computed from the filtered estimates and the result is shown in Figure 8. In the beginning the estimate is very diffuse, but after the maximum has been reached the estimate converges to nearly the correct value. The estimate is a bit less than the observed value long before reaching the maximum, which might be due to existence of two maxima in the observed data (see, Figure 4). Because the second maximum is not predicted by the model, the extra number of deaths caused by it cannot be seen in the predictions.

4 Discussion

The importance processes used in the continuous-discrete particle filtering examples are very simple and better alternatives definitely exists. In principle, the optimal importance process in the continuous-discrete particle filtering case would have the same law as the smoothing solution. Thus, constructing the importance process based on the smoothing solution instead of linearly interpolated filtering solutions, as in this arti-

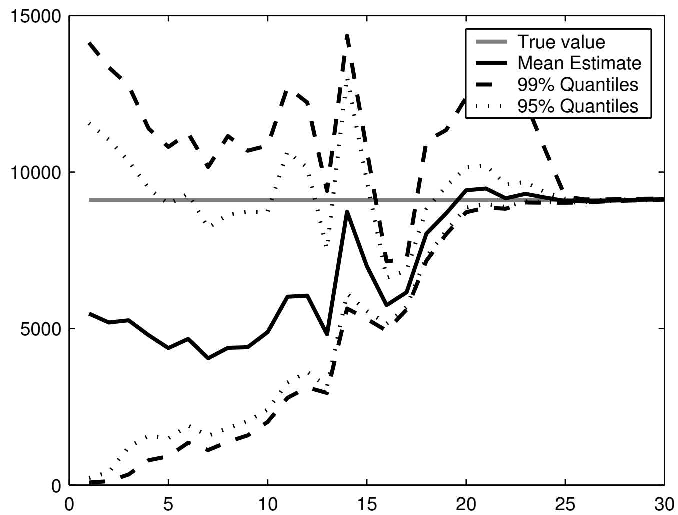

Figure 8: Bombay plague: Estimate of the number of deaths.

cle, could lead to more efficient particle filtering methods. In some cases it could be possible to construct a process, which would have exactly the same law as the optimal importance process.

A weakness in the continuous-discrete particle filtering framework is that the importance process has to be scaled before sampling. In practice, this restricts the possible forms of importance processes to those having the same dispersion matrix as the original process. However, we have not explicitly required that $\mathbf{L}(t) = \mathbf{B}(t)$, as this provides potential for the equation to be modified such that the scaling of the importance process would not be required.

Another weakness of the framework is that the time-discretization introduces bias to the estimation. The time-discretization is due to the usage of numerical integration methods for SDEs, which use discretization in time. However, there exists a method for simulating SDEs without time-discretization (Beskos et al. 2006) and maybe by using these methods this bias could be eliminated.

The continuous-discrete sequential importance resampling framework could be extended to the case of stochastic differential equations driven by more general martingales, for example, general Lévy processes such as compound Poisson processes (Applebaum 2004). This would allow modeling of sudden changes in signals. This extension could be possible by simply replacing the Brownian motion in the Girsanov theorem by a more general martingale.

It could be possible to generalize the continuous-discrete sequential importance sampling framework to continuous-time filtering problems. Then the extended Kalman-Bucy filter or the unscented Kalman-Bucy filter (Särkkä 2006b, 2007) could be used for forming the importance process and the actual filtering result would be formed by weighting the importance process samples properly.

The likelihood ratio expressions in Theorems 3 and 4 have an interesting connection to the variational method considered by Archambeau et al. (2007). If we select the processes as

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t), t)\, dt + \sqrt{\Sigma}\, d\boldsymbol{\beta}(t) \tag{53}$$

$$d\mathbf{s}(t) = \mathbf{f}_L(\mathbf{s}(t), t)\, dt + \sqrt{\Sigma}\, d\boldsymbol{\beta}(t), \tag{54}$$

where $\mathbf{x}(t)$ is a process with density $p(\cdot)$ and $\mathbf{s}(t)$ is a process with density $q(\cdot)$, then by taking the expectation of negative logarithm of (63) we get the expression for the KL-divergence between q and p:

$$\mathrm{KL}[q \,|\, p] = \mathrm{E}\Big[\frac{1}{2}\int_0^t \big\{\mathbf{f}(\mathbf{s}(t), t) - \mathbf{f}_L(\mathbf{s}(t), t)\big\}^T \Sigma^{-1} \big\{\mathbf{f}(\mathbf{s}(t), t) - \mathbf{f}_L(\mathbf{s}(t), t)\big\}\, dt\Big], \tag{55}$$

which is exactly the expression obtained heuristically in Archambeau et al. (2007). Thus the extensions to singular models would also apply to that method.

