Particle Based Smoothed Marginal MAP Estimation for General State Space Models

Saikat Saha, *Member, IEEE*, Pranab Kumar Mandal, Arunabha Bagchi, Yvo Boers, and Johannes N. Driessen

***Abstract*—We consider the smoothing problem for a general state space system using sequential Monte Carlo (SMC) methods. The marginal smoother is assumed to be available in the form of weighted random particles from the SMC output. New algorithms are developed to extract the smoothed marginal maximum *a posteriori* (MAP) estimate of the state from the existing marginal particle smoother. Our method does not need any kernel fitting to obtain the posterior density from the particle smoother. The proposed estimator is then successfully applied to find the unknown initial state of a dynamical system and to address the issue of parameter estimation problem in state space models.**

***Index Terms*—Maximum *a posteriori*, particle smoother, sequential monte carlo, unknown initial conditions.**

I. Introduction

CONSIDER a state-space model

$$x_t = f(x_{t-1}, w_t), \tag{1}$$

$$y_t = h(x_t, v_t), \quad t = 1, 2, \ldots \tag{2}$$

where x_t is the (unobserved) state with initial density $p(x_0)$ and y_t is the measurement at time step t. The process noises w_t, $t = 1, 2, \ldots$, are assumed to be independent. So are the measurement noises v_t, $t = 1, 2, \ldots$. Furthermore, (w_t) is assumed to be independent of (v_t). In this model, we assume that the probability density functions for w_t and v_t are known. The main problem related to model (1)–(2) is concerned with estimating the unknown state x_t given the set of measurements $y_{1:s} = \{y_1, \ldots, y_s\}$. The complete solution is, of course, given by the posterior probability density function $p(x_t|y_{1:s})$, which reflects all knowledge about the current state x_t. However, for a general nonlinear dynamic system, this posterior is often analytically intractable, but can be successfully approximated using a SMC approach, also known as a particle filter. In such an approach the posterior is approximated by a cloud of N weighted particles, whose empirical measure closely approximates the true posterior distribution for large N (see, e.g., [1]–[4]).

In this article[1], we focus on the smoothing problem, that is, to estimate x_t based on the measurements $y_{1:T}$, where $T > t$. In particular, we develop algorithms using SMC technique to calculate the smoothed (marginal) MAP estimate, $x_{t|T}^{\mathrm{MAP}}$, given by

$$x_{t|T}^{\mathrm{MAP}} = \arg\max_{x_t} p(x_t|y_{1:T}), \tag{3}$$

where $p(x_t|y_{1:T})$ is the (smoother) posterior density.

Only a limited number of methods exists in the literature that deal with the MAP estimation from the Monte Carlo based particle approximation. The main difficulty in obtaining the MAP lies in extracting the posterior density, whose maximizer is to be found, from the particle filter/smoother. Authors in [6], [7] use the particle with the maximum weight as the MAP estimate. This, however, does not necessarily represent the true MAP (the mode or maximizer of the posterior density) and it can actually be far from it ([8]–[10]; see also the example in Sections III.A of this article). The main reason behind this is, of course, the fact that the weights do not represent the (posterior) *density* at the particle-value.

The method proposed in [11] can be used if one is interested in the MAP sequence estimates of the whole path, $x_{0:t}$, up to the current time t. It uses the collection of all the particles up to the current time to form a trellis representation of the state space and subsequently, run a so-called Viterbi algorithm to find the path in the trellis with the highest posterior density. This, being a (joint) MAP of $x_{0:t}$, is however not necessarily the same as the *marginal* MAP of x_t, in which we are interested.

In this article we use the existing Monte Carlo based particle approximations of the marginal smoother density $p(x_t|y_{1:T})$ to provide the marginal MAP estimate. The more commonly used such marginal particle smoother in the literature is the so-called forward-backward smoother (see, for example, [12]). This smoother reuses the support points (particles) generated during the forward (filtering) pass and only recalculates the weights during the backward (smoothing) pass to derive the smoother approximation. To avoid the reliance on the particles from the forward phase, the two-filter smoother has been envisaged in [13]–[15], where one combines samples from particle filter in the forward direction with those from a so called "backward information filter" to produce the (weighted) cloud representation of $p(x_t|y_{1:T})$. As mentioned earlier, the

Manuscript received February 10, 2012; revised July 26, 2012; accepted September 25, 2012. Date of publication October 09, 2012; date of current version December 20, 2012. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Maria Sabrina Greco. The work of S. Saha was supported by a research grant from THALES Nederland B.V.

S. Saha is with the Division of Automatic Control, Department of Electrical Engineering, Linköping University, Linköping 581 83, Sweden (e-mail: saha@isy.liu.se).

P. K. Mandal and A. Bagchi are with the Department of Applied Mathematics, University of Twente, Enschede 7522 NB, The Netherlands (e-mail: p.k.mandal@ewi.utwente.nl; a.bagchi@ewi.utwente.nl

Y. Boers and J. N. Driessen are with the Thales Nederland B.V., Hengelo 7554 RR, The Netherlands (e-mail: yvo.boers@nl.thalesgroup.com; hans.driessen@nl.thalesgroup.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2012.2223691

[1]Part of the results were presented in EUSIPCO 2008 [5].

crux of the problem lies in constructing the posterior density from the weighted cloud representation of the distribution. As is known, one classic approach is the kernel method, where a kernel is fitted around each particle to approximate the posterior density [16]. The main drawback of this method is that it requires the user to choose a kernel bandwidth parameter. The density estimate is very sensitive to this parameter and the choice of an "optimal" value is not at all obvious ([17], [18]). Also, the kernel method is more suitable in a static set up than in a dynamic set up, which is the case for us. In the latter case, with the kernel density approach, one needs to go through the "optimal" selection of the bandwidth at each time-step (with each new-coming data). The time aspect will only increase if the sate-vector is multidimensional because one will then need to select an "optimal" bandwidth for each dimension.

The novel contribution of this article is to estimate the MAP, given by (3), using only the available (weighted) particle representation of the marginal smoother $p(x_t|y_{1:T})$, that is to say that, without requiring any exogenous method such as kernel fitting and thereby avoiding completely the process of choosing a non-obvious "optimal" parameter. The proposed method is simple yet elegant and uses the power of Monte Carlo samples, namely that it can be used to approximate very effectively an integral with respect to the density from which the samples are drawn. The idea is based on the fact that, even though the MAP estimate cannot be computed as integrals with respect to the posterior density, the posterior smoother density $f_{x_t|y_{1:t}}(x)$ of x_t evaluated at any point x can be expressed as an integral with respect to the posterior smoother density of x_{t+1}.

We should note that it is not our goal to compare the different existing particle smoothers as estimates of the true posterior $p(x_t|y_{1:T})$. Neither do we claim that MAP is superior to other estimators, such as the minimum mean squared estimator (MMSE). MAP estimates are, however, known to be useful [8] when the posterior is multimodal, which appears in a natural manner in many real life applications, e.g., in terrain aided navigation [19] and in target tracking problems [20]–[22]. When smoothing is also essential one would need a MAP smoother. For instance, consider the fingerprinting localization in wireless network based on received signal strength (RSS) measurements [23]. In such a situation, a so called radio map is constructed off-line using RSS measurements at different (known) locations by drive tests. The ground vehicle locations are in turn, determined by the inexpensive global positioning system (GPS) and inertial navigation system (INS) fusion platform, available in modern cars. However, the main problem with these systems are the frequent GPS outage in urban environment and the drift in INS error that grows with time. For the above problems, smoothing is shown to improve the position estimation [24], [25]. Moreover, due to multi-path propagations and non-line-of-sight (NLOS) conditions, observation error for GPS signal is typically non Gaussian, which often leads to multimodal posterior. So smoothed marginal MAP estimator can be a good candidate for such an application. Another potential application is the ground target tracking or road map extraction from smoothed tracks [26].

