Ann. Inst. Statist. Math.
Vol. 53, No. 1, 97–112 (2001)

OPTIMAL ESTIMATION AND CRAMÉR-RAO BOUNDS FOR PARTIAL NON-GAUSSIAN STATE SPACE MODELS *

NICLAS BERGMAN[1]**, ARNAUD DOUCET[2] AND NEIL GORDON[3]***

[1]*Division of Automatic Control, Linköping University, S-581 83 Linköping, Sweden*
[2]*Signal Processing Group, Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, U.K.*, e-mail: ad2@eng.cam.ac.uk
[3]*Defence Evaluation and Research Agency, St. Andrews Road, Malvern, Worcestershire WR14 3PS, U.K.*, e-mail: N.Gordon@signal.dera.gov.uk

(Received May 1, 2000; revised August 4, 2000)

Abstract. Partial non-Gaussian state-space models include many models of interest while keeping a convenient analytical structure. In this paper, two problems related to partial non-Gaussian models are addressed. First, we present an efficient sequential Monte Carlo method to perform Bayesian inference. Second, we derive simple recursions to compute posterior Cramér-Rao bounds (PCRB). An application to jump Markov linear systems (JMLS) is given.

Key words and phrases: Optimal estimation, Bayesian inference, sequential Monte Carlo methods, posterior Cramér-Rao bounds.

1. Introduction

1.1 *Background*

A partial non-Gaussian state-space model is a linear model whose parameters evolve with time according to an unobserved stochastic process s_t. Let t denote the discrete-time index, then one has

$$x_t = A_t(s_t)x_{t-1} + B_t(s_t)v_t + F_t(s_t)u_t \tag{1.1}$$

$$y_t = C_t(s_t)x_t + D_t(s_t)w_t + G_t(s_t)u_t, \tag{1.2}$$

where $x_t \in \mathbb{R}^{n_x}$, $y_t \in \mathbb{R}^{n_y}$, $u_t \in \mathbb{R}^{n_u}$, $v_t \in \mathbb{R}^{n_v}$ and $w_t \in \mathbb{R}^{n_w}$. Given s_t, $A_t(s_t)$, $B_t(s_t)$, $C_t(s_t)$, $D(s_t)$, $F_t(s_t)$ and $G_t(s_t)$ are known matrices of appropriate dimension and $D_t(s_t)D_t^T(s_t) > 0$ for any s_t. x_t is an unobserved state, y_t is the observation process and u_t is an exogenous control term. The noise sequences $v_t \overset{i.i.d}{\sim} N(0, I_{n_{v_t}})$, $w_t \overset{i.i.d}{\sim} N(0, I_{n_{w_t}})$ are independent Gaussian sequences, mutually independent and independent of the initial state $x_0 \sim N(m_0, P_0)$.

Conditional upon s_t, (1.1)–(1.2) is thus a standard linear Gaussian state-space model. However, the process s_t is itself an unobserved random process. For the sake of simplicity, it is assumed to be a first-order Markov process of initial distribution $p(s_0)$ and Markov transition kernel $p(s_t \mid s_{t-1})$.

**Now at SaabTech Systems, Nettovägen 6, 175 88 Järfälla, Sweden, e-mail: ncbe@celsiustech.se

***The research of Neil Gordon was sponsored by UK MOD Corporate Research programme TG9.

This class of models has numerous applications as illustrated in the two following examples, see Kitagawa and Gersch (1996), Shephard (1994) and West and Harrison (1997) for many other examples.

Example 1. *Jump Markov linear system.* Assume that the process s_t is a finite state-space Markov chain, then the resulting model is a so-called JMLS; that is a linear Gaussian system whose parameters evolve according to an unobserved finite state-space Markov chain. Such systems are widely used in digital communications, econometrics and target tracking (see Bar-Shalom and Li (1995)).

Example 2. *Time-varying autoregressive (TVAR) process.* The TVAR coefficients a_t are reparametrised into the partial correlation coefficients $s_t \in \mathbb{R}^{n_s}$ and s_t is assumed to follow a simple Gaussian random walk: $s_t = s_{t-1} + \varepsilon_t$ where $s_0 \sim N(0, I_{n_s})$ and $\varepsilon_t \overset{i.i.d}{\sim} N(0, I_{n_s})$. From s_t, one can compute a_t through the standard Levinson recursion. Let $u_t = \sum_{i=1}^{n_s} a_{t,i} u_{t-i} + \sigma_v v_t$, and $y_t = u_t + \sigma_w w_t$ where $v_t \overset{i.i.d}{\sim} N(0,1)$ and $w_t \overset{i.i.d}{\sim} N(0,1)$. Then, by setting $x_t \equiv (u_t, \ldots, u_{t-n_s})$, one can put (s_t, x_t, y_t) in the state-space form (1.1)–(1.2).

In this paper, we propose an efficient sequential Monte Carlo (SMC) method to perform Bayesian inference for partial non-Gaussian state-spaces and we derive simple recursions allowing easy computation of some PCRB.

1.2 *Sequential Bayesian estimation*

We denote for any sequence z_t, $z_{i:j} \equiv (z_i, z_{i+1}, \ldots, z_j)$. We are interested in estimating sequentially in time t the posterior distribution of the state of the system given by $p(x_{0:t}, s_{0:t} \mid y_{1:t})$, or some of its characteristics such as the filtering distribution $p(x_t \mid y_{1:t})$. There is no closed-form expression for this class of models and one needs to use computational methods to perform Bayesian inference. In a batch framework, several authors have exploited the structure of partial non-Gaussian state-space models so as to develop efficient Markov chain Monte Carlo (MCMC) algorithms (see Carter and Kohn (1994, 1996), Früwirth-Schnatter (1994) and Shephard (1994)). However MCMC methods are not suited to sequential estimation. Recently there has been a surge of interest in SMC methods for nonlinear/non-Gaussian time series analysis (Doucet *et al.* (2001)). These methods, initiated in Gordon *et al.* (1993) and Kitagawa (1996), utilise a random sample (or particle) based representation of the posterior probability distributions: the particles are propagated over time using a combination of sequential importance sampling and resampling steps. Related early work by West (1993*a*, 1993*b*) develops weighted mixtures of kernel densities as the proposal distribution for sequential importance sampling. However, in their standard forms, these algorithms do not use all the salient structure of partial non-Gaussian state-space models.

We show here how it is possible to use this structure to develop an efficient SMC algorithm to perform sequential Bayesian estimation. This algorithm combines sequential importance sampling, a selection scheme and MCMC methods. In particular, variance reduction is achieved by Rao-Blackwellisation using the Kalman filter as discussed in Doucet (1997) and Doucet *et al.* (2000). However, we further improve the algorithm by using other variance reduction methods and sampling schemes. A generalization of the backward-forward algorithm of Carter and Kohn (1996) is also given: it allows an exact initialization of the backward recursion, and requires neither the state covariance matrix to be strictly positive nor $A(s_t)$ to be regular.

1.3 *Posterior Cramér-Rao lower bounds*

For many real-world applications it is of great practical importance to be able to compute a bound which sets a lower limit on the average mean-square error (MSE) of the state estimate x_t. For instance, assume one is interested in tracking the position of an aircraft, then, in most civil and military applications, the user of the tracking system requires the average MSE on the position to be below a certain value. If the estimates the tracking system delivers are not precise enough, then it is considered to be of no practical use. Thus, if one obtains for the dynamic model associated to the target/tracking system, a lower bound on the average MSE that is larger than the required precision, it means that it is not even worth trying to develop an estimation scheme since not even the "true" optimal filter would achieve the required precision. Instead what would be required is more information to aid the estimation process (for example an improved sensor to provide more accurate measurements).

