SEQUENTIAL PARTICLE FILTERING IN THE PRESENCE OF ADDITIVE GAUSSIAN NOISE WITH UNKNOWN PARAMETERS

Petar M. Djurić † *Joaquín Míguez**

†Department of Electrical and Computer Engineering, SUNY at Stony Brook, NY 11794, USA
* Dept. Electrónica e Sistemas, Universidade da Coruña Campus de Elviña s/n, 15071 A Coruña, Spain
e-mail: {djuric@ece.sunysb.edu, jmiguez@udc.es}

ABSTRACT

In sequential signal processing, the main objective is to estimate evolving states. Often, however, the models under consideration contain additional unknowns, which are time invariant. When the state estimation is carried out by sequential importance sampling methods, the presence of fixed unknowns can present a nontrivial problem. In this paper, we provide a solution to this problem when the fixed unknowns are the covariance matrices of the additive Gaussian noise vectors in the state and observation equations. These matrices are first marginalized, and then the sequential processing carried out as usual. In the implementation of this approach, besides the assignment of a weight to every particle, two additional evolving quantities are required. Simulation results are provided that show the performance of the method.

1. INTRODUCTION

Sequential importance sampling is a methodology for sequential signal processing that has drawn significant interest lately [4]. It is an attractive approach for filtering, prediction and smoothing in scenarios where the signal evolution is nonlinear and the noise is non-additive and non-Gaussian. It has been shown that in many important situations it shows considerable performance advantages over the standard extended Kalman filter method.

The sequential importance sampling methods are based on representing the densities of interest by a set of samples and their associated weights. When such sets provide good approximations of their densities, various types of estimators can be used to estimate the desired unknowns. Typically, the main object of interest in sequential signal processing problems are the time varying states. The time variation of the states is usually modeled as a Markov process and is represented by a state equation. The states are not observed directly, and instead, before measurements are made, they may undergo nonlinear transformations. The measurements can be distorted and therefore they are modeled as random variables. The state and measurement observations, thus, contain noise, which follows some postulated parametric distribution.

In many practical settings, the noise parameters are not known, and one approach that has been adopted in such situations is to expand the state vector with these parameters and carry out the sequential importance sampling in the usual way. It should be noted that there is an important difference between the original state variables and the noise parameters in that the latter do not change with time. This creates a major problem in implementing the method because there is no natural mechanism to regenerate the particles of the unknown fixed parameters. As a result, the weights that correspond to the compound state degenerate very quickly. One solution to this problem is to enforce artificial evolution of the fixed parameters [8].

Often, the noise parameters are not of interest and in such cases, if at all possible, they should be integrated out. In the important problem of state estimation where the noise in the state and observation equations is additive and Gaussian, this can be done. In this paper we provide the steps of the marginalization and derive the final update weight equation. Here, the marginalization is especially useful because it significantly decreases the dimension of the state space. This has a very positive impact on the implementation of the sequential importance sampler. The number of particles that have to be used in the approximations of the relevant densities can be drastically smaller. In addition, the variance of the estimated states is reduced considerably.

The paper is organized as follows. First we state the problem, and then we provide a brief overview of the sequential importance sampling method. The main section of the paper is Section 4, where the case of unknown covariance matrices is treated and the main result is provided. Simulation results that present the performance of the proposed sequential importance sampling method are given in Section 5. Comparisons are made with the the sequential importance sampling method that uses the correct values of the covariance matrices. Finally, we conclude the paper with brief remarks.

This work was supported by the National Science Foundation under Awards No. CCR-9903120 and CCR-0082607 (†) and Xunta de Galicia, grant number PGIDT00PXI10504PR (*).