5 Conclusions

In this article, a new class of methods for continuous-discrete sequential importance sampling (particle filtering) has been presented. These methods are based on transformations of probability measures using 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. The methods have also been applied to estimation of the spread of an infectious disease based on counts of dead individuals.

The classical continuous-discrete extended Kalman filter as well as the recently developed continuous-discrete unscented Kalman filter can be used for forming importance processes for the new continuous-discrete particle filters. This way the efficiency of the Gaussian approximation based filters can be combined with the accuracy of the particle approximations. Closed form marginalization or Rao-Blackwellization can be applied if the model is conditional Gaussian or if the model contains unknown static parameters and has a suitable conjugate form. In most cases Rao-Blackwellization leads to a significant improvement in the efficiency of the particle filtering algorithm.

References

Alonso, M. and Finn, E. J. (1980). *Fundamental University Physics, Volume I: Mechanics and Thermodynamics*. Addison-Wesley, Reading, MA, 2nd edition. 567

Anderson, R. M. and May, R. M. (1991). *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, Oxford. 571

Applebaum, D. (2004). *Lévy Processes and Stochastic Calculus*. Cambridge University Press, Cambridge. 576

Archambeau, C., Cornford, D., Opper, M., and Shawe-Taylor, J. (2007). "Gaussian Process Approximations of Stochastic Differential Equations." *Journal of Machine Learning Research*, 1: 1–16. 577

Bar-Shalom, Y., Li, X.-R., and Kirubarajan, T. (2001). *Estimation with Applications to Tracking and Navigation*. Wiley, New York. 556, 560

Beskos, A., Papaspiliopoulos, O., Roberts, G., and Fearnhead, P. (2006). "Exact and Computationally Efficient Likelihood-based Estimation for Discretely Observed Diffusion Processes (with discussion)." *Journal of the Royal Statistical Society, Series B*, 68(3): 333 – 382. 556, 576

Chib, S., Pitt, M. K., and Shephard, N. (2004). "Likelihood based inference for diffusion driven models." *Economics Papers*, W20. 556

Crisan, D. (2003). "Exact Rates of Convergence for a Branching Particle Approximation to the Solution of the Zakai Equation." *Ann. Probab.*, 31(2): 693–718. 556

Crisan, D., Gaines, J., and Lyons, T. (1998). "Convergence of a Branching Particle Method to the Solution of the Zakai Equation." *SIAM Journal on Applied Mathematics*, 58(5): 1568–1590. 556

Crisan, D. and Lyons, T. (1999). "A particle approximation of the solution of the Kushner-Stratonovitch equation." *Probab. Theory Relat. Fields*, 115: 549–578. 556

Delyon, B. and Hu, Y. (2006). "Simulation of conditioned diffusion and application to parameter estimation." *Stochastic Processes and their Applications*, 116: 1660–1675. 581

Doucet, A., de Freitas, N., and Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice*. Springer, New York. 555, 556

Durham, G. B. and Gallant, A. R. (2002). "Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes." *Journal of Business & Economic Statistics*, 20(3): 297–316. 556

Fearnhead, P., Papaspiliopoulos, O., and Roberts, G. O. (2008). "Particle Filters for Partially Observed Diffusions." *Journal of the Royal Statistical Society, Series B*. To appear. 556

Gelb, A. (1974). *Applied Optimal Estimation*. The MIT Press, Cambridge, MA. 562, 568

Golightly, A. and Wilkinson, A. (2006). "Bayesian sequential inference for nonlinear multivariate diffusions." *Statist. Comput.*, 16(4): 323–338. 556