This problem has generated a large literature in the control and signal processing community; a review can be found in Kerr (1989). Although different bounds exist, most work has focused on PCRB as they are much easier to compute than the Barankin or Ziv-Zakai bounds. PCRB for discrete time filtering were initially studied by Borobsky and Zakai (1975) for scalar nonlinear models in additive Gaussian noise. Galdos (1980) extended this result to the multidimensional case, and later Doerschuk (1995) considered the class of nonlinear autoregressive processes driven by Gaussian noise with full rank covariance matrix. More recently, Bergman (1999), Bergman *et al.* (1999) and Tichavský *et al.* (1998) have independently derived general expressions for PCRB using a different and more general approach. However, their method cannot be applied to general partial non-Gaussian state-space models. Indeed, all these approaches require $p(s_t \mid s_{t-1})$ to be differentiable in argument s_t. This is obviously not the case when s_t is, for example, a discrete random variable. We derive here some original recursions to compute these bounds for state estimation in partial non-Gaussian state-space models. These bounds are very easy to estimate though numerical integration is generally required.

1.4 *Plan*

We now list the main results and the organization of this paper. Section 2 presents a general sequential Monte Carlo method, details the implementation issues and briefly reviews some sufficient conditions to ensure asymptotic convergence. Section 3 is devoted to the derivation and computation of posterior Cramér-Rao lower bounds. In Section 4, we demonstrate the performance of the proposed algorithms for jump Markov linear systems.

2. Sequential Bayesian estimation

2.1 *Particle filtering*

Given the observations $y_{1:t}$, any inference on $(x_{0:t}, s_{0:t})$ is based on the joint posterior distribution $p(x_{0:t}, s_{0:t} \mid y_{1:t})$ and its characteristics of interest such as the filtering distribution $p(x_t \mid y_{1:t})$ or the minimum mean square error (MMSE) estimate $E(x_t \mid y_{1:t})$. The joint distribution can be factorized as follows

$$p(x_{0:t}, s_{0:t} \mid y_{1:t}) = p(x_{0:t} \mid y_{1:t}, s_{0:t}) p(s_{0:t} \mid y_{1:t}), \tag{2.1}$$

where $p(x_{0:t} \mid y_{1:t}, s_{0:t})$ is a Gaussian distribution whose parameters can be evaluated using Kalman recursions. From (2.1), one can see that, once the marginal distribution

$p(s_{0:t} \mid y_{1:t})$ is estimated, it is possible to obtain easily an estimate of $p(x_{0:t}, s_{0:t} \mid y_{1:t})$ if necessary. Thus we focus in this section on developing a simulation-based method to estimate $p(s_{0:t} \mid y_{1:t})$. This distribution satisfies the following recursion

$$p(s_{0:t} \mid y_{1:t}) = \frac{p(y_t \mid y_{1:t-1}, s_{0:t}) p(s_t \mid s_{t-1})}{p(y_t \mid y_{1:t-1})} p(s_{0:t-1} \mid y_{1:t-1}).$$

We describe a general recursive algorithm which, for all $t > 0$, generates N particles/paths $(s_{0:t}^{(i)}; i = 1, \ldots, N)$ at time t with an empirical measure p_t^N,

$$p_t^N(ds_{0:t}) \equiv \frac{1}{N} \sum_{i=1}^{N} \delta_{s_{0:t}^{(i)}}(ds_{0:t}),$$

that is "close" to $p(s_{0:t} \mid y_{1:t})$, using the observation obtained at time t and the set of particles $(s_{0:t-1}^{(i)}; i = 1, \ldots, N)$ produced at time $t-1$ (whose empirical measure p_{t-1}^N was "close" to $p(s_{0:t-1} \mid y_{1:t-1})$); $\delta_{s_{0:t}^{(i)}}(ds_{0:t})$ denotes the delta-Dirac mass at $s_{0:t}^{(i)}$. This algorithm requires the introduction of an importance function $\pi(s_t \mid y_{1:t}, s_{0:t-1})$. It proceeds as follows at time t.

Particle filter for partial non-Gaussian state space models

Sequential importance sampling step

- For $i = 1, \ldots, N$, sample $\tilde{s}_t^{(i)} \sim \pi(s_t \mid y_{1:t}, s_{0:t-1}^{(i)})$ and $\tilde{s}_{0:t}^{(i)} \equiv (s_{0:t-1}^{(i)}, \tilde{s}_t^{(i)})$.
- For $i = 1, \ldots, N$, evaluate the importance weights $\tilde{w}_t^{(i)}$

$$\tilde{w}_t^{(i)} \propto \frac{p(y_t \mid y_{1:t-1}, \tilde{s}_{0:t}^{(i)}) p(\tilde{s}_t^{(i)} \mid \tilde{s}_{t-1}^{(i)})}{\pi(\tilde{s}_t^{(i)} \mid y_{1:t}, \tilde{s}_{0:t-1}^{(i)})}. \tag{2.2}$$

Selection step

- Multiply/Discard particles $(\tilde{s}_{0:t}^{(i)}; i = 1, \ldots, N)$ with respect to high/low normalised importance weights $\tilde{w}_t^{(i)}$ to obtain N particles $(s_{0:t}^{\prime(i)}; i = 1, \ldots, N)$.

MCMC step

- For $i = 1, \ldots, N$, apply to $s_{0:t}^{\prime(i)}$ a Markov transition kernel $K(ds_{0:t}^{(i)} \mid s_{0:t}^{\prime(i)})$ of invariant distribution $p(ds_{0:t} \mid y_{1:t})$ to obtain N particles $(s_{0:t}^{(i)}; i = 1, \ldots, N)$.

2.2 *Implementation issues*

2.2.1 *Sampling step*

There are infinitely many possible choices for $\pi(s_t \mid y_{1:t}, s_{0:t-1})$, the only condition being that its support includes that of $p(y_t \mid y_{1:t-1}, s_{0:t}) p(s_t \mid s_{t-1})$. A sensible selection criterion is to choose a proposal that minimises the variance of the importance weights at time t, given $s_{0:t-1}$ and $y_{1:t}$. According to this strategy, Doucet *et al.* (2000) show that the optimal distribution is $p(s_t \mid y_{1:t}, s_{0:t-1})$.

- *Optimal sampling distribution.* The optimal distribution satisfies

$$\pi(s_t \mid y_{1:t}, s_{0:t-1}) = p(s_t \mid y_{1:t}, s_{0:t-1}) \propto p(y_t \mid y_{1:t-1}, s_{0:t}) p(s_t \mid s_{t-1}),$$

and it is usually easy to sample from. The importance weight $p(y_t \mid y_{1:t-1}, s_{0:t-1})$ might not admit however an analytical expression if s_t does not lie in a finite-state space. In such cases, one has to use an alternative method.

• *Prior distribution.* If we use the prior distribution $p(s_t \mid s_{t-1})$ as importance distribution, the importance weight is proportional to

$$p(y_t \mid y_{1:t-1}, s_{0:t}) = N(\tilde{y}_{t|t-1}(s_{0:t}), S_t(s_{0:t}))$$

where $\tilde{y}_{t|t-1}(s_{0:t})$, $y_{t|t-1}(s_{0:t})$ and $S_t(s_{0:t})$ are evaluated using one step of a Kalman filter detailed in Appendix A.

• *Alternative sampling distribution.* It is possible to design a variety of alternative sampling distributions. For example, one can use the results of a suboptimal deterministic algorithm to construct an importance sampling distribution, Doucet *et al.* (2000) and Pitt and Shephard (1999). This is useful in applications where $p(s_t \mid y_{1:t}, s_{0:t-1})$ is too expensive to compute and $p(s_t \mid s_{t-1})$ is inefficient.