2. PROBLEM FORMULATION

The models of interest in this paper have the form

$$\mathbf{x}_t = f_t(\mathbf{x}_{t-1}) + \mathbf{u}_t \tag{1}$$

$$\mathbf{y}_t = g_t(\mathbf{x}_t) + \mathbf{v}_t \tag{2}$$

where $t \in \mathbb{N}$ is a time index, $\mathbf{x}_t$ is a state vector at time t, $\mathbf{u}_t$ is a state noise vector, $\mathbf{y}_t$ is an observation at time t, and $\mathbf{v}_t$ is an observation noise vector. For the vectors we assume that $\mathbf{x}_t, \mathbf{u}_t \in \mathbb{R}^n$, and $\mathbf{y}_t, \mathbf{v}_t \in \mathbb{R}^m$. The symbols $f_t(\cdot)$ and $g_t(\cdot)$ represent known functions, which in general are nonlinear in $\mathbf{x}_{t-1}$ and $\mathbf{x}_t$, respectively. In addition, the noise vectors are modeled as Gaussian, i.e.,

$$\mathbf{u}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_u)$$

$$\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_v)$$

where $\mathbf{C}_u \in \mathbb{R}^{n\times n}$ and $\mathbf{C}_v \in \mathbb{R}^{m\times m}$ are positive definite matrices with $n(n+1)/2$ and $m(m+1)/2$ different elements, respectively. These matrices are *unknown*. Also, the noise vectors $\mathbf{u}_t$ and $\mathbf{v}_t$ are assumed independent. The objective is to estimate the unobserved states $\mathbf{x}_t$ recursively from the observations $\mathbf{y}_t$.

3. SEQUENTIAL IMPORTANCE SAMPLING

A widely accepted method for estimation of $\mathbf{x}_t$ is the extended Kalman filter (EKF). It is based on the linearization of (1) and (2) around the current values of $\mathbf{x}_t$ and $\mathbf{x}_{t-1}$ and application of the standard Kalman filter [1]. It has been shown, however, that in many applications the EKF is unreliable and produces poor results. In the literature, as alternatives, several other methods have been proposed, including Gaussian sum approximations [10] and grid-based filters [3]. At about the same time, a method based on sequential Monte Carlo integration was also proposed [6], but due to lack of computing power, it had a similar destiny as other numerically based approaches; it was almost forgotten. Later in the late eighties and early nineties, the interest in computationally intensive methods for sequential signal processing was rekindled [5], [7], and ever since, this interest has steadily increased.

In this paper, we address the sequential importance sampling method [4]. Here we briefly summarize it and outline its implementation. Let $\mathbf{x}_{1:t}$ and $\mathbf{y}_{1:t}$ denote the sets $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_t\}$ and $\{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_t\}$, respectively. When the data $\mathbf{y}_{1:t}$ have been observed, all the information about the states $\mathbf{x}_{1:t}$ is summarized by the posterior density $p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})$. If the only unknowns were the states, the recursive equation that describes the transition from $p(\mathbf{x}_{1:t-1}|\mathbf{y}_{1:t-1})$ to $p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})$ is given by

$$p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t|\mathbf{x}_{1:t}, \mathbf{y}_{1:t-1})p(\mathbf{x}_t|\mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})}$$

$$\times p(\mathbf{x}_{1:t-1}|\mathbf{y}_{1:t-1}). \tag{3}$$

The computation of the "new" given the "old" posterior in (3) is often intractable for analytical evaluation. In such cases, one usually resorts to a computational approach. Suppose that at time $t-1$, the posterior density $p(\mathbf{x}_{1:t-1}|\mathbf{y}_{1:t-1})$ is given by the set of particles and weights $\{\mathbf{x}_{1:t-1}^{(i)}, w_{t-1}^{(i)}\}_{i=1}^N$. One straightforward approach to obtaining the particles and weights $\{\mathbf{x}_{1:t}^{(i)}, w_t^{(i)}\}_{i=1}^N$ from $\{\mathbf{x}_{1:t-1}^{(i)}, w_{t-1}^{(i)}\}_{i=1}^N$ would be to apply the sequential importance sampling algorithm [4]. The algorithm runs as follows:

- Draw $\mathbf{x}_t^{(i)}$ from $\pi(\mathbf{x}_t|\mathbf{x}_{1:t-1}^{(i)}, y_{1:t})$, where $\pi(\cdot)$ is an importance function.
- Compute the importance weights $w_t^{*(i)}$ according to

$$w_t^{*(i)} = w_{t-1}^{(i)} \frac{p(\mathbf{y}_t|\mathbf{x}_{1:t}^{(i)}, \mathbf{y}_{1:t-1})p(\mathbf{x}_t^{(i)}|\mathbf{x}_{1:t-1}^{(i)}, \mathbf{y}_{1:t-1})}{\pi(\mathbf{x}_t^{(i)}|\mathbf{x}_{1:t-1}^{(i)}, \mathbf{y}_{1:t})}. \tag{4}$$

- Normalize the weights by

$$w_t^{(i)} = \frac{w_t^{*(i)}}{\sum_{j=1}^N w_t^{(j)}}.$$

4. THE CASE OF UNKNOWN COVARIANCE MATRICES

In our problem, the unknowns also include the covariance matrices $\mathbf{C}_u$ and $\mathbf{C}_v$. They are fixed, and do not evolve with time. Since they are not of interest, they are not estimated and are instead integrated out. We show how one can obtain the expressions for the factors in (4), which are needed for the computation of the new weights.

First, $\pi(\mathbf{x}_t^{(i)}|\mathbf{x}_{1:t-1}^{(i)}, \mathbf{y}_{1:t})$ is selected so that, besides providing good sample candidates, it is also easily computable. Second, for the factor $p(\mathbf{y}_t|\mathbf{x}_{1:t}^{(i)}, \mathbf{y}_{1:t-1})$ we can write

$$p(\mathbf{y}_t|\mathbf{x}_{1:t}^{(i)}, \mathbf{y}_{1:t-1}) = \int p(\mathbf{y}_t|\mathbf{x}_{1:t}^{(i)}, \mathbf{y}_{1:t-1}, \mathbf{C}_v)$$

$$\times\, p(\mathbf{C}_v|\mathbf{x}_{1:t}^{(i)}, \mathbf{y}_{1:t-1}) d\mathbf{C}_v \tag{5}$$

where

$$p(\mathbf{y}_t|\mathbf{x}_{1:t}^{(i)}, \mathbf{y}_{1:t-1}, \mathbf{C}_v) \propto \frac{1}{|\mathbf{C}_v|^{\frac{1}{2}}}$$

$$\times\, e^{-\frac{1}{2}(\mathbf{y}_t - g_t(\mathbf{x}_t))^{\mathrm{T}} \mathbf{C}_v^{-1} (\mathbf{y}_t - g_t(\mathbf{x}_t))}.$$

Another way of writing this density is

$$p(\mathbf{y}_t|\mathbf{x}_{1:t}^{(i)}, \mathbf{y}_{1:t-1}, \mathbf{C}_v) \propto \frac{1}{|\mathbf{C}_v|^{\frac{1}{2}}}$$

$$e^{-\frac{1}{2}\mathrm{tr}(\mathbf{C}_v^{-1}(\mathbf{y}_t - g(\mathbf{x}_t))(\mathbf{y}_t - g(\mathbf{x}_t))^{\mathrm{T}}}. \tag{6}$$

