BACKWARD HIDDEN MARKOV CHAIN FOR OUTLIER-ROBUST FILTERING AND FIXED-INTERVAL SMOOTHING

Boujemaa Ait-El-Fquih *Cédric Gouy-Pailler*

CEA, LIST, Data Analysis Tools Laboratory, 91191 Gif-sur-Yvette CEDEX, FRANCE

ABSTRACT

This paper addresses the problem of recursive estimation of a process in presence of outliers among the observations. It focuses on deriving robust approximate Kalman-like backward filtering and backward-forward fixed-interval smoothing algorithms in the context of linear hidden Markov chain with heavy-tailed measurement noise. The proposed algorithms are derived based on the backward Markovianity of the model as well as the variational Bayesian approach. In a simulation design, our algorithms are shown to outperform the classical Kalman filter in the presence of outliers.

***Index Terms*—** Backward Markovian models, Robust filtering, Robust smoothing, Kalman-like algorithms, Variational Bayes.

1. INTRODUCTION

The estimation problem of an unobservable process $\mathbf{x} = \{\mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_N\}$ from an observed process $\mathbf{y} = \{\mathbf{y}_0, \mathbf{y}_1, \cdots, \mathbf{y}_N\}$ has a particular interest in a number of areas such as target tracking, autonomous navigation or wireless communications [1] [2] [3]. This problem is usually performed in the Linear and Gaussian Hidden Markov Chain (LGHMC) and thereby, a number of efficient *recursive* algorithms have been developed based on the dynamic structure of this model. Let $\mathbf{x}_n \in \mathbb{R}^{n_\mathbf{x}}$ and $\mathbf{y}_n \in \mathbb{R}^{n_\mathbf{y}}$. It has been shown that a LGHMC $(\mathbf{x}, \mathbf{y})$ is indeed a state-space model [4] [5],

$$\left\{ \begin{array}{lcl} \mathbf{x}_{n+1} & = & \mathbf{F}_n \mathbf{x}_n + \mathbf{u}_n \\ \mathbf{y}_n & = & \mathbf{H}_n \mathbf{x}_n + \mathbf{v}_n \end{array} \right. ; \quad (1)$$

the input noise $\mathbf{u} = \{\mathbf{u}_n\}_{n=0}^N$ and the measurement noise $\mathbf{v} = \{\mathbf{v}_n\}_{n=0}^N$ are assumed to be independent, jointly independent and independent of $\mathbf{x}_0$; and $\mathbf{x}_0$, $\mathbf{u}_n$ and $\mathbf{v}_n$ are Gaussian. Let $\mathbf{x}_0 \sim \mathcal{N}(\hat{\mathbf{x}}_0, \mathbf{P}_0)$, $\mathbf{u}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_n)$, $\mathbf{v}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_n)$, $\mathbf{x}_{0:n} = \{\mathbf{x}_i\}_{i=0}^n$ and $\mathbf{y}_{0:n} = \{\mathbf{y}_i\}_{i=0}^n$. Let also $p(\mathbf{x}_n)$ and $p(\mathbf{x}_n|\mathbf{y}_{0:n})$, say, denote the probability density function (pdf) (w.r.t. Lebesgue measure) of $\mathbf{x}_n$ and the pdf of $\mathbf{x}_n$ conditional on $\mathbf{y}_{0:n}$, respectively ; the other pdf's are defined similarly. A fundamental problem associated to model (1) (the so-called *filtering*) consists in estimating, in each time n, the state $\mathbf{x}_n$ from the measurements $\mathbf{y}_{0:n}$. The classical solution is given by the *a posteriori* mean which minimizes the mean square error. On the other hand, it happens that the Kalman Filter (KF) algorithm has been introduced as an indisputable tool insofar as it allows an exact and recursive computation of this solution [6] [7]. However, in presence of outliers among the observations, *i.e.*, if there exist some samples that lie outside the data set $\mathbf{y}_{0:N}$ [8], the Gaussian assumption under $\mathbf{v}_n$ breaks down leading to the degradation of the performance of KF. This problem originates from the lightweight tails of the Gaussian distribution due to which the KF considers that none outlier is present.