2.2.2 *Selection step*

The aim of the selection step is to obtain an "unweighted" approximation

$$\hat{p}_N(s_{0:t} \mid y_{1:t}) = N^{-1} \sum_{i=1}^{N} N_i \delta_{\tilde{s}_{0:t}^{(i)}}(ds_{0:t}) = N^{-1} \sum_{i=1}^{N} \delta_{s_{0:t}'^{(i)}}(ds_{0:t})$$

of the weighted distribution

$$\tilde{p}_N(s_{0:t} \mid y_{1:t}) = \sum_{i=1}^{N} \tilde{w}_t^{(i)} \delta_{\tilde{s}_{0:t}^{(i)}}(ds_{0:t}),$$

that is a selection procedure associates with each particle $(\tilde{s}_{0:t}^{(i)} : i = 1, \ldots, N)$ a number of offspring $N_i \in N$, such that $\sum_{i=1}^{N} N_i = N$, to obtain N new particles $s_{0:t}'^{(i)}$. If $N_i = 0$, then $\tilde{s}_{0:t}^{(i)}$ is discarded, otherwise it has N_i offspring at time t.

We briefly describe here some selection schemes of complexity $O(N)$.

• *Sampling Importance Resampling (SIR)/multinomial sampling procedure.* This procedure, introduced originally by Gordon *et al.* (1993), is the most popular one. One samples N times from $\tilde{p}_N(s_{0:t} \mid y_{1:t})$ to obtain $(s_{0:t}'^{(i)}; i = 1, \ldots, N)$. This is equivalent to drawing jointly $(N_i; i = 1, \ldots, N)$ according to a multinomial distribution of parameters N and $\tilde{w}_t^{(i)}$. In this case, one has $E[N_i] = N\tilde{w}_t^{(i)}$ and $\text{var}[N_i] = N\tilde{w}_t^{(i)}(1 - \tilde{w}_t^{(i)})$. It is better to use selection schemes with a reduced variance.

• *Residual resampling.* This was first presented by Higuchi (1997) and Liu and Chen (1998). Set $\tilde{N}_i = \lfloor N\tilde{w}_t^{(i)} \rfloor$ then perform a SIR procedure to select the remaining $\bar{N}_t = N - \sum_{i=1}^{N} \tilde{N}_i$ samples with the new weights $w_t'^{(i)} = (\tilde{w}_t^{(i)} N - \tilde{N}_i)/\bar{N}_t$; add the results to the current $\tilde{N}_i$. In this case, we obtain $E[N_i] = N\tilde{w}_t^{(i)}$ but $\text{var}[N_i] = \bar{N}_t w_t'^{(i)}(1 - w_t'^{(i)})$. As outlined by Higuchi (1997), this procedure has been introduced earlier in the genetic algorithm literature.

• *Minimum variance sampling.* This class includes the stratified sampling procedure introduced in Kitagawa (1996) where a set U of N points is generated in the interval $[0, 1]$, each of the points a distance N^{-1} apart. The number N_i is taken to be the number of points in U that lie between $\sum_{j=1}^{i-1} \tilde{w}_t^{(j)}$ and $\sum_{j=1}^{i} \tilde{w}_t^{(j)}$. If we denote $\{N\tilde{w}_t^{(i)}\} \equiv N\tilde{w}_t^{(i)} - \lfloor N\tilde{w}_t^{(i)} \rfloor$, then the variance of all the algorithms in this class is $\text{var}[N_i] = \{N\tilde{w}_t^{(i)}\}(1 - \{N\tilde{w}_t^{(i)}\})$. This strategy also includes the Tree Based Branching Algorithm presented in Crisan (2001).

Recent theoretical results obtained in Crisan and Doucet (2000) show that it is not necessary to design unbiased selection schemes, *i.e.* we can have $E[N_i] \neq N\tilde{w}_t^{(i)}$. Using randomised selection schemes is also part of the "folklore" of particle filters: the deterministic selection scheme proposed by Kitagawa (1996) is efficient and theoretically valid.

2.2.3 *MCMC step*

If the distribution of the importance weights is skewed then the particles $(\tilde{s}_{0:t}^{(i)}, i = 1, \ldots, N)$ which have high importance weights $\tilde{w}_t^{(i)}$ are selected many times and thus numerous particles $s_{0:t}^{\prime(i)}$ and $s_{0:t}^{\prime(j)}$ are in fact equal for $i \neq j$; there is a depletion of samples. One way to perform sample regeneration consists of considering a mixture approximation, Gordon *et al.* (1993). An alternative method based on MCMC methods has been recently proposed by Gilks and Berzuini (1998) in the context of fixed-parameter estimation. The rationale behind the use of MCMC moves is based on the following remark. Assume that the particles $s_{0:t}^{\prime(i)}$ are distributed marginally according to $p(s_{0:t} \mid y_{1:t})$. Then if we apply to each particle a Markov transition kernel $K(ds_{0:t} \mid s'_{0:t})$ of invariant distribution $p(s_{0:t} \mid y_{1:t})$, *i.e.* such that $\int K(ds_{0:t} \mid s'_{0:t}) p(ds'_{0:t} \mid y_{1:t}) = p(ds_{0:t} \mid y_{1:t})$, then the new particles $s_{0:t}^{(i)}$ are still distributed according to the posterior distribution of interest. So if $K(ds_{0:t} \mid s'_{0:t})$ is a kernel that stochastically updates $s_{0:t}^{\prime(i)}$ to obtain $s_{0:t}^{(i)}$ then we have a theoretically valid way of introducing diversity amongst the samples. It is possible to use all the standard MCMC methods such as the Metropolis-Hastings (M-H) or the Gibbs samplers, Robert and Casella (1999), the main difference being that we do not require the kernel to be ergodic.

There is an infinite number of possible choices for the MCMC transition kernel. Carter and Kohn (1996) proposed an efficient MCMC algorithm based on a backward-forward recursion where the state x_t is integrated out and s_t is sampled one-at-a-time. They empirically demonstrated that their sampler was more efficient than the one where x_t is sampled (Carter and Kohn (1994), Früwirth-Schnatter (1994) and Shephard (1994)). We propose here a similar strategy but relax the restrictions of Carter and Kohn (1996) (approximate initialization of the backward recursion, state covariance matrix strictly positive and $A(s_t)$ regular). Our algorithm, based on the backward information filter (Anderson and Moore (1979)), can be applied to any state-space model (1.1)–(1.2), see Doucet (1997) and Doucet and Andrieu (1999) for details. The derivation is not presented here, due to its length.

We set $s_{0:t-L}^{(i)} = s_{0:t-L}^{\prime(i)}$. Then, to sample $s_k^{(i)}$ for $k = t - L + 1, \ldots, t$ according to $p(s_k \mid y_{1:t}, s_{-k}^{(i)})$ where $s_{-k}^{(i)} \equiv (s_{0:t-L}^{(i)}, s_{t-L+1}^{(i)}, \ldots, s_{k-1}^{(i)}, s_{k+1}^{\prime(i)}, \ldots, s_t^{\prime(i)})$, the algorithm proceeds as follows at time t for the particle i.

Backward-forward procedure

Backward step

For $k = t-L+1, \ldots, t$, compute and store $P_{k|k+1}^{\prime -1}(s_{k+1:t}^{\prime(i)})$ and $P_{k|k+1}^{\prime -1}(s_{k+1:t}^{\prime(i)}) m'_{k|k+1}(s_{k+1:t}^{\prime(i)})$ using (A.1)–(A.3) given in Appendix A.