The rest of the article is organized as follows. We describe in Section II the proposed methodologies to obtain the MAP for both the particle smoother mentioned above. In Section II.A we first review briefly the method used to obtain the particle smoother based on forward-backward smoothing. Subsequently, we describe how to obtain the MAP and demonstrate the performance of this MAP estimator through a generic nonlinear time series model. The same is done for two filter smoothing in Section II.B. Section III deals with the applications of the proposed marginal MAP smoother. We begin, in Section III.A, by validating the proposed estimator based on forward-backward particle smoother using a linear Gaussian model. We also confirm in this example that the particle with the maximum weight does not represent the true MAP estimator (mode of the posterior density). In Section III.B, the proposed MAP smoother is applied to estimate the unknown initial state of a given dynamic system, which is subsequently used in Section III.C in connection with the parameter estimation problem of a dynamic system. Finally, we conclude the article in Section IV.

II. Particle Based Smoothed Marginal MAP Estimator (PS-MAP)

As mentioned earlier, our starting point in this article is that there already exists a (weighted) particle cloud for the marginal smoother. Based on the weighted cloud representation, we calculate the smoothed marginal density and subsequently, extract the MAP from it. We start with the most commonly used forward-backward smoother and then describe our algorithm for the two filter smoother. We emphasize once again that our purpose is not to compare these different smoothing methods. Rather we focus on extracting the marginal MAP from the available particle cloud generated by either of these methods.

A. Forward-Backward Smoothing (FBS)

The marginal smoother by forward-backward algorithm is based on the relationship (see, e.g., [12])

$$p(x_t|y_{1:T}) = p(x_t|y_{1:t}) \int \frac{p(x_{t+1}|y_{1:T})p(x_{t+1}|x_t)}{p(x_{t+1}|y_{1:t})} dx_{t+1}, \quad (4)$$

where, $p(x_t|y_{1:t})$ and $p(x_{t+1}|y_{1:t})$ are the filtering density and one step ahead predictive density respectively, at time t. Thus, starting with $p(x_T|y_{1:T})$, one can recursively obtain $p(x_t|y_{1:T})$ from $p(x_{t+1}|y_{1:T})$. Using the above recursion, the marginal smoothing distribution can now be approximated by the weighted particle cloud as described, for example, in [13]. Here, one starts with the forward filtering pass for computing the filtered distribution at each time step using the particle filter as

$$\hat{P}(dx_t|y_{1:t}) = \sum_{i=1}^{N} \omega_t^{(i)} \delta_{x_t^{(i)}}(dx_t), \quad (5)$$

where $\delta_{x_t}(dx_t)$ denotes the Dirac delta mass located at x_t. Then one performs the backward smoothing pass as given by (4) to approximate the smoothing distribution

$$\hat{P}(dx_t|y_{1:T}) = \sum_{i=1}^{N} \omega_{t|T}^{(i)} \delta_{x_t^{(i)}}(dx_t), \quad (6)$$

where the smoothing weights are obtained through the following backward recursion:

$$\omega_{t|T}^{(i)} = \omega_t^{(i)} \sum_{j=1}^{N} \left[\omega_{t+1|T}^{(j)} \frac{p\left(x_{t+1}^{(j)}|x_t^{(i)}\right)}{\sum_{k=1}^{N} p\left(x_{t+1}^{(j)}|x_t^{(k)}\right) \omega_t^{(k)}} \right] \quad (7)$$

with $\omega_{T|T}^{(i)} = \omega_T^{(i)}$. It is important to note that the forward-backward smoother keeps the same particle support as used in filtering step and re-weights the particles to obtain the approximated particle based smoothed distribution. Thus, success of this method crucially hinges on the filtered distribution having supports where the smoothed distribution is significant.

To obtain the smoothed marginal MAP, one needs the posterior density $p(x_t|y_{1:T})$ from the above cloud representation. Here, we proceed as follows. Using the Bayes' rule, one can write the one step ahead predictive density in (4) as

$$p(x_{t+1}|y_{1:t}) = \frac{p(x_{t+1}|y_{1:t+1})p(y_{t+1}|y_{1:t})}{p(y_{t+1}|x_{t+1})}. \quad (8)$$

Then (4) becomes

$$\begin{aligned} p(x_t|y_{1:T}) &= p(x_t|y_{1:t}) \\ &\quad \times \int \frac{p(x_{t+1}|y_{1:T})p(x_{t+1}|x_t)p(y_{t+1}|x_{t+1})}{p(x_{t+1}|y_{1:t+1})p(y_{t+1}|y_{1:t})} dx_{t+1} \\ &= \frac{p(x_t|y_{1:t})}{p(y_{t+1}|y_{1:t})} \int \left[\frac{p(x_{t+1}|x_t)p(y_{t+1}|x_{t+1})}{p(x_{t+1}|y_{1:t+1})} \right] \\ &\qquad \times p(x_{t+1}|y_{1:T}) dx_{t+1} \\ &\approx \frac{p(x_t|y_{1:t})}{p(y_{t+1}|y_{1:t})} \int \left[\frac{p(x_{t+1}|x_t)p(y_{t+1}|x_{t+1})}{p(x_{t+1}|y_{1:t+1})} \right] \\ &\qquad \times \hat{P}(dx_{t+1}|y_{1:T}). \end{aligned}$$

Making use of the particle representation of $\hat{P}(dx_t|y_{1:T})$, given by (6), and subsequently approximating the above integration by a Monte Carlo integration method, one obtains

$$\begin{aligned} p(x_t|y_{1:T}) &\approx \frac{p(x_t|y_{1:t})}{p(y_{t+1}|y_{1:t})} \\ &\quad \times \sum_{j=1}^{N} \left[\frac{p\left(x_{t+1}^{(j)}|x_t\right) p\left(y_{t+1}|x_{t+1}^{(j)}\right)}{p\left(x_{t+1}^{(j)}|y_{1:t+1}\right)} \right] \omega_{t+1|T}^{(j)}. \end{aligned} \quad (9)$$

Further approximating the filtered density $p(x_{t+1}|y_{1:t+1})$ from the running particle filter [8] as

$$p(x_{t+1}|y_{1:t+1}) \approx \frac{p(y_{t+1}|x_{t+1}) \sum_k p\left(x_{t+1}|x_t^{(k)}\right) w_t^{(k)}}{p(y_{t+1}|y_{1:t})} \quad (10)$$

we can rewrite (9) as

$$p(x_t|y_{1:T}) \approx p(x_t|y_{1:t}) \sum_{j=1}^{N} \left[\frac{p\left(x_{t+1}^{(j)}|x_t\right)}{\sum_{k=1}^{N} p\left(x_{t+1}^{(j)}|x_t^{(k)}\right) \omega_t^{(k)}} \right] \omega_{t+1|T}^{(j)}. \quad (11)$$

The smoothed marginal density, $p(x_t|y_{1:T})$ is obtained at any support point x_t and the corresponding MAP estimate can then be extracted by finding the location of its global maximum. At this point, one can in principle, employ any standard optimization technique to arrive at the MAP estimate. In general, however, this maximization step is nontrivial due to the possible multimodalities arising from the non Gaussian nature of the posterior.