To overcome this drawback and improve the robustness to outliers, the measurement noise should be modelled by a heavy-tailed distribution. In this perspective, an auxiliary independent Gamma process $w = \{w_n\}_{n\in\mathbb{N}}$ has been introduced in [9] (see also [10] and references therein) such that $\mathbf{v}$ is Gaussian conditional on w and

$$p(\mathbf{v}_n|w_n) \sim \mathcal{N}\left(\mathbf{0}, \frac{\mathbf{R}_n}{w_n}\right), \; p(w_n) \sim \Gamma(\frac{\alpha_n}{2}, \frac{\beta_n}{2}). \quad (2)$$

More precisely, the model used in [9] can be considered as a particular case of (1)-(2), insofar $\mathbf{F}_n$, $\mathbf{H}_n$, $\mathbf{Q}_n$ and $\mathbf{R}_n$ are constant, and that $\mathbf{Q}_n$ and $\mathbf{R}_n$ are diagonal. Now, it has been shown that (2) implies that $\mathbf{v}_n$ is Student's *t*-distributed rather than Gaussian (see e.g. [11]). Moreover, the ratio $\frac{\mathbf{R}_n}{w_n}$ describes the fact that a weight w_n is assigned to the variance [1] of each data sample $\mathbf{y}_n$, and thus, as we will see below, indicates its contribution in the estimation of $\mathbf{x}_n$. On the other hand, an approximate robust KF-like algorithm has been proposed in [9] based on the Variational Bayesian (VB) approach [12] [13]. This work was then extended in [14] [15] to the case in which $\mathbf{Q}_n$ and $\mathbf{R}_n$ are no longer diagonal, and therefore, the so-called *fixed-interval smoothing* problem, which consists in estimating $\mathbf{x}_n$, $n = 0, 1, \cdots, N$ from the whole data set $\mathbf{y}_{0:N}$ has been also treated. For this problem an approximate *forward-backward* KF-like smoother was derived.

Our contribution is based on the fact that model (1)-(2) is a conditionally LGHMC in the backward direction, *i.e.*, for decreasing values of n. We thus exploit the backward Markovianity of our model in order to address two problems. The first is

1. Note that $p(\mathbf{v}_n|w_n)$ and $p(\mathbf{y}_n|\mathbf{x}_n, w_n)$ have the same covariance matrix (or the same variance in the scalar case).

the so-called *backward filtering* whose the aim is to estimate $\mathbf{x}_n$ from $\mathbf{y}_{n:N}$ in the backward direction. The second is the fixed-interval smoothing for which a new *backward-forward* algorithm will be derived, in contrast to [14], who proposes a reversed algorithm (a *forward-backward* algorithm). Let us turn to the content of this paper. In section 2 we introduce the backward model associated to (1)-(2). We then derive in section 3 a robust VB KF-like algorithm allowing the propagation of $p(\mathbf{x}_n, w_n|\mathbf{y}_{n:N})$ in the backward direction. A robust VB KF-like smoother is introduced in section 4 for propagating $p(\mathbf{x}_n, w_n|\mathbf{y}_{0:N})$ in the forward direction. Some simulations are provided in section 5. We finally emphasize on positioning our contributions compared to previous works in section 6.

2. BACKWARD FILTERING IN BACKWARD HMC WITH HEAVY-TAILED MEASUREMENT NOISE

Let us turn back to model (1)-(2). Let $\mathbf{z}_n = [\mathbf{x}_n^T, w_n]^T$. Then the following properties hold :

$$\begin{aligned} p(\mathbf{z}_n|\mathbf{z}_{n+1:N}) &= p(\mathbf{z}_n|\mathbf{z}_{n+1}), & (3)\\ &= p(\mathbf{x}_n|\mathbf{x}_{n+1})p(w_n), & (4)\\ p(\mathbf{y}_{n:N}|\mathbf{z}_{n:N}) &= \prod_{i=n}^{N} p(\mathbf{y}_i|\mathbf{z}_i). & (5)\end{aligned}$$

In other words, $\mathbf{z}$ is a Markov Chain (MC) in the backward direction, and since $\mathbf{z}$ is known only through the observed process $\mathbf{y}$, $(\mathbf{z}, \mathbf{y})$ is a backward HMC model. Now from (3) and (5) we get

$$\begin{aligned} p(\mathbf{z}_n|\mathbf{z}_{n+1}, \mathbf{y}_{n+1:N}) &= p(\mathbf{z}_n|\mathbf{z}_{n+1}), & (6)\\ p(\mathbf{y}_n|\mathbf{z}_n, \mathbf{y}_{n+1:N}) &= p(\mathbf{y}_n|\mathbf{z}_n). & (7)\end{aligned}$$

As a consequence, the backward filtering pdf $p(\mathbf{z}_n|\mathbf{y}_{n:N})$ is *recursively* computed in the backward direction by :

$$\begin{cases} p(\mathbf{z}_n|\mathbf{y}_{n+1:N}) = \int p(\mathbf{z}_n|\mathbf{z}_{n+1})p(\mathbf{z}_{n+1}|\mathbf{y}_{n+1:N})d\mathbf{z}_{n+1} \\ p(\mathbf{z}_n|\mathbf{y}_{n:N}) \quad \propto p(\mathbf{y}_n|\mathbf{z}_n)p(\mathbf{z}_n|\mathbf{y}_{n+1:N}) \end{cases}. \quad (8)$$