Forward step

For $k = t - L + 1, \ldots, t$, sample $s_k^{(i)}$ using a Gibbs or a M-H step of invariant distribution $p(s_k \mid y_{1:t}, s_{-k}^{(i)})$ (eq. (2.3)) and store $m_{k|k}(s_{0:k-1}^{(i)}, s_k^{(i)})$ and $P_{k|k}(s_{0:k-1}^{(i)}, s_k^{(i)})$.

In this algorithm, $P_{k|k+1}^{\prime -1}(s_{k+1:t}^{\prime(i)})$ and $P_{k|k+1}^{\prime -1}(s_{k+1:t}^{\prime(i)}) m'_{k|k+1}(s_{k+1:t}^{\prime(i)})$ are given by the

backward information filter recursion given in Appendix A and, for any $k = t - L + 1, \ldots, t$, we have

$$
\begin{aligned}
(2.3)\ p(s_k \mid y_{1:t}, s_{-k}) &\propto p(s_k \mid s_{k-1}) p(s_{k+1} \mid s_k) N(\tilde{y}_{k|k-1}(s_{0:k}), S_k(s_{0:k})) \\
&\times |\tilde{\Pi}_{k|k}(s_{0:k}) \tilde{Q}^T_{k|k}(s_{0:k}) P'^{-1}_{k|k+1}(s_{k+1:t}) \tilde{Q}_{k|k}(s_{0:k}) + I_{n_k}|^{-1/2} \\
&\times \exp\Bigg(-\frac{1}{2} [m^T_{k|k}(s_{0:k}) P'^{-1}_{k|k+1}(s_{k+1:t}) m_{k|k}(s_{0:k}) \\
&\quad - 2 m^T_{k|k}(s_{0:k}) P'^{-1}_{k|k+1}(s_{k+1:t}) m'_{k|k+1}(s_{k+1:t}) \\
&\quad - (m'_{k|k+1}(s_{k+1:t}) - m_{k|k}(s_{0:k}))^T P'^{-1}_{k|k+1}(s_{k+1:t}) \\
&\quad \times R_{k|k}(s_{0:t}) P'^{-1}_{k|k+1}(s_{k+1:t}) (m'_{k|k+1}(s_{k+1:t}) - m_{k|k}(s_{0:k}))] \Bigg).
\end{aligned}
$$

Here one has $P_{k|k}(s_{0:k}) = \tilde{Q}_{k|k}(s_{0:k}) \tilde{\Pi}_{k|k}(s_{0:k}) \tilde{Q}^T_{k|k}(s_{0:k})$ where $\tilde{\Pi}_{k|k}(s_{0:k})$ is a $n_k \times n_k$ $(1 \leq n_k \leq n_x)$ diagonal matrix with the non-zero eigenvalues of $P_{k|k}(s_{0:k})$ as elements, and

$$
R_{k|k}(s_{0:t}) = \tilde{Q}_{k|k}(s_{0:k}) [\tilde{\Pi}^{-1}_{k|k}(s_{0:k}) + \tilde{Q}^T_{k|k}(s_{0:k}) P'^{-1}_{k|k+1}(s_{k+1:t}) \tilde{Q}_{k|k}(s_{0:k})]^{-1} \tilde{Q}^T_{k|k}(s_{0:k}).
$$

The matrices $\tilde{Q}_{k|k}(s_{0:k})$ and $\tilde{\Pi}_{k|k}(s_{0:k})$ are straightforwardly obtained using the singular value decomposition of $P_{k|k}(s_{0:k})$. The computational complexity of the resulting Gibbs sampling algorithm at each iteration is $O(LN)$ and one needs to keep in memory the paths of all trajectories $(s'^{(i)}_{t-L+1:t}; i = 1, \ldots, N)$ as well as $(m_{k|k}(s^{(i)}_{0:k}), P_{k|k}(s^{(i)}_{0:k}); i = 1, \ldots, N)$ over the time interval $k = t - L + 1, \ldots, t$.

2.3 *Convergence issues*

Let $B(\mathbb{R}^n)$ be the space of bounded, Borel measurable functions on $\mathbb{R}^n$. We denote $\|f\| \equiv \sup_{\in \mathbb{R}^n} |f(x)|$. The following theorem is a straightforward consequence of Theorem 1 in Crisan and Doucet (2000), which is an extension of previous results in Crisan *et al.* (1999).

THEOREM 2.3.1. *If the importance weights*, (2.2), *are upper bounded and if one uses one of the selection schemes described previously, then, for all* $t \geq 0$, *there exists* c_t *independent of* N *such that for any* $f_t \in B(\mathbb{R}^{n_s \times (t+1)})$

$$
E\left[\left(\frac{1}{N} \sum_{i=1}^{N} f_t(s^{(i)}_{0:t}) - \int f_t(s_{0:t}) p(ds_{0:t} \mid y_{1:t}) \right)^2 \right] \leq c_t \frac{\|f_t\|^2}{N}.
$$

This result shows that, under very mild assumptions, convergence of this general particle filtering method is ensured and that the convergence rate of the method is independent of the dimension of the state-space; c_t usually increases exponentially with time however. If additional assumptions on the dynamic system under study are made, it is possible to get uniform convergence results for the filtering distribution $p(x_t \mid y_{1:t})$, as demonstrated by Del Moral and Guionnet (1998).

Another severe restriction of this convergence result is that it is limited to bounded functions: this excludes the MMSE estimate $E(x_t \mid y_{1:t})$. In conclusion, although the

particle filter is guaranteed to converge asymptotically (as $N \to +\infty$) and seems to stabilize in practice, it is difficult to assess rigorous convergence towards the true filter when N is finite.

3. Posterior Cramér-Rao lower bounds

In this section, we derive lower bounds on the average mean-square error of the state estimate of x_t given by

$$E((E(x_t \mid y_{1:n}) - x_t)(E(x_t \mid y_{1:n}) - x_t)^T), \tag{3.1}$$

where $E(x_t \mid y_{1:n})$ is the MMSE estimate of x_t given the observations $y_{1:n}$ and the expectation is with respect to $p(x_t, y_{1:n})$. A Monte Carlo method to estimate (3.1) would consist of sampling a large number P of realizations $\{(x_{1:t}(i), y_{1:n}(i)); i = 1, \ldots, P\}$, computing an estimate $\hat{x}_{t|n}(i)$ of $E(x_t \mid y_{1:n}^{(i)})$ for each realization using a particle filtering method and then finally approximating (3.1) by

$$\frac{1}{P}\sum_{i=1}^{P}(\hat{x}_{t|n}(i) - x_t(i))(\hat{x}_{t|n}(i) - x_t(i))^T. \tag{3.2}$$

This method, however, is not only very computationally intensive, but it is also difficult to quantify the error one commits by approximating $E(x_t \mid y_{1:n}^{(i)})$ by $\hat{x}_{t|n}(i)$. Instead, we derive below a posterior Cramér-Rao lower bound on (3.1), which is much easier to compute.

3.1 *Review of the posterior Cramér-Rao bound*

Let $\theta \in \mathbb{R}^n$ be a random parameter with prior density $p(\theta)$, and let $\hat{\psi}(z) : \mathcal{Z} \to \mathbb{R}^p$ be an estimate of an absolutely continuous function of this parameter, $\psi(\theta) : \mathbb{R}^n \to \mathbb{R}^p$, based on the observation $z \in \mathcal{Z}$. Then we have the following Bayesian version of the Cramér-Rao bound. It is originally due to van Trees (1968); see also Gill and Levit (1995) for different extensions and detailed regularity conditions.