Following the argument, as used in [11], that the particles form a randomized adaptive grid approximation of the values of the posterior, we select the particle at which the density is the highest as the MAP, i.e.,

$$\begin{aligned} x_{t|T}^{\mathrm{MAP}} &\approx \arg\max_{x_t^{(i)}} p\left(x_t^{(i)}|y_{1:t}\right) \\ &\quad \times \sum_{j=1}^{N} \left[\frac{p\left(x_{t+1}^{(j)}|x_t^{(i)}\right)}{\sum_{k=1}^{N} p\left(x_{t+1}^{(j)}|x_t^{(k)}\right) \omega_t^{(k)}} \right] \omega_{t+1|T}^{(j)}, \end{aligned} \quad (12)$$

where N is the number of particles used at each time step. By using (7), the estimator can be further simplified to

$$x_{t|T}^{\mathrm{MAP}} \approx \arg\max_{x_t^{(i)}} p\left(x_t^{(i)}|y_{1:t}\right) \frac{\omega_{t|T}^{(i)}}{\omega_t^{(i)}}, \quad (13)$$

where the filtered density $p(x_t|y_{1:t})$ at the particle cloud $\{x_t^{(i)}\}_{i=1}^{N}$ can be evaluated during the forward filtering step [8] as

$$p\left(x_t^{(i)}|y_{1:t}\right) \approx \frac{p\left(y_t|x_t^{(i)}\right) \sum_j p\left(x_t^{(i)}|x_{t-1}^{(j)}\right) w_{t-1}^{(j)}}{p(y_t|y_{1:t-1})}. \quad (14)$$

Since $p(y_t|y_{1:t-1})$ in (14) is independent of $x_t^{(i)}$, to obtain $x_{t|T}^{\mathrm{MAP}}$, one can replace $p(x_t^{(i)}|y_{1:t})$ in (13) by the unnormalized filtered density

$$q\left(x_t^{(i)}|y_{1:t}\right) = p\left(y_t|x_t^{(i)}\right) \sum_j p\left(x_t^{(i)}|x_{t-1}^{(j)}\right) w_{t-1}^{(j)}. \quad (15)$$

The summary of the procedure is presented in Algorithm 1.

Algorithm 1: FBS marginal MAP

Input: $y_0, \ldots, y_T$

1. FBS initialization:
   a) set $p(x_0|x_{-1}) = p(x_0)$ (known)
   b) draw $x_0^{(i)} \sim p(x_0),\ i = 1, \ldots, N$
   c) compute $\omega_0^{(i)} = p(y_0|x_0^{(i)})$ and normalize
2. run FBS
3. available FBS output: $\{x_t^{(i)},\ \omega_t^{(i)},\ \omega_{t|T}^{(i)}\}, t = 0, \ldots, T$, where $\omega_{T|T}^{(i)} = \omega_T^{(i)}$
4. estimate the marginal MAP using (13) and (15)
   a) for $t = 1, \ldots, T$,

$$x_{t|T}^{\mathrm{MAP}} \approx \arg\max_{x_t^{(i)}} \left\{ p\left(y_t|x_t^{(i)}\right) \sum_j p\left(x_t^{(i)}|x_{t-1}^{(j)}\right) w_{t-1}^{(j)} \right\} \frac{\omega_{t|T}^{(i)}}{\omega_t^{(i)}}.$$

   b) for $t = 0$,

$$x_{0|T}^{\mathrm{MAP}} \approx \arg\max_{x_0^{(i)}} p\left(x_0^{(i)}\right) \omega_{0|T}^{(i)}$$

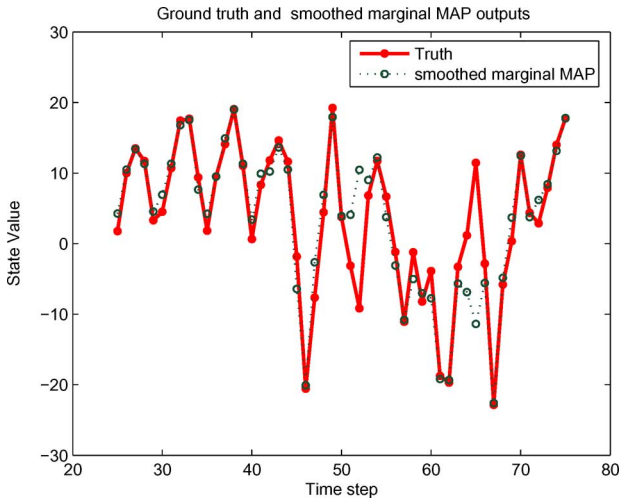


Fig. 1. Ground truth and smoothed marginal MAP outputs.

We note here that a numerical problem may arise in evaluating (13) if the filtered weights attached to some particles are very small. This may happen when the "particle degeneracy" occurs. This problem can be effectively addressed using a combination of efficient importance proposal (see, e.g., [27]) along with resampling steps.

To demonstrate the computation of FBS based smoothed marginal MAP, we consider the nonlinear time series model:

$$x_t = \frac{x_{t-1}}{2} + \frac{25x_{t-1}}{1+x_{t-1}^2} + 8\cos(1.2t) + w_t, \tag{16}$$

$$y_t = \frac{x_t^2}{20} + v_t, \quad t = 1, 2, \ldots \tag{17}$$

where $w_t \sim N(0, 10)$ and $v_t \sim N(0, 1)$ with initial prior $p(x_0) \sim N(0, 5)$. The above model is highly nonlinear and when the measured data is (large) positive, state density may become symmetric bimodal. In fact, this model has become a *de facto* benchmark problem in the particle filtering community due to the attractive nonlinear and/or non Gaussian characteristics (see e.g., [1], [9], [2]). For this nonlinear problem, we use the "Exact Moment matching (EMM) proposal" as in [28] during forward filtering step with particle sample size $N = 500$ and $T = 200$. The smoothed marginal MAP outputs along with the ground truth for $t = 25, \ldots, 75$ is shown in Fig. 1.

B. Two-Filter Smoothing (TFS)

We describe, in this section, how the smoothed marginal MAP can be obtained from the particle cloud generated by the generalized two-filter smoother. We start with a brief description of how two filter particle smoother is obtained. For this, we follow [13].

In the two-filter smoother framework, the so-called backward information filter $p(y_{t:T}|x_t)$ is calculated sequentially from $p(y_{t+1:T}|x_{t+1})$ as

$$p(y_{t:T}|x_t) = p(y_t|x_t) \int p(x_{t+1}|x_t) p(y_{t+1:T}|x_{t+1}) dx_{t+1}. \tag{18}$$

As noted by [13], $p(y_{t:T}|x_t)$ is not a probability density function in x_t and actually, its integral over x_t may not even be finite. The smoothing algorithm in [14], [15] assumes implicitly that $\int p(y_{t:T}|x_t)dx_t < \infty$. However, if this assumption does not hold, SMC based methods, which can only approximate finite measures, will not work anymore. To avoid this, "generalized two-filter smoothing" has been proposed by [13], where the smoothing distributions are computed through a combination of forward filter and an auxiliary probability distribution $\tilde{p}(x_t|y_{t:T})$ in argument x_t. This auxiliary density is defined through a sequence of artificial distributions $\gamma_t(x_t)$ as

$$\tilde{p}(x_t|y_{t:T}) \propto \gamma_t(x_t) p(y_{t:T}|x_t).$$

It then follows from (18) that

$$\tilde{p}(x_t|y_{t:T}) \propto \gamma_t(x_t) p(y_t|x_t) \times \int p(x_{t+1}|x_t) \frac{\tilde{p}(x_{t+1}|y_{t+1:T})}{\gamma_{t+1}(x_{t+1})} dx_{t+1}. \tag{19}$$

This in turn, is used to generate recursively the weighted particle representation of the backward information filter

$$\tilde{p}(dx_t|y_{t:T}) \simeq \sum_{k=1}^{N} \tilde{\omega}_t^{(k)} \delta_{\tilde{x}_t^{(k)}}(dx_t). \tag{20}$$

The marginal smoother $p(x_t|y_{1:T})$ is then computed by combining the outputs of the forward filter (FF) and the backward information filter (BIF) as

$$p(x_t|y_{1:T}) \propto p(x_t|y_{1:t-1}) p(y_{t:T}|x_t) = \left(\int p(x_t|x_{t-1}) p(x_{t-1}|y_{1:t-1}) dx_{t-1} \right) \left(\frac{\tilde{p}(x_t|y_{t:T})}{\gamma_t(x_t)} \right). \tag{21}$$