Aside from the linear and Gaussian case for which a backward KF algorithm has been already introduced [16] [17], the exact computation of (8) is generally impossible in practice. On the other hand, one can verify that

$$\begin{aligned} p(\mathbf{x}_n|\mathbf{x}_{n+1}) &\sim \mathcal{N}(\tilde{\mathbf{F}}_{n+1}\mathbf{x}_{n+1} + \tilde{\mathbf{b}}_{n+1}, \tilde{\mathbf{Q}}_{n+1}), & (9)\\ p(\mathbf{y}_n|\mathbf{x}_n, w_n) &\sim \mathcal{N}(\mathbf{H}_n\mathbf{x}_n, \frac{\mathbf{R}_n}{w_n}), & (10)\\ p(w_n) &\sim \Gamma(\frac{\alpha_n}{2}, \frac{\beta_n}{2}), & (11)\end{aligned}$$

with $p(\mathbf{x}_n) \sim \mathcal{N}(\hat{\mathbf{x}}_n, \mathbf{P}_n)$, $\tilde{\mathbf{F}}_{n+1} = \mathbf{P}_n\mathbf{F}_n^T\mathbf{P}_{n+1}^{-1}$, $\tilde{\mathbf{b}}_{n+1} = (\mathbf{I}_{n_\mathbf{x}} - \tilde{\mathbf{F}}_{n+1}\mathbf{F}_n)\hat{\mathbf{x}}_n$, $\tilde{\mathbf{Q}}_{n+1} = (\mathbf{I}_{n_\mathbf{x}} - \tilde{\mathbf{F}}_{n+1}\mathbf{F}_n)\mathbf{P}_n$ and $\mathbf{I}_{n_\mathbf{x}}$ is the $n_\mathbf{x} \times n_\mathbf{x}$ identity matrix. Now, as far as the estimation of $\mathbf{x}_n$ is concerned, the development of a backward KF-like algorithm allowing the computation of the mean and the covariance matrix of $p(\mathbf{x}_n|w_n, \mathbf{y}_{n:N})$ is feasible, since, as we can see, $(\mathbf{x}, \mathbf{y})$ is a backward LGHMC conditional on w. Similarly, since the Gamma distribution belongs to the family of conjugate priors, analytic computation of the mean of $p(w_n|\mathbf{x}_n, \mathbf{y}_{n:N})$ is tractable. However, our aim is rather the calculation of parameters of $p(\mathbf{x}_n|\mathbf{y}_{n:N})$ and $p(w_n|\mathbf{y}_{n:N})$, and consequently, it is most convenient to propose tools allowing the suppression of the conditional dependence of $\mathbf{x}_n$ and w_n. For this purpose, we use VB approach [12] [13] and we develop a robust backward KF-like algorithm.

3. A ROBUST BACKWARD VB KALMAN FILTER

The aim of VB approach is to approximate the joint pdf $p(\mathbf{x}_n, w_n|\mathbf{y}_{n:N})$ by a separable product of marginal densities $q(\mathbf{x}_n|\mathbf{y}_{n:N}) \times q(w_n|\mathbf{y}_{n:N})$ in the sense of the minimization of the Kullback-Leibler (KL) divergence. Due to the exponential nature of our problem we get [13] :

$$q(\mathbf{x}_n|\mathbf{y}_{n:N}) \propto \exp\left(\int \ln(p(\mathbf{z}_n, \mathbf{y}_{n:N}))q(w_n|\mathbf{y}_{n:N})dw_n\right), \quad (12)$$

$$q(w_n|\mathbf{y}_{n:N}) \propto \exp\left(\int \ln(p(\mathbf{z}_n, \mathbf{y}_{n:N}))q(\mathbf{x}_n|\mathbf{y}_{n:N})d\mathbf{x}_n\right). \quad (13)$$

It remains to compute (12)-(13) in a recursive way based on the dynamical structure of model (9)-(11). The proposed algorithm holds in two steps : time-update and measurement-update.