THEOREM 3.1.1. *The PCRB for estimating θ using z is given by*

$$E_{p(z,\theta)}((\hat{\psi}(z) - \psi(\theta))(\hat{\psi}(z) - \psi(\theta))^T) \geq MPM^T, \tag{3.3}$$

where

$$M^T = E_{p(z,\theta)}(\nabla_\theta \log p(z,\theta)(\hat{\psi}(z) - \psi(\theta))^T) = \int \nabla_\theta p(z,\theta)(\hat{\psi}(z) - \psi(\theta))^T dz d\theta$$

$$P^{-1} = E_{p(z,\theta)}(\nabla_\theta \log p(z,\theta) \nabla_\theta^T \log p(z,\theta)) = \int \nabla_\theta p(z,\theta) \nabla_\theta^T \log p(z,\theta) dz d\theta.$$

The PCRB differs from the traditional, likelihood based, Cramér-Rao bound in several ways. The bound (3.3) is a matrix inequality on the estimator mean square error correlation matrix and it holds even if the estimator $\hat{\psi}(z)$ is biased. Several interesting scalar bounds can be derived from (3.3), and a number of different weighted matrix bounds can also be formed: see Gill and Levit (1995). The bound (3.3) does not depend on any true unknown parameter θ_0, instead it is computed by expectation with respect to the joint density $p(z, \theta)$.

3.2 *PCRB for partial non-Gaussian state space models*

We now derive PCRB for estimation of the state x_t. Conditional upon $s_{0:t}$ the model (1.1)–(1.2) is a linear Gaussian state-space model and the optimal MMSE estimate of x_t is given by the Kalman filter. The error covariance from the Kalman filter equations would therefore give the fundamental lower bound on the estimation performance if the sequence $s_{0:t}$ were known. Given $y_{1:n}$, the PCRB for the state x_t is given by

$$P_{t|n}^{-1} = E_{p(x_t,y_{1:n})}(-\Delta_{x_t}^{x_t} \log p(x_t, y_{1:n})), \tag{3.4}$$

where $\Delta_{x_t}^{x_t}$ is the second order (Laplacian) partial differential operator with respect to the vector x_t. This fundamental bound (3.4) is valid under very mild conditions on the prior $p(x_t)$.

The PCRB can be used to determine a lower bound on the estimation error of x_t based on the prior distribution of the sequence $s_{0:t}$. We get the following results.

THEOREM 3.2.1. *Filtering. The PCRB for estimating x_t in the model* (1.1)–(1.2) *using $y_{1:t}$ is given by*

$$E((\hat{x}_t(y_{1:t}) - x_t)(\hat{x}_t(y_{1:t}) - x_t)^T) \geq P_{t|t},$$

where the matrix $P_{t|t}$ satisfies

$$P_{t|t} = A_t \Upsilon_t A_t^T + (A_t \Upsilon_t S_t - B_t)\Delta_t^{-1}(S_t^T \Upsilon_t A_t^T - B_t^T), \tag{3.5}$$

with $\Upsilon_t = (P_{t-1|t-1}^{-1} + R_t^{-1})^{-1}$ and $\Delta_t = Q_t^{-1} - S_t^T \Upsilon_t^{-1} S_t$. The matrices A_t, B_t, Q_t and R_t are formed by averaging the matrices from the model (1.1)–(1.2) *over the prior distribution of s_t, i.e.*

$$\begin{aligned}
A_t &= E(A_t(s_t)) \\
B_t &= E(B_t(s_t)) \\
Q_t^{-1} &= E(-\Delta_{v_t}^{v_t} \log p(y_t \,|x_{t-1}, s_t, v_t) p(v_t)) \\
&= E(B_t^T(s_t) C_t^T(s_t)(D_t(s_t) D_t^T(s_t))^{-1} C_t(s_t) B_t(s_t)) + I_{n_{v_t}} \\
R_t^{-1} &= E(-\Delta_{x_{t-1}}^{x_{t-1}} \log p(y_t \,|x_{t-1}, s_t, v_t)) \\
&= E(A_t^T(s_t) C_t^T(s_t)(D_t(s_t) D_t^T(s_t))^{-1} C_t(s_t) A_t(s_t)) \\
S_t^T &= E(-\Delta_{v_t}^{x_{t-1}} \log p(y_t \,|x_{t-1}, s_t, v_t)) \\
&= E(B_t^T(s_t) C_t^T(s_t)(D_t(s_t) D_t^T(s_t))^{-1} C_t(s_t) A_t(s_t)).
\end{aligned} \tag{3.6}$$

PROOF. The proof is given in Appendix B.

A bound in the prediction case can be obtained in a similar way.

THEOREM 3.2.2. *Prediction. The PCRB for estimating x_t in the model* (1.1)–(1.2) *using $y_{1:t-1}$ is given by*

$$E((\hat{x}_t(y_{1:t-1}) - x_t)(\hat{x}_t(y_{1:t-1}) - x_t)^T) \geq P_{t|t-1},$$

where the matrix $P_{t|t-1}$ satisfies

$$P_{t|t-1} = A_t(P_{t-1|t-2} - P_{t-1|t-2}(P_{t-1|t-2} + R_t)^{-1} P_{t-1|t-2}) A_t^T + B_t B_t^T. \tag{3.7}$$

PROOF. The proof is given in Appendix B.

These bounds require numerical integration methods as A_t, B_t, Q_t, R_t and S_t in (3.6), do not usually admit any closed-form expression. These quantities can be computed using Monte Carlo methods by simulating realizations of s_t according to its *prior* distribution. Quantifying the error of these Monte Carlo estimates is straightforward.

4. Application

We address the problem of tracking a maneuvering target in noise. The difficulty in this problem arises from the uncertainty in the maneuvering command driving the target. The state of the target at time t is denoted as $x_t \equiv (l_{x,t}, r_{x,t}, l_{y,t}, r_{y,t})^T$ where $l_{x,t}$ ($l_{y,t}$) and $r_{x,t}$ ($r_{y,t}$) represent the position and velocity of the target in the x (resp. in the y) direction. The state evolves according to a JMLS model of parameters (see Bar-Shalom and Li (1995)),

$$A = \begin{pmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad B = 0.1 I_4, \quad C = I_4, \quad G = 0_{4 \times n_u},$$

and $D = \sqrt{3}\,\mathrm{diag}(20, 1, 20, 1)$. The switching term is $F(s_t)u_t$, where s_t is 3-state Markov chain corresponding to the three possible maneuver commands: straight, left turn, right turn. The chain has the following transition probabilities: $p_{m,m} = 0.9$ and $p_{m,n} = 0.05$ for $m \neq n$; its initial distribution is the invariant distribution of the transition matrix. For any t, we have

$$\begin{aligned} F(1)u_t &= (0, 0, 0, 0)^T, \\ F(2)u_t &= (-1.225, -0.35, 1.225, 0.35)^T, \\ F(3)u_t &= (1.225, 0.35, -1.225, -0.35)^T. \end{aligned}$$

We implement the particle filtering method described in Section 2 to estimate $E(x_t \mid y_{1:t})$. We use as importance distribution $p(s_t \mid y_{1:t}, s_{1:t-1})$ and perform selection according to the stratified sampling scheme. We perform $M = 100$ different measurement realizations and compare our results with the Interacting Multiple Model (IMM) algorithm (Bar-Shalom and Li (1995)) and the standard Sampling Importance Resampling (SIR) filter (Gordon *et al.* (1993) and Kitagawa (1996)). The performance measure is the root mean square (RMS) position error, computed as follows from the MMSE estimates with respect to the true simulated trajectories:

$$RMS = \sqrt{\frac{1}{MT} \sum_{m=1}^{M} \sum_{t=1}^{T} [(l_{x,t} - l_{x,t}^{MMSE}(m))^2 + (l_{y,t} - l_{y,t}^{MMSE}(m))^2]},$$

where $l_{x,t}^{MMSE}(m)$ is the MMSE target position estimate in the x direction at time t of the m-th Monte Carlo simulation. We present in Table 1 the performance of our Monte Carlo (MC) filter, the IMM and the SIR filter. The MC filter is more precise than the other methods.