Evaluating the integral in (21) by Monte Carlo integration using the forward filter cloud $(x_{t-1}^{(j)}, \omega_{t-1}^{(j)})$ one obtains

$$p(x_t|y_{1:T}) \propto \left(\sum_{j=1}^{N} p\left(x_t|x_{t-1}^{(j)}\right) \omega_{t-1}^{(j)} \right) \left(\frac{\tilde{p}(x_t|y_{t:T})}{\gamma_t(x_t)} \right). \tag{22}$$

Finally, the particle cloud representation is obtained using the cloud $(\tilde{x}_t^{(k)}, \tilde{\omega}_t^{(k)})$ from the backward filter:

$$p(dx_t|y_{1:T}) \simeq \sum_{k=1}^{N} \tilde{\omega}_{t|T}^{(k)} \delta_{\tilde{x}_t^{(k)}}(dx_t) \tag{23}$$

where

$$\tilde{\omega}_{t|T}^{(k)} \propto \frac{\tilde{\omega}_t^{(k)}}{\gamma_t\left(\tilde{x}_t^{(k)}\right)} \sum_{j=1}^{N} p\left(\tilde{x}_t^{(k)} \middle| x_{t-1}^{(j)}\right) \omega_{t-1}^{(j)}. \tag{24}$$

Thus, in essence the particles from the forward filter are used to re-weight those from the backward filter so that they represent the marginal smoother distribution. We refer the readers to the original article by [13] for more details.

Algorithm 2: TFS marginal MAP

Input: $y_0, \ldots, y_T$

1. FF initialization:
   a) set $p(x_0|x_{-1}) = p(x_0)$ (known)
   b) draw $x_0^{(i)} \sim p(x_0),\ i = 1, \ldots, N$
   c) compute $\omega_0^{(i)} = p(y_0|x_0^{(i)})$ and normalize
2. run FF for $t = 1, \ldots, T$
3. available FF output: $\{x_t^{(i)},\ \omega_t^{(i)}\},\ t = 0, \ldots, T$
4. BIF initialization:
   a) $\{\tilde{x}_T^{(j)}, \tilde{\omega}_T^{(j)}\}_{j=1}^N$ obtained by resampling $\{x_T^{(i)}, \omega_T^{(i)}\}_{i=1}^N$
5. select $\gamma_t(\cdot)$ for $t = 0, \ldots, T$
6. run BIF for $t = T-1, \ldots, 0$
7. available BIF output: $\{\tilde{x}_t^{(j)}, \tilde{\omega}_t^{(j)}\}_{j=1}^N,\ t = 0, \ldots, T$
8. TFS output:
   a) (from (23)–(24)) $\{\tilde{x}_t^{(j)}, \tilde{\omega}_{t|T}^{(j)}\}_{j=1}^N$ for $t = 1, \ldots, T$
   b) $\{\tilde{x}_0^{(j)}, \tilde{\omega}_{0|T}^{(j)}\}_{j=1}^N$ for $t = 0$ where
$$\tilde{\omega}_{0|T}^{(k)} \propto \frac{\tilde{\omega}_0^{(k)}}{\gamma_0\left(\tilde{x}_0^{(k)}\right)} p\left(\tilde{x}_0^{(k)}\right).$$
9. estimate the marginal MAP (using (28) and (25)) as
$$x_{t|T}^{\mathrm{MAP}} = \arg\max_{\tilde{x}_t^{(i)}} \left\{ \gamma_t\left(\tilde{x}_t^{(i)}\right) p\left(y_t|\tilde{x}_t^{(i)}\right) \times \sum_{k=1}^N \frac{p\left(\tilde{x}_{t+1}^{(k)}|\tilde{x}_t^{(i)}\right)}{\gamma_{t+1}\left(\tilde{x}_{t+1}^{(k)}\right)} \tilde{\omega}_{t+1}^{(k)} \right\} \frac{\tilde{\omega}_{t|T}^{(i)}}{\tilde{\omega}_t^{(i)}}.$$

Now we describe how to derive the smoothing density from the particle smoother obtained as above. Note that using (20) one can rewrite (19) as

$$\tilde{p}(x_t|y_{t:T}) \propto \gamma_t(x_t) p(y_t|x_t) \sum_{k=1}^N \frac{p\left(\tilde{x}_{t+1}^{(k)}\middle| x_t\right)}{\gamma_{t+1}\left(\tilde{x}_{t+1}^{(k)}\right)} \tilde{\omega}_{t+1}^{(k)}. \quad (25)$$

It then follows from (22) that

$$p(x_t|y_{1:T}) \propto \left(\sum_{j=1}^N p\left(x_t|x_{t-1}^{(j)}\right) \omega_{t-1}^{(j)} \right) \times \left(p(y_t|x_t) \sum_{k=1}^N \frac{p\left(\tilde{x}_{t+1}^{(k)}\middle| x_t\right)}{\gamma_{t+1}\left(\tilde{x}_{t+1}^{(k)}\right)} \tilde{\omega}_{t+1}^{(k)} \right). \quad (26)$$

The required smoothed marginal MAP can now be obtained by maximizing the unnormalized smoothing density, given by the right hand side of (26). Furthermore, when this maximization is done along the particles $\tilde{x}_t^{(i)}$, we have

$$p\left(\tilde{x}_t^{(i)}\middle| y_{1:T}\right) \propto \left(\sum_{j=1}^N p\left(\tilde{x}_t^{(i)}\middle| x_{t-1}^{(j)}\right) \omega_{t-1}^{(j)} \right) \times \left(p\left(y_t|\tilde{x}_t^{(i)}\right) \sum_{k=1}^N \frac{p\left(\tilde{x}_{t+1}^{(k)}\middle|\tilde{x}_t^{(i)}\right)}{\gamma_{t+1}\left(\tilde{x}_{t+1}^{(k)}\right)} \tilde{\omega}_{t+1}^{(k)} \right)$$

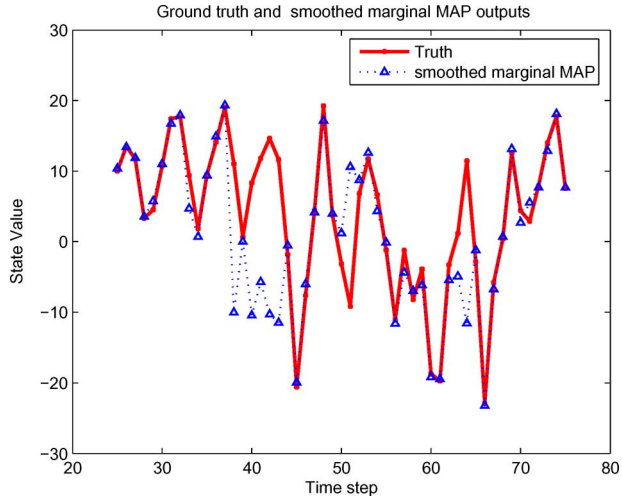


Fig. 2. Ground truth and smoothed marginal MAP outputs.

$$= \left(\frac{1}{\gamma_t\left(\tilde{x}_t^{(i)}\right)} \sum_{j=1}^N p\left(\tilde{x}_t^{(i)}\middle| x_{t-1}^{(j)}\right) \omega_{t-1}^{(j)} \right) \times \left(\gamma_t\left(\tilde{x}_t^{(i)}\right) p\left(y_t|\tilde{x}_t^{(i)}\right) \sum_{k=1}^N \frac{p\left(\tilde{x}_{t+1}^{(k)}\middle|\tilde{x}_t^{(i)}\right)}{\gamma_{t+1}\left(\tilde{x}_{t+1}^{(k)}\right)} \tilde{\omega}_{t+1}^{(k)} \right).$$

From (24) and (25) this reduces to

$$p\left(\tilde{x}_t^{(i)}|y_{1:T}\right) \propto \left(\frac{\tilde{\omega}_{t|T}^{(i)}}{\tilde{\omega}_t^{(i)}}\right) \left(\tilde{p}\left(\tilde{x}_t^{(i)}\middle| y_{t:T}\right)\right). \quad (27)$$