Gordon, N. J., Salmond, D. J., and Smith, A. F. M. (1993). "Novel approach to nonlinear/non-Gaussian Bayesian state estimation." In *IEEE Proceedings on Radar and Signal Processing*, volume 140, 107–113. 556

Grewal, M. S., Weill, L. R., and Andrews, A. P. (2001). *Global Positioning Systems, Inertial Navigation and Integration*. Wiley, New York. 556, 560

Hethcote, H. W. (2000). "The Mathematics of Infectious Diseases." *SIAM Review*, 42(4): 599–653. 571, 573

Ionides, E. L. (2004). "Inference and Filtering for Partially Observed Diffusion Processes via Sequential Monte Carlo." Technical report, University of Michigan Statistics Department Technical Report #405. 556

Jazwinski, A. H. (1966). "Filtering for nonlinear dynamical systems." *IEEE Transactions on Automatic Control*, 11(4): 765–766. 556

— (1970). *Stochastic Processes and Filtering Theory*. Academic Press, New York. 555, 556, 568

Kallianpur, G. (1980). *Stochastic Filtering Theory*. Springer-Verlag, New York. 556, 581

Karatzas, I. and Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York. 555, 556, 560, 581

Kermack, W. O. and McKendrick, A. C. (1927). "A Contribution to the Mathematical Theory of Epidemics." *Proceedings of the Royal Society of London, Series A*, 115: 700–721. 571, 572, 573

Kitagawa, G. (1996). "Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models." *Journal of Computational and Graphical Statistics*, 5: 1–25. 555, 556

Kloeden, P. E. and Platen, E. (1999). *Numerical Solution to Stochastic Differential Equations*. Springer, New York. 556, 557, 560

Moral, P. D. and Miclo, L. (2000). "Branching and interacting particle systems. Approximations of Feynman-Kac Formulae with applications to non-linear filtering." In Azéma, E. J., Emery, M., Ledoux, M., and Yor, M. (eds.), *Séminaire de Probabilités XXXIV*. Springer-Verlag, New York. 556

Murray, J. D. (1993). *Mathematical Biology*. Springer, New York. 571

Øksendal, B. (2003). *Stochastic Differential Equations: An Introduction with Applications*. Springer, New York, 6th edition. 555, 556, 560, 581

Pitt, M. K. and Shephard, N. (1999). "Filtering via Simulation: Auxiliary Particle Filters." *Journal of the American Statistical Association*, 94(446): 590–599. 556

Ristic, B., Arulampalam, S., and Gordon, N. (2004). *Beyond the Kalman Filter*. Artech House, Norwood, MA. 556

Särkkä, S. (2006a). "On Sequential Monte Carlo Sampling of Discretely Observed Stochastic Differential Equations." In *Proceedings of Nonlinear Statistical Signal Processing Workshop, Cambridge, September 2006*. 556

— (2006b). "Recursive Bayesian Inference on Stochastic Differential Equations." Doctoral dissertation, Helsinki University of Technology, Department of Electrical and Communications Engineering. 556, 567, 571, 572, 576

— (2007). "On Unscented Kalman Filtering for State Estimation of Continuous-Time Nonlinear Systems." *IEEE Transactions on Automatic Control*, 52(9): 1631–1641. 576

Särkkä, S., Vehtari, A., and Lampinen, J. (2007). "Rao-Blackwellized Particle Filter for Multiple Target Tracking." *Information Fusion Journal*, 8(1): 2–15. 565

Stengel, R. F. (1994). *Optimal Control and Estimation*. Dover Publications, New York. 556, 560

Storvik, G. (2002). "Particle filters in state space models with the presence of unknown static parameters." *IEEE Transactions on Signal Processing*, 50(2): 281–289. 565

Van Trees, H. L. (1968). *Detection, Estimation, and Modulation Theory Part I*. John Wiley & Sons, New York. 556, 560

— (1971). *Detection, Estimation, and Modulation Theory Part II*. John Wiley & Sons, New York. 556, 560

Appendices

1 Likelihood Ratios of SDEs

In the computation of the likelihood ratios of stochastic differential equations we need a slightly generalized version of the Girsanov theorem (Kallianpur 1980; Karatzas and Shreve 1991; Øksendal 2003). The generalized theorem can be obtained, for example, as a special case from the theorems presented in Delyon and Hu (2006).