Table 1. RMS for IMM, SIR filter and MC filter.

Algorithm/N	50	100	250	500	1000
IMM Filter	24.69	-	-	-	-
SIR Filter	26.22	24.76	24.02	23.88	23.45
MC Filter	22.95	22.74	22.69	22.64	22.62

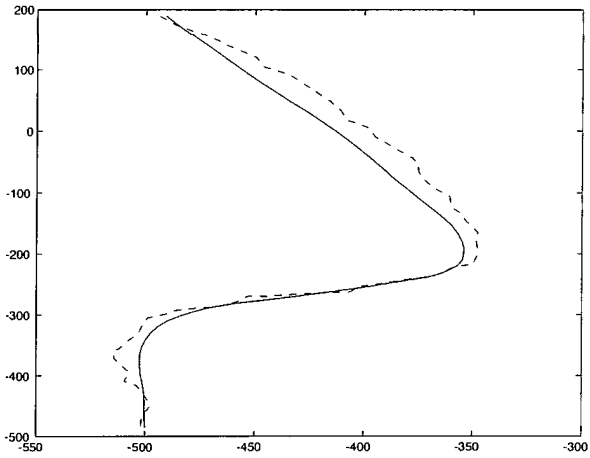

Fig. 1. Simulated path of the target $(l_{x,t}, l_{y,t})$ (solid line) and MMSE estimate (dashed line).

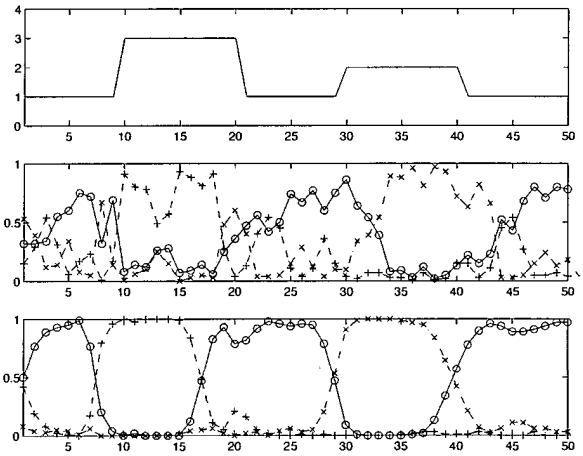

Fig. 2. From top to bottom: simulated sequence r_t, estimation of $\hat{p}(s_t \mid y_{1:t})$ ($s_t = 1$ circle solid, $s_t = 2$ cross dashdot, $s_t = 3$ plus dashed), $\hat{p}(s_t \mid y_{1:t+10})$ (same convention for $s_t = 1, 2, 3$).

In Fig. 1, we display a realization of $(l_{x,t}, l_{y,t})$ and its MMSE estimate computed using $N = 500$ particles. In Fig. 2, we present for the same realization the simulated sequence s_t, $\hat{p}(s_t \mid y_{1:t})$ and $\hat{p}(s_t \mid y_{1:t+L})$ $(L = 10)$. Fixed-lag smoothing significantly improves the detection of occurrences with respect to filtering. The performance of the

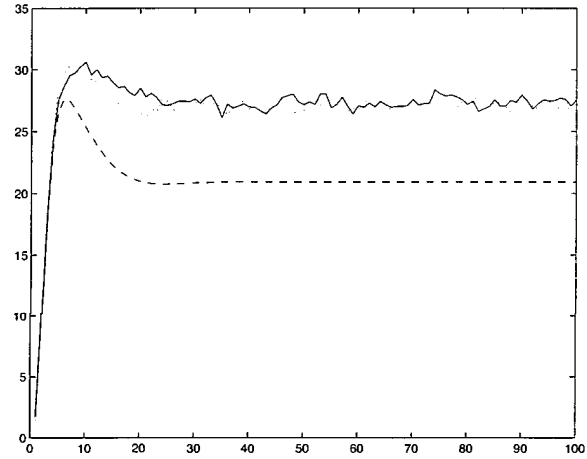

Fig. 3. Monte Carlo approximation of the average mean-square error for the Monte Carlo filter (solid line) and SIR filter (dotted line) with the posterior Cramér-Rao bound (dashed line) for $l_{x,t}$.

fixed-lag smoothing approximation and of a Gibbs sampler to estimate $p(s_t \mid y_{1:T})$ (not displayed here) appears very similar. Finally in Fig. 3, we compute the PCRB given by (3.5) and compare it to the empirical estimate (3.2) of the RMS (3.1) obtained from both our algorithm ($N = 500$) and the SIR filter ($N = 10000$). The PCRB is of the same order of magnitude as the RMS achieved by the particle filter algorithms.

5. Discussion

In this paper, we have addressed the problem of optimal estimation for partial non-Gaussian state space models. A particle filtering method based on several variance reduction methods has been proposed and general posterior Cramér-Rao bounds have been established. The performance of this algorithm and bounds have been demonstrated on a target tracking problem.

Appendix A: Kalman filter and backward information filter

We consider the system (1.1)–(1.2). The sequence $s_{0:t}$ being here assumed known, the Kalman filter equations are as follows. Set $m_{0|0} = \hat{x}_0$ and $P_{0|0} = P_0$, then, for any t, compute

$$
\begin{aligned}
m_{t|t-1}(s_{0:t}) &= A(s_t)m_{t-1|t-1}(s_{1:t-1}) + F(s_t)u_t \\
P_{t|t-1}(s_{0:t}) &= A(s_t)P_{t-1|t-1}(s_{1:t-1})A^T(s_t) + B(s_t)B^T(s_t) \\
y_{t|t-1}(s_{0:t}) &= C(s_t)m_{t|t-1}(s_{0:t}) + G(s_t)u_t \\
S_t(s_{0:t}) &= C(s_t)P_{t|t-1}(s_{0:t})C^T(s_t) + D(s_t)D^T(s_t) \\
m_{t|t}(s_{0:t}) &= m_{t|t-1}(s_{0:t}) + P_{t|-1}(s_{0:t})C^T(s_t)S_t^{-1}(s_{0:t})\tilde{y}_{t|t-1}(s_{0:t}) \\
P_{t|t}(s_{0:t}) &= P_{t|t-1}(s_{0:t}) - P_{t|-1}(s_{0:t})C^T(s_t)S_t^{-1}(s_{0:t})C(s_t)P_{t|t-1}(s_{0:t}),
\end{aligned}
$$

where $m_{t|t-1}(s_{0:t}) = E\{x_t \mid y_{1:t-1}, s_{0:t}\}$, $m_{t|t}(s_{0:t}) = E\{x_t \mid y_{1:t}, s_{0:t}\}$, $\tilde{y}_{t|t-1}(s_{0:t}) = y_t - y_{t|t-1}(s_{0:t})$, $P_{t|t-1}(s_{0:t}) = \mathrm{cov}\{x_t \mid y_{1:t-1}, s_{0:t}\}$, $P_{t|t}(s_{0:t}) = \mathrm{cov}\{x_t \mid y_{1:t}, s_{0:t}\}$, $y_{t|t-1}$

$(s_{0:t}) = E\{y_t \mid y_{1:t-1}, s_{0:t}\}$ and $S_t(s_{0:t}) = \operatorname{cov}\{y_t \mid y_{1:t-1}, s_{0:t}\}$.