Hence, the required MAP can be obtained as

$$x_{t|T}^{\mathrm{MAP}} = \arg\max_{\tilde{x}_t^{(i)}} \tilde{p}\left(\tilde{x}_t^{(i)}\middle| y_{t:T}\right) \frac{\tilde{\omega}_{t|T}^{(i)}}{\tilde{\omega}_t^{(i)}}, \quad (28)$$

where $\tilde{p}(\tilde{x}_t^{(i)}|y_{1:T})$ is evaluated using (25). A summary of the procedure is presented in Algorithm 2. We now demonstrate the computation of TFS based smoothed marginal MAP on the same time series model as given in (16)–(17) with particle sample size $N = 1000$ and $T = 200$. We use the EMM proposal during forward filtering step. For this example, we select the artificial density $\gamma_t(x_t)$ to be a time-invariant density $\gamma(x_t)$ as in [13], but approximated by a mixture of two Gaussian densities $\gamma(x_t) = W_1 N(\mu_1, \Sigma_1) + (1 - W_1) N(\mu_2, \Sigma_2)$. For calculating $\gamma(x_t)$, a long sample path $x_{0:T_f}$ ($T_f \gg T$) is simulated using (16) and then the Gaussian mixture is fitted to the empirical measure $\hat{\pi}(dx) = (1)/(T_f - T_0) \sum_{t=T_0+1}^{T_f} \delta X_t(dx)$, where T_0 is the burn-in period. We select $T_0 = 2000$ and $T_f = 20000$. Next we use an EM like algorithm ([29], with $\beta = 0$) to select the parameters of the mixture. Starting with $W_1^{(\mathrm{ini})} = 0.5$, $\mu_1^{(\mathrm{ini})} = 5$, $\Sigma_1^{(\mathrm{ini})} = 200$, $\mu_2^{(\mathrm{ini})} = -5$ and $\Sigma_2^{(\mathrm{ini})} = 200$, the final estimates are obtained as $W_1^{(*)} = 2.3987 \cdot 10^{-007}$, $\mu_1^{(*)} = 16.2565$, $\Sigma_1^{(*)} = 0.6273$, $\mu_2^{(*)} = -0.0343$ and $\Sigma_2^{(*)} = 109.2576$. The backward proposal is taken to be a Gaussian approximation of the optimal backward proposal $p_{\mathrm{opt}}(x_t|x_{t+1}, y_t)$. This is obtained as follows. We first approximate the joint distribution of x_t, x_{t+1} and y_t by a Gaussian distribution with matching moments up to second order, calculated numerically over 50 Monte Carlo runs. From the theory

of multivariate Gaussian, it then follows that $p(x_t|x_{t+1}, y_t)$ is Gaussian. The smoothed marginal MAP outputs along with the ground truth for $t = 25, \ldots, 75$ is shown in Fig. 2.

III. Numerical Examples

A. Validation of the FBS Based Smoothed Marginal MAP and Comparison With Other Estimators

In this section we validate the proposed MAP estimator[2]. We do this on the basis of a linear-Gaussian model. In particular, we verify numerically that the proposed estimator converges to the true MAP (the analytical solution), given by the Kalman smoother. For this we consider a simplified one dimensional target tracking [20] problem where a nearly constant velocity model is used for the state dynamics. We use the state vector $x_t \triangleq [p_t, v_t]^T$, where the scalar variables p_t and v_t denote the position and velocity, respectively, of the target and y_t is the measurement at time step t. The discrete time state space model is given by:

$$x_{t+1} = \begin{pmatrix} 1 & \Delta \\ 0 & 1 \end{pmatrix} x_t + \begin{pmatrix} \Delta^2/2 \\ \Delta \end{pmatrix} w_t \tag{29}$$

$$y_t = (\,1\;0\,)x_t + e_t, \quad t = 1, 2, \ldots, T. \tag{30}$$

Here $\Delta = 4$ is the measurement scan interval, w_t and e_t are the process and measurement noises, respectively, given by $w_t \sim \mathcal{N}(0, 5^2)$ and $e_t \sim \mathcal{N}(0, 20^2)$. The noise sequences are serially independent and also independent of each other. The initial state is assumed to be distributed according to a zero mean Gaussian random variable with a diagonal covariance matrix $\begin{pmatrix} 100 & 0 \\ 0 & 1 \end{pmatrix}$. For the simulation we have used $T = 30$. Since the above state space model is linear Gaussian, the exact smoothed marginal MAP can be obtained analytically using a Kalman smoother running on the same data. For the particle filter, we use state transition density as proposal with resampling at every step. Next, we compute the smoothed marginal MAP using different number of particles and compare them with the exact smoothed marginal MAP as obtained from the Kalman smoother. The accuracy of the proposed MAP is assessed in terms of the root mean square error (RMSE), given by

$$\text{RMSE} = \left(\frac{1}{M} \sum_{j=1}^{M} \left(\hat{a}_t^j - a_t^j \right)^2 \right)^{\frac{1}{2}} \tag{31}$$

where M is the number of Monte Carlo runs, $\hat{a}_t^j$ is our proposed estimate of a desired quantity (say position p_t or velocity v_t) at time t for j-th (Monte Carlo) run and a_t^j is the corresponding output from the Kalman smoother. The mean and standard deviation (Std) of the RMSE values (over the $T = 30$ time steps) against different number of particles are shown in Table I below:

The results indicate that with increasing number of particles, the RMSE values converge to a limit. It should not be surprising that the limit is not zero. From the definition (31) it is clear that even if the MAP estimates converge to the true MAP, the RMSE would not converge to zero. Since both the MAP estimate and the true MAP are time dependent and stochastic, it would converge, under ergodicity conditions, to some sort of standard deviation of the error of the estimate, which is determined by the variances of the noise processes. From the convergence of RMSE values we conclude that the errors of the estimates remain within bound.

[2]We stress again our assumption that the particle cloud representation is given to us. Thus any specific particle filter implementation is irrelevant to our problem formulation.

TABLE I
RMSE as a Function of Number of Particles Used

Nr. of particles	**Position**		**Velocity**	
	mean	Std	mean	Std
50	25.6564	11.6652	18.2127	4.1586
250	9.6446	5.0851	17.1901	4.3301
500	7.0625	1.6735	16.1167	4.1403
1000	6.6446	1.6810	16.4073	4.1379
2000	6.0519	1.6673	15.5771	3.9074

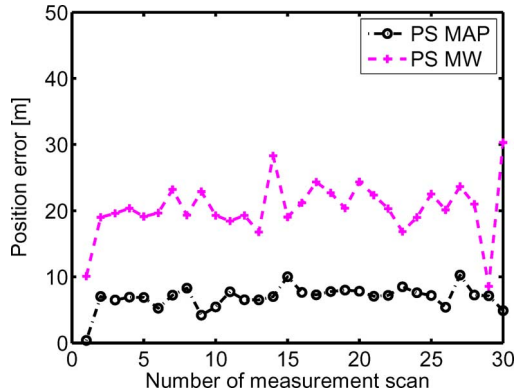


Fig. 3. RMSE position error.

For the sake of completeness, we also compare the performance of the proposed smoothed marginal MAP (PS MAP) with that of the smoother particle with the maximum weight (PS MW). We use a particle smoother with $N = 2000$ particles. We estimate the (squared) errors of the estimators with respect to the exact smoothed marginal MAP (Kalman smoother). The root mean squared position and velocity errors over 30 Monte Carlo runs and over 30 time steps are shown in Figs. 3 and 4, respectively. We see in Fig. 3 that the RMSE of the position estimates when using the PS-MW is quite far off as compared to that of the PS-MAP. This is a clear indication that the PS MW estimator does not represent the true marginal MAP. On the other hand, Fig. 4 shows that the two estimators behave almost similarly for the velocity component of the RMSE, with PS MW performing a bit worse than PS MAP.