Theorem 2 *(Girsanov).* Let $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)$ be a Brownian motion with diffusion matrix $\mathbf{Q}(t)$ under the probability measure $\mathbb{P}$. Let $\boldsymbol{\theta} : \Omega \times \mathbb{R}_+ \mapsto \mathbb{R}^d$ be an adapted process such that the process Z defined as

$$Z(t) = \exp\left\{ \int_0^t \boldsymbol{\theta}^T(t) d\boldsymbol{\beta}(t) - \frac{1}{2} \int_0^t \boldsymbol{\theta}^T(t)\, \mathbf{Q}(t)\, \boldsymbol{\theta}(t) dt \right\}, \tag{56}$$

satisfies $\mathrm{E}[Z(t)] = 1$. Then the process

$$d\tilde{\boldsymbol{\beta}}(t) = d\boldsymbol{\beta}(t) - \mathbf{Q}(t)\, \boldsymbol{\theta}(t) dt \tag{57}$$

is a Brownian motion with diffusion matrix $\mathbf{Q}(t)$ under the probability measure $\tilde{\mathbb{P}}$ defined via the relation

$$\mathrm{E}\left[\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} \,\middle|\, \mathcal{F}_t \right] = Z(t), \tag{58}$$

where $\mathcal{F}_t$ is the natural filtration of the Brownian motion $\boldsymbol{\beta}(t)$.

Proof. See, for example, Delyon and Hu (2006). □

Theorem 3 *(Transformation of SDE Solutions I).* Let

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t), t)\, dt + \mathbf{L}(t)\, d\boldsymbol{\beta}(t), \qquad \mathbf{x}(0) = x_0 \tag{59}$$

$$d\mathbf{s}(t) = \mathbf{g}(\mathbf{s}(t), t)\, dt + \mathbf{B}(t)\, d\boldsymbol{\beta}(t), \qquad \mathbf{s}(0) = x_0, \tag{60}$$

where $\boldsymbol{\beta}(t)$ is a Brownian motion with diffusion matrix $\mathbf{Q}(t)$ with respect to measure $\mathbb{P}$. Let $\mathcal{F}_t$ be its natural filtration. The matrices $\mathbf{L}(t)$ and $\mathbf{B}(t)$ are assumed to be invertible for all t. Now the process $\mathbf{s}^*(t)$ defined as

$$d\mathbf{s}^* = \mathbf{L}(t)\, \mathbf{B}^{-1}(t)\, d\mathbf{s}, \qquad \mathbf{s}(0) = \mathbf{x}_0 \tag{61}$$

is a weak solution to the Equation (59) under the measure $\tilde{\mathbb{P}}$ defined by the relation

$$\mathrm{E}\left[\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}\,\middle|\,\mathcal{F}_t\right] = Z(t), \tag{62}$$

where

$$\begin{aligned} Z(t) = \exp\Big[& \int_0^t \big\{\mathbf{f}(\mathbf{s}^*(t),t) - \mathbf{L}(t)\,\mathbf{B}^{-1}(t)\,\mathbf{g}(\mathbf{s}(t),t)\big\}^T\,\mathbf{L}^{-T}(t)\,\mathbf{Q}^{-1}(t)\,d\boldsymbol{\beta}(t) \\ & - \frac{1}{2}\int_0^t \big\{\mathbf{f}(\mathbf{s}^*(t),t) - \mathbf{L}(t)\,\mathbf{B}^{-1}(t)\,\mathbf{g}(\mathbf{s}(t),t)\big\}^T \\ & \quad \times \big\{\mathbf{L}(t)\,\mathbf{Q}(t)\,\mathbf{L}^T(t)\big\}^{-1}\big\{\mathbf{f}(\mathbf{s}^*(t),t) - \mathbf{L}(t)\,\mathbf{B}^{-1}(t)\,\mathbf{g}(\mathbf{s}(t),t)\big\}\,dt\Big] \end{aligned} \tag{63}$$