The backward information filter proceeds as follows from time t to $t-L+1$,

$$P'^{-1}_{t|t}(s_t) = C^T(s_t)(D(s_t)D^T(s_t))^{-1}C(s_t) \tag{A.1}$$

$$P'^{-1}_{t|t}(s_t)m'_{t|t}(s_t) = C^T(s_t)(D(s_t)D^T(s_t))^{-1}(y_t - G(s_t)u_t), \tag{A.2}$$

and for $k = t, \ldots, t-L+1$,

$$\begin{aligned}
&\Delta_{k+1}(s_{k+1:t}) = [I_{n_v} + B^T(s_{k+1})P'^{-1}_{k+1|k+1}(s_{k+1:t})B(s_{k+1})]^{-1}\\
&P'^{-1}_{k|k+1}(s_{k+1:t})\\
&\quad = A^T(s_{k+1})P'^{-1}_{k+1|k+1}(s_{k+1:t})\\
&\qquad \times (I_{n_x} - B(s_{k+1})\Delta_{k+1}(s_{k+1:t})B^T(s_{k+1})P'^{-1}_{k+1|k+1}(s_{k+1:t}))A(s_{k+1})\\
&P'^{-1}_{k|k+1}(s_{k+1:t})m'_{k|k+1}(s_{k+1:t})\\
&\quad = A^T(s_{k+1})(I_{n_x} - P'^{-1}_{k+1|k+1}(s_{k+1:t})B(s_{k+1})\Delta_{k+1}(s_{k+1:t})B^T(s_{k+1}))\\
&\qquad \times P'^{-1}_{k+1|k+1}(s_{k+1:t})(m'_{k+1|k+1}(s_{k+1:t}) - F(s_{k+1})u_{k+1})\\
&P'^{-1}_{k|k}(s_{k:t}) = P'^{-1}_{k|k+1}(s_{k+1:t}) + C^T(s_k)(D(s_k)D^T(s_k))^{-1}C(s_k)\\
&P'^{-1}_{k|k}(s_{k:t})m'_{k|k}(s_{k:t}) = P'^{-1}_{k|k+1}(s_{k+1:t})m'_{k|k+1}(s_{k+1:t})\\
&\qquad\qquad + C^T(s_k)(D(s_k)D^T(s_k))^{-1}(y_k - G(s_k)u_k).
\end{aligned} \tag{A.3}$$

Appendix B: Posterior Cramér-Rao bounds derivation

PROOF OF THEOREM 3.2.1. Consider the stochastic vector

$$\xi_t = \begin{pmatrix} \hat{x}_{t-1}(y_{1:t}) - x_{t-1} \\ \hat{x}_t(y_{1:t}) - A_t(s_t)x_{t-1} - B_t(s_t)v_t - F_t(s_t)u_t \\ \nabla_{x_{t-1}} \log p(v_t, x_{t-1}, s_{t-1:t}, y_{1:t}) \\ \nabla_{v_t} \log p(v_t, x_{t-1}, s_{t-1:t}, y_{1:t}) \end{pmatrix},$$

where

$$p(v_t, x_{t-1}, s_{t-1:t}, y_{1:t}) = p(v_t)p(s_t \,|s_{t-1})p(y_t \,|x_{t-1}, s_t, v_t)p(x_{t-1}, s_{t-1}, y_{1:t-1}).$$

The correlation matrix of ξ_t is

$$E(\xi_t\xi_t^T) = \begin{pmatrix} H_t & M_t \\ M_t^T & K_t \end{pmatrix} \geq 0, \tag{B.1}$$

where

$$H_t = \begin{pmatrix} H_{11,t} & H_{12,t} \\ H_{21,t} & H_{22,t} \end{pmatrix} \quad M_t = \begin{pmatrix} I & 0 \\ A_t & B_t \end{pmatrix} \quad K_t = \begin{pmatrix} J_{t-1} + R_t^{-1} & S_t \\ S_t^T & Q_t^{-1} \end{pmatrix},$$

$$J_{t-1} = E(-\Delta^{x_{t-1}}_{x_{t-1}} \log p(x_{t-1}, s_{t-1}, y_{1:t-1})).$$

Assume that a bound on the filter error covariance for the previous time instant is known; $E((\hat{x}_{t-1}(y_{1:t-1}) - x_{t-1})(\hat{x}_{t-1}(y_{1:t-1}) - x_{t-1})^T) \geq P_{t-1|t-1}$. It follows from the PCRB that $P_{t-1|t-1}^{-1} \geq J_{t-1}$, and hence from (B.1) that

$$\begin{pmatrix} H_{11,t} & H_{12,t} & I & 0 \\ H_{21,t} & H_{22,t} & A_t & B_t \\ I & A_t^T & P_{t-1|t-1}^{-1} + R_t^{-1} & S_t \\ 0 & B_t^T & S_t^T & Q_t^{-1} \end{pmatrix} \geq 0.$$

Thus, the following matrix inequality holds

$$\text{(B.2)} \qquad \begin{pmatrix} H_{11,t} & H_{12,t} \\ H_{21,t} & H_{22,t} \end{pmatrix} \geq \begin{pmatrix} I & 0 \\ A_t & B_t \end{pmatrix} \begin{pmatrix} P_{t-1|t-1}^{-1} + R_t^{-1} & S_t \\ S_t^T & Q_t^{-1} \end{pmatrix}^{-1} \begin{pmatrix} I & A_t^T \\ 0 & B_t^T \end{pmatrix},$$

and the lower right block induces a lower bound on the filtering error covariance at the current time instant: $H_{22,t} = E((\hat{x}_t(y_{1:t}) - x_t)(\hat{x}_t(y_{1:t}) - x_t)^T) \geq P_{t|t}$. This matrix bound follows from (B.2) since

$$\begin{pmatrix} P_{t-1|t-1}^{-1} + R_t^{-1} & S_t \\ S_t^T & Q_t^{-1} \end{pmatrix}^{-1} = \begin{pmatrix} (\Upsilon_t^{-1} - S_t Q_t^{-1} S_t^T)^{-1} & -\Upsilon_t S_t \Delta_t^{-1} \\ -\Delta_t^{-1} S_t^T \Upsilon_t & \Delta_t^{-1} \end{pmatrix},$$

where $\Delta_t = Q_t^{-1} - S_t^T \Upsilon_t^{-1} S_t$, and $\Upsilon_t^{-1} = P_{t-1|t-1}^{-1} + R_t^{-1}$ yields that

$$\begin{aligned} P_{t|t} &= A_t \Upsilon_t A_t^T + A_t \Upsilon_t S_t \Delta_t^{-1} S_t^T \Upsilon_t A_t^T - A_t \Upsilon_t S_t \Delta_t^{-1} B_t^T + B_t \Delta_t^{-1} B_t^T - B_t \Delta_t^{-1} S_t^T \Upsilon_t A_t^T \\ &= A_t \Upsilon_t A_t^T + (A_t \Upsilon_t S_t - B_t) \Delta_t^{-1} (S_t^T \Upsilon_t A_t^T - B_t^T). \end{aligned}$$

The result (3.5) follows by induction since $J_0^{-1} = P_{0|0}$.