For the density $p(\mathbf{C}_v|\mathbf{x}_{1:t}^{(i)}, \mathbf{y}_{1:t-1})$, we have

$$p(\mathbf{C}_v|\mathbf{x}_{1:t}^{(i)}, \mathbf{y}_{1:t-1}) \propto p(\mathbf{y}_{1:t-1}|\mathbf{x}_{1:t-1}^{(i)}, \mathbf{C}_v)p(\mathbf{C}_v)$$

where $p(\mathbf{C}_v)$ is the prior of $\mathbf{C}_v$. We choose the noninformative prior for $\mathbf{C}_v$ [2], i.e.,

$$p(\mathbf{C}_v) \propto \frac{1}{|\mathbf{C}_v|^{\frac{m+1}{2}}}.$$

Then

$$p(\mathbf{C}_v|\mathbf{x}_{1:t}^{(i)}, \mathbf{y}_{1:t-1}) \propto \frac{1}{|\mathbf{C}_v|^{\frac{t-1}{2}}} e^{-\frac{1}{2}\mathrm{tr}(\mathbf{C}_v^{-1}\mathbf{R}_{t-1}^{(i)})} \frac{1}{|\mathbf{C}_v|^{\frac{m+1}{2}}}$$

where

$$\mathbf{R}_{t-1}^{(i)} = \sum_{j=1}^{t-1}(\mathbf{y}_j - g_j(\mathbf{x}_j^{(i)}))(\mathbf{y}_j - g_j(\mathbf{x}_j^{(i)}))^{\mathsf{T}}$$

or

$$p(\mathbf{C}_v|\mathbf{x}_{1:t}^{(i)}, \mathbf{y}_{1:t-1}) \propto \frac{1}{|\mathbf{C}_v|^{\frac{t+m}{2}}} e^{-\frac{1}{2}\mathrm{tr}(\mathbf{C}_v^{-1}\mathbf{R}_{t-1}^{(i)})}. \quad (7)$$

Next, we substitute (6) and (7) in (5) and obtain

$$p(\mathbf{y}_t|\mathbf{x}_{1:t}^{(i)}, \mathbf{y}_{1:t-1}) \propto \int \frac{1}{|\mathbf{C}_v|^{\frac{1}{2}}} \frac{1}{|\mathbf{C}_v|^{\frac{t+m}{2}}}$$

$$\times e^{-\frac{1}{2}\mathrm{tr}(\mathbf{C}_v^{-1}(\mathbf{y}_t - g(\mathbf{x}_t))(\mathbf{y}_t - g(\mathbf{x}_t))^{\mathsf{T}})} e^{-\frac{1}{2}\mathrm{tr}(\mathbf{C}_v^{-1}\mathbf{R}_{t-1}^{(i)})} d\mathbf{C}_v$$

or

$$p(\mathbf{y}_t|\mathbf{x}_{1:t}^{(i)}, \mathbf{y}_{1:t-1}) \propto \int \frac{1}{|\mathbf{C}_v|^{\frac{t+m+1}{2}}} e^{-\frac{1}{2}\mathrm{tr}(\mathbf{C}_v^{-1}\mathbf{R}_t^{(i)})} d\mathbf{C}_v.$$

Using the integral [2]

$$\int_{\mathbf{Z}>\mathbf{0}} |\mathbf{Z}|^{\frac{1}{2}k-1} e^{-\frac{1}{2}\mathrm{tr}\mathbf{Z}\mathbf{B}} d\mathbf{Z} = |\mathbf{B}|^{-\frac{1}{2}(k+m-1)} 2^{\frac{1}{2}m(k+m-1)}$$

$$\times \Gamma_m\left(\frac{k+m-1}{2}\right)$$

where $\mathbf{Z}$, $\mathbf{B}$ are positive definite symmetric $m \times m$ matrices, and $\Gamma_m(\cdot)$ is the generalized gamma function [9]

$$\Gamma_m(b) = \left(\Gamma\left(\frac{1}{2}\right)\right)^{\frac{1}{2}m(m-1)} \prod_{l=1}^{m} \Gamma\left(b + \frac{l-m}{2}\right)$$

where $b > \frac{m-1}{2}$, we can show that

$$p(\mathbf{y}_t|\mathbf{x}_{1:t}^{(i)}, \mathbf{y}_{1:t-1}) \propto |\mathbf{R}_t^{(i)}|^{-\frac{1}{2}t}.$$