Proof. By substituting the expression (60) into Equation (61), solving for $d\boldsymbol{\beta}(t)$, we get

$$d\boldsymbol{\beta}(t) = \mathbf{L}^{-1}(t)\,d\mathbf{s}^* - \mathbf{B}^{-1}(t)\,\mathbf{g}(\mathbf{s}(t),t)\,dt. \tag{64}$$

If we now define

$$\boldsymbol{\theta}(t) = \mathbf{Q}^{-1}(t)\,\mathbf{L}^{-1}(t)\,\mathbf{f}(\mathbf{s}^*(t),t) - \mathbf{Q}^{-1}(t)\,\mathbf{B}^{-1}(t)\,\mathbf{g}(\mathbf{s}(t),t), \tag{65}$$

then under the measure $\tilde{\mathbb{P}}$ defined by (62) and (63) with the process $\boldsymbol{\theta}(t)$ defined as above, the following process is a Brownian motion with diffusion matrix $\mathbf{Q}(t)$:

$$\begin{aligned} d\tilde{\boldsymbol{\beta}}(t) &= d\boldsymbol{\beta}(t) - \mathbf{Q}(t)\,\boldsymbol{\theta}(t)dt \\ &= \mathbf{L}^{-1}(t)\,d\mathbf{s}^* - \mathbf{B}^{-1}(t)\,\mathbf{g}(\mathbf{s}(t),t)\,dt \\ &- \mathbf{Q}(t)\,\mathbf{Q}^{-1}(t)\,\mathbf{L}^{-1}(t)\,\mathbf{f}(\mathbf{s}^*(t),t)\,dt + \mathbf{Q}(t)\,\mathbf{Q}^{-1}(t)\,\mathbf{B}^{-1}(t)\,\mathbf{g}(\mathbf{s}(t),t)\,dt \\ &= \mathbf{L}^{-1}(t)\,d\mathbf{s}^* - \mathbf{L}^{-1}(t)\,\mathbf{f}(\mathbf{s}^*(t),t)\,dt \end{aligned} \tag{66}$$

By rearranging we get that

$$d\mathbf{s}^* = \mathbf{f}(\mathbf{s}^*(t),t)\,dt + \mathbf{L}(t)\,d\tilde{\boldsymbol{\beta}}(t) \tag{67}$$

and thus the result follows. The explicit expression for the likelihood ratio is given as follows:

$$
\begin{aligned}
Z(t) &= \exp\Big[\int_0^t \big\{\mathbf{Q}^{-1}(t)\,\mathbf{L}^{-1}(t)\,\mathbf{f}(\mathbf{s}^*(t),t) - \mathbf{Q}^{-1}(t)\,\mathbf{B}^{-1}(t)\,\mathbf{g}(\mathbf{s}(t),t)\big\}^T\, d\boldsymbol{\beta}(t) \\
&\quad - \frac{1}{2}\int_0^t \big\{\mathbf{Q}^{-1}(t)\,\mathbf{L}^{-1}(t)\,\mathbf{f}(\mathbf{s}^*(t),t) - \mathbf{Q}^{-1}(t)\,\mathbf{B}^{-1}(t)\,\mathbf{g}(\mathbf{s}(t),t)\big\}^T \\
&\quad \times \mathbf{Q}(t)\big\{\mathbf{Q}^{-1}(t)\,\mathbf{L}^{-1}(t)\,\mathbf{f}(\mathbf{s}^*(t),t) - \mathbf{Q}^{-1}(t)\,\mathbf{B}^{-1}(t)\,\mathbf{g}(\mathbf{s}(t),t)\big\}\, dt\Big] \\
&= \exp\Big[\int_0^t \big\{\mathbf{f}(\mathbf{s}^*(t),t) - \mathbf{L}(t)\,\mathbf{B}^{-1}(t)\,\mathbf{g}(\mathbf{s}(t),t)\big\}^T\,\mathbf{L}^{-T}(t)\,\mathbf{Q}^{-1}(t)\, d\boldsymbol{\beta}(t) \\
&\qquad - \frac{1}{2}\int_0^t \big\{\mathbf{L}^{-1}(t)\,\mathbf{f}(\mathbf{s}^*(t),t) - \mathbf{L}(t)\,\mathbf{B}^{-1}(t)\,\mathbf{g}(\mathbf{s}(t),t)\big\}^T \\
&\qquad \times \big\{\mathbf{L}^{-T}(t)\,\mathbf{Q}^{-1}(t)\,\mathbf{Q}(t)\mathbf{Q}(t)^{-1}\,\mathbf{L}^{-1}(t)\big\} \\
&\qquad \times \big\{\mathbf{f}(\mathbf{s}^*(t),t) - \mathbf{L}(t)\,\mathbf{B}^{-1}(t)\,\mathbf{g}(\mathbf{s}(t),t)\big\}\, dt\Big] \\
&= \exp\Big[\int_0^t \big\{\mathbf{f}(\mathbf{s}^*(t),t) - \mathbf{L}(t)\,\mathbf{B}^{-1}(t)\,\mathbf{g}(\mathbf{s}(t),t)\big\}^T\,\mathbf{L}^{-T}(t)\,\mathbf{Q}^{-1}(t)\, d\boldsymbol{\beta}(t) \\
&\qquad - \frac{1}{2}\int_0^t \big\{\mathbf{f}(\mathbf{s}^*(t),t) - \mathbf{L}(t)\,\mathbf{B}^{-1}(t)\,\mathbf{g}(\mathbf{s}(t),t)\big\}^T \\
&\qquad \times \big\{\mathbf{L}(t)\,\mathbf{Q}(t)\,\mathbf{L}^T(t)\big\}^{-1}\big\{\mathbf{f}(\mathbf{s}^*(t),t) - \mathbf{L}(t)\,\mathbf{B}^{-1}(t)\,\mathbf{g}(\mathbf{s}(t),t)\big\}\, dt\Big]
\end{aligned} \tag{68}
$$