Next, we apply our marginal smoother MAP estimator to estimate the unknown initial condition of the state. Subsequently, using the same approach, we have addressed parameter estimation problems by considering the parameter as an additional state.

B. Estimation of (Unknown) Initial Condition

1) Linear Model: We consider the following

$$x_t = 0.8x_{t-1} + w_t \tag{32}$$

$$y_t = x_t + v_t \tag{33}$$

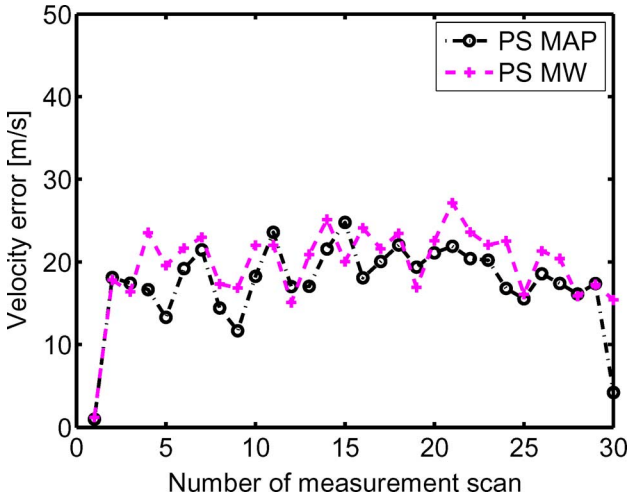


Fig. 4. RMSE velocity error.

TABLE II
MEAN AND VARIANCE OF ESTIMATED INITIAL STATE

$Mean(x_{0\|500}^{MAP})$	$Var(x_{0\|500}^{MAP})$
9.9726	0.0915

TABLE III
MEAN AND VARIANCE OF ESTIMATED INITIAL STATE

$Mean(x_{0\|500}^{MAP})$	$Var(x_{0\|500}^{MAP})$
9.7165	0.9236

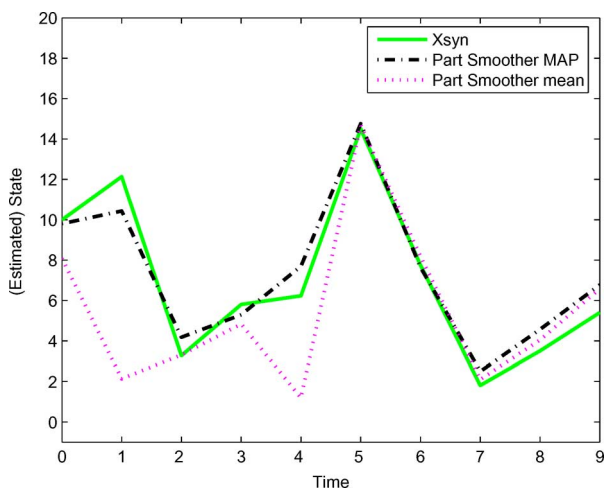


Fig. 5. Simulated state (Xsyn), MAP and mean of the marginal smoothed posterior for the first 10 time steps.

with $w_t \sim N(0,1)$ and $v_t \sim N(0,0.1)$. The initial state x_0 is assumed to be unknown (constant). The simulated data $\{x_t, y_t\}_{t=0:500}$ is generated starting with $x_0^* = 10$. To estimate the unknown initial state x_0, we start with initial prior $p(x_0) \sim U[0,\ 20]$ where $U[a,\ b]$ denotes uniform probability density function with lower bound a and upper bound b respectively. We use $N = 500$ particles and the optimal proposal as given in [2] in the forward filtering step. The estimate of the initial unknown state is taken as the MAP of $p(x_0|y_{0:T})$[3]. The mean and variance of the estimator over 30 Monte Carlo runs are shown in Table II. The result shows that the smoothed initial density peaks around the true initial state, even though we have started with a pretty wide uniform initial prior.

Though not needed for this exercise, we have nonetheless calculated for all $0 \le t \le T = 500$, the proposed MAP for the smoothed marginal density $p(x_t|y_{0:T})$ and the corresponding mean. We notice in the simulations that the MAP and the mean of the smoothed marginal density at a given time step are the same, as expected for a linear Gaussian model.

2) Nonlinear Model: Here, we consider the nonlinear time series model as given in (16)–(17). The simulated data $\{x_t, y_t\}_{t=0:500}$ is generated starting with $x_0^* = 10$. As in the previous case, we start with initial prior $p(x_0) \sim U[0,\ 20]$. For this nonlinear problem, we use the EMM proposal during forward filtering step with particle sample size $N = 500$. The estimate of the initial unknown state is given by the particle based MAP of $p(x_0|y_{0:T})$. We repeat this MAP state estimate for 30 Monte Carlo runs. The mean and variance of the estimator are shown in Table III. The result in Table III is really remarkable as we can see by comparing with Table II. Even for highly nonlinear model as considered above and with wide uniform initial prior, the result is almost as good as in linear case. Of course the variance is somewhat larger, but that is to be expected given the highly nonlinear nature of the problem.

It is also interesting to study the behavior of the smoother when the initial distribution is supported on a larger interval. Starting with $p(x_0) \sim U[-40,\ 40]$, we have calculated for all $0 \le t \le T = 500$, the proposed MAP for the smoothed marginal density $p(x_t|y_{0:T})$ and the corresponding mean. These estimates for the first 10 time steps for a particular realization are shown in Fig. 5 while the corresponding filtered and smoothed pdfs (unnormalized versions) for x_0 are shown in Fig. 6. We notice that the filtered as well as the smoothed pdf are bimodal with (local) peaks around the true initial state, $x_0^* = 10$ and its reflected value -10. In the filtered version both peaks are equally high (suggesting the inability of the filter to decide between these two values), whereas in the smoothed version, the peak at $+10$ is much higher. This shows the improved performance of the smoother in comparison with the filtered density. Furthermore, although the dominant mode of the smoother density is very close to the true initial state $x_0^* = 10$, the contribution from the weaker mode, shifts the smoothed mean away from x_0^* (as seen in Fig. 5, the smoothed mean is near 8 here). This further strengthens the justification of using the MAP in such a scenario.

C. Parameter Estimation

One of the common approaches of estimating a parameter in a state-space model is to augment the parameter as an extra state with a small artificial dynamics and then take the filtered estimate as the estimate of the parameter. The artificial evolution, however, in effect, renders the fixed parameter into a slowly

[3]With an uniform prior $p(x_0)$, note from (13) that for estimating the initial condition, we are essentially picking the (smoothed) particle with the highest weight. However, with other choice of prior, the estimate (i.e., the smoothed marginal MAP) is different from the particle with highest weight.

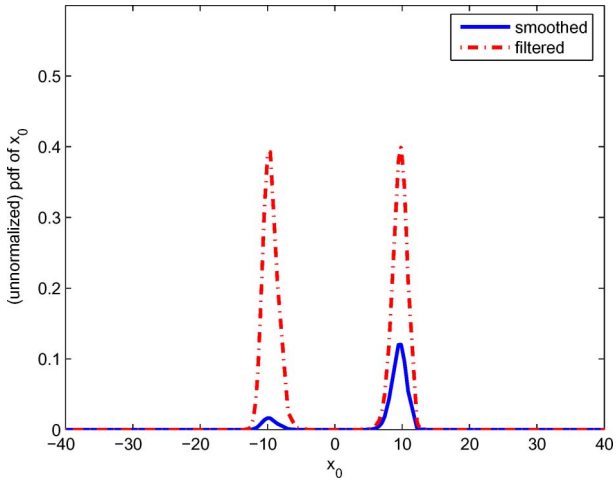


Fig. 6. (Unnormalized) posterior probability density functions for the initial state x_0 (filtered: $p(x_0|y_0)$ and smoothed: $p(x_0|y_{0:500})$).