The derivation of the expression for $p(\mathbf{x}_t^{(i)}|\mathbf{x}_{1:t-1}^{(i)}, \mathbf{y}_{1:t-1})$ is similar. The obtained result is

$$p(\mathbf{x}_t^{(i)}|\mathbf{x}_{1:t-1}^{(i)}, \mathbf{y}_{1:t-1}) \propto |\mathbf{Q}_t^{(i)}|^{-\frac{1}{2}t}$$

where

$$\mathbf{Q}_t^{(i)} = \sum_{j=1}^{t}(\mathbf{x}_j^{(i)} - f_j(\mathbf{x}_{j-1}^{(i)}))(\mathbf{x}_j^{(i)} - f_j(\mathbf{x}_{j-1}^{(i)}))^{\mathsf{T}}.$$

So, the computation of the weights simplifies to

$$w_t^{*(i)} = w_{t-1}^{(i)} \frac{|\mathbf{R}_t^{(i)}|^{-\frac{t}{2}} |\mathbf{Q}_t^{(i)}|^{-\frac{t}{2}}}{\pi(\mathbf{x}_t^{(i)}|\mathbf{x}_{1:t-1}^{(i)}, \mathbf{y}_{1:t})}. \quad (8)$$

Thus, in the implementation of the sequential importance sampling method, when keeping track of the particles and their weights, one also needs to monitor the quantities $\mathbf{R}_t^{(i)}$ and $\mathbf{Q}_t^{(i)}$. Modifications of (8) can be implemented to avoid the imminent problem of the monotonically increasing $\mathbf{R}_t^{(i)}$, $\mathbf{Q}_t^{(i)}$, and t.

5. SIMULATION RESULTS

We conducted several experiments in which we evaluated the performance of the sequential importance sampling method proposed here. We compared it with the performance of the sequential importance sampling approach when the covariance matrices are known. Some of the results of these experiments are shown here.

The model that was used in the simulations is given by

$$\begin{aligned} \mathbf{x}_t &= \mathbf{x}_{t-1} + \mathbf{u}_t \\ \mathbf{y}_t &= g(\mathbf{x}_t) + \mathbf{v}_t \\ &= \begin{bmatrix} e^{x_{t1}} \\ e^{x_{t2}} \\ \vdots \\ e^{x_{tm}} \end{bmatrix} + \begin{bmatrix} v_{t1} \\ v_{t1} \\ \vdots \\ v_{tn} \end{bmatrix}. \end{aligned}$$

We used $n = m = 2$, and the covariance matrices $\mathbf{C}_u$ and $\mathbf{C}_v$ were

$$\mathbf{C}_u = \begin{bmatrix} 0.020 & -0.014 \\ -0.014 & 0.020 \end{bmatrix} \quad \mathbf{C}_v = \begin{bmatrix} 0.05 & 0.01 \\ 0.01 & 0.05 \end{bmatrix}.$$

The realizations had $T = 100$ samples, and in the simulations we used $M = 300$ particles. In Figure 1 we displayed the two state trajectories $x_{1,1:t}$ and $x_{2,1:t}$ of a specific realization and their estimates obtained by the two methods. The figure also contains the graphs of the observations used for the estimation. In Figure 2, we provided statistical results of the performance of the methods. The subplots show the mean square error (MSE) as a function of time of the first and second state trajectories obtained by the two sequential importance sampling methods. The results were computed from 50 estimated trajectories of the state and observation realizations shown in Figure 1.