PROOF OF THEOREM 3.2.2. Consider the stochastic vector

$$\xi_t = \begin{pmatrix} \hat{x}_{t-1} - x_{t-1} \\ \hat{x}_t - A_t(s_t) x_{t-1} - B_t(s_t) v_t - F_t(s_t) u_t \\ \nabla_{x_{t-1}} \log p(v_t, x_{t-1}, s_{t-1:t}, y_{1:t-1}) \\ \nabla_{v_t} \log p(v_t, x_{t-1}, s_{t-1:t}, y_{1:t-1}) \end{pmatrix},$$

where

$$p(v_t, x_{t-1}, s_{t-1:t}, y_{1:t-1}) = p(v_t) p(s_t \,|s_{t-1}) p(y_{t-1} \,|x_{t-1}, s_t) p(x_{t-1}, s_{t-1}, y_{1:t-2}).$$

The correlation matrix of ξ_t is

$$\text{(B.3)} \qquad E(\xi_t \xi_t^T) = \begin{pmatrix} H_t & M_t^T \\ M_t & K_t \end{pmatrix} \geq 0,$$

where

$$H_t = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \qquad M_t = \begin{pmatrix} I & 0 \\ A_t & B_t \end{pmatrix} \qquad K_t = \begin{pmatrix} J_t + R_t^{-1} & 0 \\ 0 & I \end{pmatrix}.$$

Assume that a bound $E((\hat{x}_{t-1}(y_{1:t-2})-x_{t-1})(\hat{x}_{t-1}(y_{1:t-2})-x_{t-1})^T) \geq P_{t-1|t-2}$ is known. It follows from the PCRB that $P_{t-1|t-2}^{-1} \geq J_t$, and hence from (B.3) that

$$\begin{pmatrix} H_{11} & H_{12} & I & A_t^T \\ H_{21} & H_{22} & 0 & B_t^T \\ I & 0 & P_{t-1|t-2}^{-1} + R_t^{-1} & 0 \\ A_t & B_t & 0 & I \end{pmatrix} \geq 0.$$

Thus we have that

$$\begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \geq \begin{pmatrix} I & 0 \\ A_t & B_t \end{pmatrix} \begin{pmatrix} (P_{t-1|t-2}^{-1} + R_t^{-1})^{-1} & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} I & A_t^T \\ 0 & B_t^T \end{pmatrix},$$

and hence that

$$\begin{aligned} &E((\hat{x}_t(y_{1:t-1}) - x_t)(\hat{x}_t(y_{1:t-1}) - x_t)^T) \\ &\quad \geq A_t(P_{t-1|t-2} - P_{t-1|t-2}(P_{t-1|t-2} + R_t)^{-1}P_{t-1|t-2})A_t^T + B_t B_t^T. \end{aligned}$$

The result (3.7) follows by induction.

REFERENCES

Anderson, B. D. O. and Moore, J. B. (1979). *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, New Jersey.

Bar-Shalom, Y. and Li, X. R. (1995). *Multitarget-multisensor Tracking: Principles and Techniques*, YBS Publishing.

Bergman, N. (1999). Recursive Bayesian estimation: Navigation and tracking applications, Ph.D. Thesis, Department of Automatic Control, University of Linköping, Sweden.

Bergman, N., Ljung, L. and Gustafsson, F. (1999). Terrain navigation using Bayesian statistics, *IEEE Control Systems Magazine*, **19**, 33–40.

Borobsky, B. Z. and Zakai, M. (1975). A lower bound on the estimation error for Markov processes, *IEEE Trans. Automat. Control*, **20** (6), 785–788.

Carter, C. K. and Kohn, R. (1994). On Gibbs sampling for state space models, *Biometrika*, **81** (3), 541–553.

Carter, C. K. and Kohn, R. (1996). Markov chain Monte Carlo methods in conditionally Gaussian state space models, *Biometrika*, **83**, 589–601.

Crisan, D. (2001). Particle filters—A theoretical perspective, *Sequential Monte Carlo Methods in Practice* (eds. A. Doucet, De Freitas and N. J. Gordon), ISBN: 0-387-95146-6, Springer, New York.

Crisan, D. and Doucet, A. (2000). Convergence of sequential Monte Carlo methods, CUED/F-INFENG/TR381, Cambridge University.

Crisan, D., Del Moral, P. and Lyons, T. (1999). Discrete filtering using branching and interacting particle systems, *Markov Processes and Related Fields*, **5** (3), 293–318.

Del Moral, P. and Guionnet, A. (1998). *On the Stability of Measure Valued Processes. Applications to Non Linear Filtering and Interacting Particle Systems*, Laboratoire de Statistiques et Probabilités, Toulouse.

Doerschuk, P. C. (1995). Cramér-Rao bounds for discrete-time nonlinear filtering problems, *IEEE Trans. Automat. Control*, **40** (8), 1465–1469.

Doucet, A. (1997). Monte Carlo methods for Bayesian estimation of hidden Markov models, Ph.D. Thesis, Department of Electrical Engineering, University of Paris-Sud, Orsay, France (in French).

Doucet, A. and Andrieu, C. (1999). Iterative algorithms for state estimation of jump Markov linear systems, CUED/F-INFENG/TR353, Cambridge University.

Doucet, A., de Freitas, J. F. G. and Gordon N. J. (eds.) (2001). *Sequential Monte Carlo Methods in Practice*, ISBN: 0-387-95146-6, Springer, New York.

Doucet, A., Godsill S. J. and Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering, *Statist. Comput.*, **10** (3), 197–208.

Früwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models, *J. Time Ser. Anal.*, **15**, 183–202.

Galdos, J. I. (1980). A Cramér-Rao bound for discrete-time nonlinear filtering problems, *IEEE Trans. Automat. Control*, **25** (1), 117–119.

Gilks, W. R. and Berzuini, C. (1998). Following a moving target—Monte Carlo inference for dynamic Bayesian models (*submitted to J. Roy. Statist. Soc. Ser. B*).

Gill, R. D. and Levit, B. Y. (1995). Applications of the Van Trees inequality: a Bayesian Cramér Rao bound, *Bernoulli*, **1**, 59–79.

Gordon, N. J., Salmond, D. J. and Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *IEE Proceedings-F*, **140** (2), 107–113.

Higuchi, T. (1997). Monte Carlo filter using the genetic algorithm operators, *J. Statist. Comput. Simulation*, **59** (1), 1–23.

Kerr, T. (1989). Status of Cramér-Rao lower bounds for nonlinear filtering, *IEEE Transactions on Aerospace and Electronic Systems*, **25**, 590–601.

Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, *J. Comput. Graph. Statist.*, **5** (1), 1–25.

Kitagawa, G. and Gersch, W. (1996). *Smoothness Priors Analysis of Time Series*, Springer, New York.

Liu, J. S. and Chen, R. (1998). Sequential Monte Carlo methods for dynamic systems, *J. Amer. Statist. Assoc.*, **93**, 1032–1044.

Pitt, M. K. and Shephard, N. (1999). Filtering via simulation: auxiliary particle filters, *J. Amer. Statist. Assoc.*, **94**, 590–599.

Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*, Springer, New York.

Shephard, N. (1994). Partial non-Gaussian times series models, *Biometrika*, **81**, 115–131.

Tichavský, P., Muravchik, C. and Nehorai, A. (1998). Posterior Cramér-Rao bounds for discrete-time nonlinear filtering. *IEEE Transactions on Signal Processing*, **46** (5), 1386–1396.

van Trees, H. L. (1968). *Detection, Estimation and Modulation Theory*, Wiley, New York.

West, M. (1993*a*). Approximating posterior distributions by mixtures, *J. Roy. Statist. Soc. Ser. B*, **55**, 409–422.

West, M. (1993*b*). Mixture models, Monte Carlo, Bayesian updating and dynamic models, *Computing Science and Statistics*, **24**, 325–333.

West, M. and Harrison, P. J. (1997). *Bayesian Forecasting and Dynamic Models*, 2nd ed. Springer, New York.