TABLE IV
TRUE PARAMETER, MEAN AND STANDARD DEVIATION OF THE ESTIMATED PARAMETER

True parameter	$Mean(\theta_{0\|500}^{MAP})$	$Std(\theta_{0\|500}^{MAP})$
0.5	0.422	0.265

varying one. As a result, the variance of the filtered estimate of the parameter increases over time [30], which limits the precision of the resulting estimate. Looking from another perspective at this augmented framework, one may observe that only the initial augmented state is not corrupted by artificial noise.

Hence in our approach, we consider the marginal smoother of the initial augmented state to be the estimate of the true (fixed) parameter. It is expected that as more and more observations are available, the smoothed estimate would converge to the true parameter value. We proceed here with the following dynamic system:

$$x_{t+1} = f(x_t, w_{t+1}; \theta), \tag{34}$$
$$y_t = h(x_t, v_t), \quad t = 0, 1, \ldots \tag{35}$$

where θ is a fixed unknown parameter, (x_t) are the unobservable state with (known) initial prior density $p(x_0)$ and (y_t) are the observation. The process noises (w_t) are assumed to be independent of the measurement noises (v_t). We start with the usual procedure of augmenting the state space by treating the parameter as additional state. Note that the dimension of the state increases by the numbers of parameters augmented. Now the augmented state space can be written as

$$x_{t+1} = f(x_t, \theta_t, w_{t+1}) \tag{36}$$
$$\theta_{t+1} = \theta_t + \eta_{t+1} \tag{37}$$
$$y_t = h(x_t, v_t), \qquad t = 0, 1, \ldots \tag{38}$$

with $\theta_0 = \theta$, which is unknown here. Now, using notation $X_{t+1} = [x_{t+1}\ \theta_{t+1}]'$ and $W_{t+1} = [w_{t+1}\ \eta_{t+1}]'$, where $[\cdots]'$ denotes vector transpose, the above model can be rewritten as

$$X_{t+1} = \bar{f}(X_t, W_{t+1})$$

TABLE V
TRUE PARAMETER, MEAN AND STANDARD DEVIATION OF THE ESTIMATED PARAMETER

True parameter	$Mean(\theta_{0\|500}^{MAP})$	$Std(\theta_{0\|500}^{MAP})$
25.0	27.259	1.241

$$y_t = \bar{h}(X_t, v_t)$$

for some $\bar{f}$ and $\bar{h}$. We estimate the initial state vector X_0 using marginal the MAP smoother. The corresponding estimation for the augmented state θ_0 is taken as the estimated parameter. We consider the following two numerical examples for this parameter estimation approach. We begin with a linear example:

$$x_t = \theta x_{t-1} + w_t \tag{39}$$
$$y_t = x_t + v_t \tag{40}$$

with $w_t \sim N(0, 1)$ and $v_t \sim N(0, 0.1)$ and (unknown) true parameter $\theta = \theta^* = 0.5$. We take $\eta_t \sim N(0, 0.0025)$. Note that θ_0 is independent of x_0. With $p(x_0) \sim N(0, 5)$, we started with $p(\theta_0) \sim U[-5,\ 5]$. We use $N = 1000$ particles and state transition density as our proposal during forward filtering step. The mean and the standard deviation of the estimator of θ over 30 Monte Carlo runs are shown in Table IV. Although the assumption of uniform initial prior is radically different from the knowledge of exact initial condition (parameter), we see the parameter estimate to be quite good.

Next we consider the following nonlinear example:

$$x_t = \frac{x_{t-1}}{2} + \frac{\theta x_{t-1}}{1 + x_{t-1}^2} + 8\cos(1.2t) + w_t, \tag{41}$$
$$y_t = \frac{x_t^2}{20} + v_t, \tag{42}$$

where $w_t \sim N(0, 10)$ and $v_t \sim N(0, 1)$. The true parameter is $\theta = \theta^* = 25$. With known $p(x_0) \sim N(0, 5)$, we started with $p(\theta_0) \sim U[-50,\ 50]$. We use $N = 1000$ particles and state transition density as proposal during forward filtering step. We set $\eta_t \sim N(0, 5)$. The estimate of θ for 30 Monte Carlo runs is shown in Table V. As remarked after Table IV, we see the same pattern in a nonlinear problem as well. We observed that this estimation procedure works quite well even in nonlinear cases. However, the computational burden with the growing memory requirement is a major stumbling block here. Additionally, when the number of parameters is large, the dimension of X_{t+1} also increases and the effective exploration of the state space in region where the joint probability of X_{t+1} is high, becomes difficult with a finite number of (relatively small) samples.

IV. CONCLUSION

In this article we have considered the problem of estimating the smoothed marginal MAP $x_{t|T}^{\mathrm{MAP}}$, given by (3), of unobserved x_t from all the observations, $y_{1:T}$, up to time $T(>t)$, where (x_t), (y_t) follow a general state space model, given by (1)–(2). In doing so, we assume that a marginal particle smoother for the posterior $p(x_t|y_{1:T})$ already exists. The naive choice of the particle with maximum (smoothed) weight does not represent

the true MAP estimator as observed by the authors in [8] and [9]; and confirmed further by the example in Section III.A. The newly proposed estimator for the marginal smoother MAP is based on the theoretically sound fact that the posterior density evaluated at any arbitrary point can be expressed as an integral with respect to the posterior density from the "previous" time-step. The proposed method is self-sufficient in the sense that it does not need any exogenous method such as kernel fitting and thereby avoiding the non-obvious and computationally expensive choice of the optimal kernel bandwidth. The algorithm corresponding to the most commonly used forward-backward particle smoother is developed in Section II.A and that for the two filter smoother in Section II.B. We have performed a quick validation of the proposed estimator (using forward-backward smoother) in Section III.A. Here we have considered a linear Gaussian model for which the true MAP is given by the Kalman smoother. A numerical comparison of our estimator with the true MAP suggests that as the number of particle increases the proposed MAP estimates stay close to the true MAPs (the errors remain within bound). After the successful validation step we have applied the proposed MAP estimator to find the unknown initial state of a given dynamical system (Section III.B). We notice that even for highly nonlinear model with wide uniform initial prior the result is very good. This is subsequently applied (Section III.C) to address the parameter estimation problem in dynamical systems. We observe reasonably good results in this application as well.

REFERENCES

[1] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

[2] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statist. Comput.*, vol. 10, pp. 197–208, 2000.

[3] F. Gustafsson, "Particle filter theory and practice with positioning applications," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 25, no. 7, pp. 53–81, Mar. 2010.

[4] P. M. Djuric, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Máguez, "Particle filtering," *IEEE Signal Process. Mag.*, vol. 20, no. 5, pp. 19–38, 2003.

[5] S. Saha, P. K. Mandal, and A. Bagchi, "A new approach to particle based smoothed marginal MAP," presented at the EUSIPCO, Lausanne, Switzerland, Aug. 25–29, 2008.

[6] S. K. Zhou, R. Chellappa, and B. Moghaddam, "Visual tracking and recognition using appearance-adaptive models in particle filters," *IEEE Trans. Image Process.*, vol. 13, no. 11, pp. 1491–1506, Nov. 2004.

[7] J. V. Candy, "Bootstrap particle filtering," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 73–85, 2007.

[8] J. N. Driessen and Y. Boers, "MAP estimation in particle filter tracking," in *IET Seminar on Target Tracking and Data Fusion: Algorithms and Applications*, Birmingham, U.K., Apr. 2008, pp. 41–45.

[9] O. Cappé, S. J. Godsill, and E. Moulines, "An overview of existing methods and recent advances in sequential Monte Carlo," *IEEE Proc.*, vol. 95, no. 5, pp. 899–924, 2007.