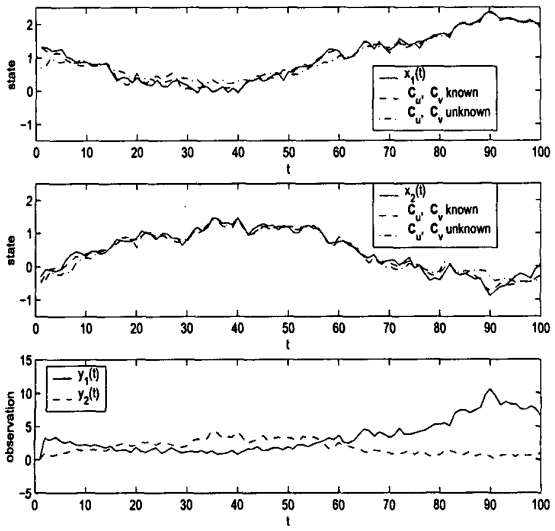

Fig. 1. Top plot: Trajectories of the first state and their estimates. Middle plot: Trajectories of the second state and their estimates. Bottom plot: Trajectories of the observations. One set of the estimates were obtained with the assumption that the covariance matrices were known, and the other, when they were unknown quantities.

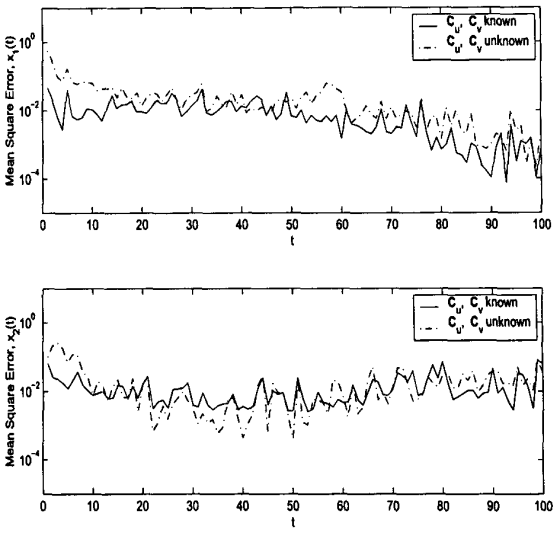

Fig. 2. MSE's obtained from 50 realizations as functions of time. They were computed for the realization shown in Figure 1. The applied methods were sequential importance sampling when the covariance matrices were known and when they were unknown quantities.

6. CONCLUSIONS

In most of the published papers on nonlinear state estimation problems, where the noise is additive and Gaussian, a typical assumption is that the noise parameters are known. Here we address the problem when this assumption is not valid. The noise parameters are not estimated but are integrated out, which decreases the complexity of the method substantially and reduces the variance of the estimates. Simulation results are provided that indicate the performance of the method.

7. REFERENCES

[1] B. D. Anderson and J. B. Moore, *Optimal Filtering*, Prentice-Hall, New Jersey, 1979.

[2] G. E. P. Box an G. C. Tiao, *Bayesian Inference in Statistical Analysis*, 1973.

[3] R. C. Bucy and K. D. Senne, "Digital synthesis of nonlinear filters," *Automatica*, vol. 7, pp. 287–298, 1971.

[4] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, Springer, New York, 2001.

[5] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings-F*, vol. 140, no. 2, pp. 107–113, Apr. 1993.

[6] J. E. Handschin and D. Q. Mayne, "Monte Carlo techniques to estimate the conditional expectation in multi-stage non-linear filtering," *International Journal of Control*, vol. 9, no. 5, pp. 547–559, 1969.

[7] G. Kitagawa, "Non-Gaussian state-space modeling of nonstationary time series," *Journal of the American Statistical Association*, vol. 82, pp. 1032–1063, 1987.

[8] J. Liu and M. West, "Combined parameter and state estimation in simulation based filtering," in *Sequential Monte Carlo Methods in Practice*, A. Doucet, N. de Freitas, and N. Gordon, Eds., chapter 10, pp. 197–213. Springer, 2001.

[9] C. L. Siegel, "Ueber die analytische theorie der quadratischen formen," *Ann. Math.*, vol. 36, pp. 527, 1935.

[10] H. W. Sorenson and D. L. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica*, vol. 7, pp. 465–479, 1971.