□

Theorem 4 (Transformation of SDE Solutions II). Assume that processes $\mathbf{x}_1(t)$, $\mathbf{x}_2(t)$, $\mathbf{s}_1(t)$ and $\mathbf{s}_2(t)$ are generated by the stochastic differential equations

$$\frac{d\mathbf{x}_1}{dt} = \mathbf{f}_1(\mathbf{x}_1, \mathbf{x}_2, t), \qquad \mathbf{x}_1(0) = \mathbf{x}_{1,0} \tag{69}$$

$$d\mathbf{x}_2 = \mathbf{f}_2(\mathbf{x}_1, \mathbf{x}_2, t)\, dt + \mathbf{L}(t)\, d\boldsymbol{\beta}, \qquad \mathbf{x}_2(0) = \mathbf{x}_{2,0} \tag{70}$$

$$\frac{d\mathbf{s}_1}{dt} = \mathbf{f}_1(\mathbf{s}_1, \mathbf{s}_2, t), \qquad \mathbf{s}_1(0) = \mathbf{x}_{1,0} \tag{71}$$

$$d\mathbf{s}_2 = \mathbf{g}_2(\mathbf{s}_1, \mathbf{s}_2, t)\, dt + \mathbf{B}(t)\, d\boldsymbol{\beta}, \qquad \mathbf{s}_2(0) = \mathbf{x}_{2,0}, \tag{72}$$

where $\mathbf{L}(t)$ and $\mathbf{B}(t)$ are invertible matrices for all $t \geq 0$ and under the measure $\mathbb{P}$, $\boldsymbol{\beta}(t)$ is a Brownian motion with diffusion matrix $\mathbf{Q}(t)$. Then the processes $\mathbf{s}_1$ and $\mathbf{s}_2$ defined as

$$\frac{d\mathbf{s}_1^*}{dt} = \mathbf{f}_1(\mathbf{s}_1^*, \mathbf{s}_2^*, t), \qquad \mathbf{s}_1^*(0) = \mathbf{x}_{1,0} \tag{73}$$

$$d\mathbf{s}_2^* = \mathbf{L}(t)\,\mathbf{B}^{-1}(t)\, ds_2, \qquad \mathbf{s}_2^*(0) = \mathbf{x}_{2,0} \tag{74}$$

are weak solutions to the Equations (69) and (70) under the measure $\tilde{\mathbb{P}}$ defined by the relation