[10] S. Saha, Y. Boers, H. Driessen, P. K. Mandal, and A. Bagchi, "Particle based MAP state estimation: A comparison," in *Proc. 12th Int. Conf. Inf. Fusion (FUSION)*, Seattle, WA, Jul. 6–9, 2009, pp. 278–283.

[11] S. Godsill, A. Doucet, and M. West, "Maximum a posteriori sequence estimation using Monte Carlo particle filters," *Ann. Inst. Statist. Math.*, vol. 53, pp. 82–96, 2001.

[12] G. Kitagawa, "Non-Gaussian state-space modeling of nonstationary time series," *J. Amer. Statist. Assoc.*, vol. 82, no. 400, pp. 1032–1063, 1987.

[13] M. Briers, A. Doucet, and S. Maskell, "Smoothing algorithm for state space models," *Ann. Inst. Statist. Math.*, vol. 62, no. 1, pp. 61–89, 2010.

[14] G. Kitagawa, "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models," *J. Comput. Graphic. Statist.*, vol. 5, no. 1, pp. 1–25, 1996.

[15] M. Isard and A. Blake, "A smoothing filter for condensation," in *Proc. 5th Eur. Conf. Comput. Vis.*, 1998, pp. 767–781.

[16] B. Silverman, *Density Estimation for Statistics and Data Analysis*. London, U.K.: Chapman & Hall/CRC, 1986.

[17] B. A. Turlach, "Bandwidth selection in kernel density estimation: A review," CORE and Institut de Statistique, Universite Catholique de Louvain, Louvain la Neuve, Belgium, Tech. Rep., 1993.

[18] M. C. Jones, J. S. Marron, and S. J. Sheather, "A brief survey of bandwidth selection for density estimation," *J. Amer. Statist. Assoc.*, vol. 91, no. 433, pp. 401–407, 1996.

[19] P.-J. Nordlund, "Efficient estimation and detection methods for airborne applications," Ph.D. dissertation, Electr. Eng. Dept., Linköping Univ., Linköping, Sweden, 2008.

[20] Y. Bar-Shalom and X. Li, *Multitarget-Multisensor Tracking: Principles and Techniques*. New York: Academic, 1995.

[21] H. A. P. Blom, E. A. Bloem, Y. Boers, and J. N. Driessen, "Tracking closely spaced targets: Bayes outperformed by an approximation?," in *Proc. 11th Int. Conf. Inf. Fusion*, Jun. 30–Jul. 3, 2008, pp. 1–8.

[22] Y. Boers, E. Sviestins, and J. N. Driessen, "Mixed labelling in multi target particle filtering," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 2, pp. 792–802, 2010.

[23] M. Bshara, U. Orguner, F. Gustafsson, and L. Biesen, "Fingerprinting localization in wireless networks based on received signal strength measurements: A case study on wimax networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 1, pp. 283–294, 2010.

[24] S. Godha and M. Cannon, "GPS/MEMS INS integrated system for navigation in urban areas," *GPS Solutions*, vol. 11, no. 3, pp. 193–203, 2007.

[25] H. Liu, S. Nassar, and N. Elsheimy, "Two-filter smoothing for accurate INS/GPS land-vehicle navigation in urban centers," *IEEE Trans. Veh. Technol.*, vol. 59, no. 9, pp. 4256–4267, 2010.

[26] M. Ulmke and W. Koch, "Road map extraction using GMTI tracking," in *Proc. 9th Int. Conf. Inf. Fusion*, Jul. 2006, pp. 1–7.

[27] S. Saha, P. K. Mandal, Y. Boers, H. Driessen, and A. Bagchi, "Gaussian proposal density using moment matching in SMC methods," *Statist. Comput.*, vol. 19, no. 2, pp. 203–208, Jun. 2009.

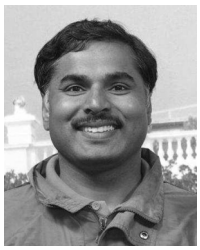
[28] S. Saha, P. K. Mandal, Y. Boers, and H. Driessen, "Exact moment matching for efficient importance functions in SMC methods," presented at the SSPW 2006, Cambridge, U.K., Sep. 13–15, 2006.

[29] H. Fujisawa and S. Eguchi, "Robust estimation in the normal mixture model," *J. Statist. Planning Inference*, vol. 136, no. 11, pp. 3989–4011, 2006.

[30] J. Liu and M. West, "Combined parameter and state estimation in simulation-based filtering," in *Sequential Monte Carlo Methods in Practice*, A. Doucet, J. de Freitas, and N. Gordon, Eds. New York: Springer-Verlag, 2001, pp. 197–217.

Saikat Saha (M'10) received the M.Sc. (Engg.) degree from the Indian Institute of Science, Bangalore, in 2003 and the Ph.D. degree from the University of Twente, The Netherlands, in 2009.

Between 2009 and 2011, he previously held a postdoctoral position with the Division of Automatic Control, Linköping University, Linköping, Sweden, where he is currently an Assistant Professor. His research interests include statistical signal processing, information fusion, system identification, and computational finance.



Pranab Kumar Mandal received the Master's degree in statistics from the Indian Statistical Institute, Kolkata, in 1992 and the Ph.D. degree in statistics from the University of North Carolina at Chapel Hill in 1997.

After his Ph.D., he held a visiting position at Michigan State University, East Lansing, and a Postdoctoral Research position at EURANDOM in The Netherlands. He is currently an Assistant Professor at the University of Twente, The Netherlands. His research interests include mathematical statistics and (nonlinear) filtering, in particular particle filtering with applications to statistical signal processing, system identification, and financial mathematics.

Arunabha Bagchi was born in Calcutta, India, in September 1947. He received the M.Sc. degree in applied mathematics from Calcutta University, India, in 1969 and the M.S. and Ph.D. degrees in engineering from the University of California, Los Angeles (UCLA), in 1970 and 1974, respectively.

Since 1974, he has been with the University of Twente, The Netherlands, where he is currently Professor of applied mathematics. He has been the founder and head of the FELab (Financial Engineering Laboratory) of the University of Twente. His current research interest mainly lies in particle filtering, distributed sensor network, and financial engineering. He is the author of *Optimal Control of Stochastic Systems* (Prentice-Hall International, 1993) and *Stackelberg Differential Games in Economic Models* (LCCIS 64, Springer-Verlag, 1984).

Dr. Bagchi has been Associate editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL and of *Automatica*. He was a Fulbright Scholar in 1998–1999 and has been Visiting Professor at UCLA, SUNY Stony Brook, Northeastern University, and the Indian Statistical Institute.

Yvo Boers received the M.Sc. degree in applied mathematics from Twente University, The Netherlands, in 1994 and the Ph.D. degree in electrical engineering from the Technical University Eindhoven, The Netherlands, in 1999.

Since 1999, he has been employed at Thales Nederland B.V. His research interests are in the areas of detection, (particle) filtering, target tracking, sensor networks, and control. He was an NWO-Casimir research fellow from 2008 to 2011 at the University of Twente in the field of distributed sensor systems.

Dr. Boers received, together with H. Driessen, the Best Paper Award at the FUSION 2006 conference in Florence, Italy. He has coedited several special issues for different journals. He was an Associate Editor for the International Society of Information Fusion (ISIF) journal, *Advances on Information Fusion*, from 2005 to 2008 and currently is an elected member of the Board of Directors for the ISIF, serving the term 2011 to 2013.

Johannes N. Driessen received the M.Sc. and Ph.D. degrees, both in electrical engineering, from the Delft University of Technology, Delft, The Netherlands, in 1987 and 1992, respectively.

In 1993, he joined Thales Nederland B.V. (formerly known as Hollandse Signaalapparaten B.V.) as a Design Engineer of plot processing and target tracking systems. He is currently R&D Manager in the area of signal and data processing and sensor management. His professional interests are in developing innovative sensor system concepts applying modern multi-target stochastic detection, estimation, classification, information, and control theory.