$$\mathrm{E}\left[\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}\,\middle|\,\mathcal{F}_t\right] = Z(t). \tag{75}$$

where

$$\begin{aligned} Z(t) = \exp \Big[\int_0^t & \big\{ \mathbf{f}_2(\mathbf{s}_1^*(t), \mathbf{s}_2^*(t), t) - \mathbf{L}(t)\, \mathbf{B}^{-1}(t)\, \mathbf{g}_2(\mathbf{s}_1(t), \mathbf{s}_2(t), t) \big\}^T \\ & \times \mathbf{L}^{-T}(t)\, \mathbf{Q}^{-1}(t)\, d\boldsymbol{\beta}(t) \\ - \frac{1}{2} \int_0^t & \big\{ \mathbf{f}_2(\mathbf{s}_1^*(t), \mathbf{s}_2^*(t), t) - \mathbf{L}(t)\, \mathbf{B}^{-1}(t)\, \mathbf{g}_2(\mathbf{s}_1(t), \mathbf{s}_2(t), t) \big\}^T \\ & \times \big\{ \mathbf{L}^T(t)\, \mathbf{Q}(t)\, \mathbf{L}(t) \big\}^{-1} \\ & \times \big\{ \mathbf{f}_2(\mathbf{s}_1^*(t), \mathbf{s}_2^*(t), t) - \mathbf{L}(t)\, \mathbf{B}^{-1}(t)\, \mathbf{g}_2(\mathbf{s}_1(t), \mathbf{s}_2(t), t) \big\}\, dt \Big] \end{aligned} \tag{76}$$

Proof. From equations (71), (72), (73) and (74) we get that

$$d\boldsymbol{\beta}(t) = \mathbf{L}^{-1}(t)\, d\mathbf{s}_2^* - \mathbf{B}^{-1}(t)\, \mathbf{g}_2(\mathbf{s}_1(t), \mathbf{s}_2(t), t)\, dt. \tag{77}$$

If we now define

$$\boldsymbol{\theta}(t) = \mathbf{Q}^{-1}(t)\, \mathbf{L}^{-1}(t)\, \mathbf{f}_2(\mathbf{s}_1^*(t), \mathbf{s}_2^*(t), t) - \mathbf{Q}^{-1}(t)\, \mathbf{B}^{-1}(t)\, \mathbf{g}_2(\mathbf{s}_1(t), \mathbf{s}_2(t), t), \tag{78}$$

then similarly as in proof of Theorem 3, we get that the process $\tilde{\boldsymbol{\beta}}(t)$ defined as

$$\begin{aligned} d\tilde{\boldsymbol{\beta}}(t) &= d\boldsymbol{\beta}(t) - \mathbf{Q}(t)\, \boldsymbol{\theta}(t) dt \\ &= \mathbf{L}^{-1}(t)\, d\mathbf{s}_2^* - \mathbf{B}^{-1}(t)\, \mathbf{g}_2(\mathbf{s}_1(t), \mathbf{s}_2(t), t)\, dt \\ &\quad - \mathbf{Q}(t)\, \mathbf{Q}^{-1}(t)\, \mathbf{L}^{-1}(t)\, \mathbf{f}_2(\mathbf{s}_1^*(t), \mathbf{s}_2^*(t), t)\, dt \\ &\quad + \mathbf{Q}(t)\, \mathbf{Q}^{-1}(t)\, \mathbf{B}^{-1}(t)\, \mathbf{g}_2(\mathbf{s}_1(t), \mathbf{s}_2(t), t)\, dt \\ &= \mathbf{L}^{-1}(t)\, d\mathbf{s}_2^* - \mathbf{L}^{-1}(t)\, \mathbf{f}_2(\mathbf{s}_1^*(t), \mathbf{s}_2^*(t), t)\, dt \end{aligned} \tag{79}$$

is a Brownian motion with respect to measure $\tilde{\mathbb{P}}$ and thus $\mathbf{s}_1^*$ and $\mathbf{s}_2^*$ are the weak solutions to the equations (69) and (70). □

Acknowledgments

The authors would like to thank Aki Vehtari, Jouko Lampinen, Ilkka Kalliomäki and Steve Ashton for helpful discussions and comments on the manuscript.