3.1. Time-update step

We assume that the joint backward filtering pdf is approximated as $p(\mathbf{z}_{n+1}|\mathbf{y}_{n+1:N}) \approx q(\mathbf{x}_{n+1}|\mathbf{y}_{n+1:N})q(w_{n+1}|\mathbf{y}_{n+1:N})$ and we wish to compute an approximation

$$p(\mathbf{z}_n|\mathbf{y}_{n+1:N}) \approx q(\mathbf{x}_n|\mathbf{y}_{n+1:N})q(w_n|\mathbf{y}_{n+1:N}), \quad (14)$$

of the so-called backward prediction pdf. By using (4) in the 1^{st} eq. of (8) we get

$$\begin{cases} q(\mathbf{x}_n|\mathbf{y}_{n+1:N}) = \int p(\mathbf{x}_n|\mathbf{x}_{n+1})q(\mathbf{x}_{n+1}|\mathbf{y}_{n+1:N})d\mathbf{x}_{n+1} \\ q(w_n|\mathbf{y}_{n+1:N}) = p(w_n) \end{cases} \quad (15)$$

Equalities (15) show that the separability of the joint filtering pdf leads to that of the prediction pdf. On the other hand, computing $p(\mathbf{z}_n|\mathbf{y}_{n+1:N})$ by the 1^{st} eq. of (8) amounts to compute the marginal pdf $q(\mathbf{x}_n|\mathbf{y}_{n+1:N})$ by (15). From now on we set $\hat{w}_{n|i:j} = \mathbb{E}_{q(w_n|\mathbf{y}_{i:j})}[w_n]$, $\hat{\mathbf{x}}_{n|i:j} = \mathbb{E}_{q(\mathbf{x}_n|\mathbf{y}_{i:j})}[\mathbf{x}_n]$ and $\mathbf{P}_{n|i:j} = \mathbb{E}_{q(\mathbf{x}_n|\mathbf{y}_{i:j})}[(\mathbf{x}_n - \hat{\mathbf{x}}_{n|i:j})(\mathbf{x}_n - \hat{\mathbf{x}}_{n|i:j})^T]$, for all n, i, j with $0 \leq n \leq N$ and $0 \leq i \leq j \leq N$.

One can show that the 1^{st} eq. of (15) reduces to the time-update step of the backward KF [16] [17] :

$$\begin{aligned} \hat{\mathbf{x}}_{n|n+1:N} &= \tilde{\mathbf{F}}_{n+1}\hat{\mathbf{x}}_{n+1|n+1:N} + \tilde{\mathbf{b}}_{n+1}, & (16)\\ \mathbf{P}_{n|n+1:N} &= \tilde{\mathbf{F}}_{n+1}\mathbf{P}_{n+1|n+1:N}\tilde{\mathbf{F}}_{n+1}^T + \tilde{\mathbf{Q}}_{n+1}. & (17)\end{aligned}$$

3.2. Measurement-update step

We now address the computation of (12) and (13). We start by computing $p(\mathbf{z}_n, \mathbf{y}_{n:N})$. From (7), (14) and 2^{nd} eq. of (15) the following factorization hold :

$$p(\mathbf{z}_n, \mathbf{y}_{n:N}) \approx \mathcal{C}_1 p(\mathbf{y}_n|\mathbf{x}_n, w_n) q(\mathbf{x}_n|\mathbf{y}_{n+1:N}) p(w_n), \quad (18)$$

$\mathcal{C}_1$ is independent of $\mathbf{x}_n$ and w_n. Furthermore, by injecting (10) and (11) in (18) we get

$$\ln(p(\mathbf{z}_n, \mathbf{y}_{n:N})) \approx \mathcal{C}_2 - \frac{1}{2}\left[(-n_{\mathbf{y}} - \alpha_n + 2)\ln(w_n) + w_n(\beta_n + ||\mathbf{y}_n - \mathbf{H}_n\mathbf{x}_n||^2_{\mathbf{R}_n^{-1}}) + ||\mathbf{x}_n - \hat{\mathbf{x}}_{n|n+1:N}||^2_{\mathbf{P}^{-1}_{n|n+1:N}}\right] \quad (19)$$

with $\mathcal{C}_2$ is independent of $\mathbf{x}_n$ and w_n, and for some vector $\mathbf{v}$ and positive definite matrix $\mathbf{M}$ we set $||\mathbf{v}||^2_{\mathbf{M}} \stackrel{\text{def}}{=} \mathbf{v}^T\mathbf{M}\mathbf{v}$.

Now, as far as the computation of $q(\mathbf{x}_n|\mathbf{y}_{n:N})$ is concerned, by injecting (19) and (11) in (12) one can show that $q(\mathbf{x}_n|\mathbf{y}_{n:N})$ is Gaussian with

$$\mathbf{P}_{n|n:N} = \left[\mathbf{P}^{-1}_{n|n+1:N} + \hat{w}_{n|n:N}\mathbf{H}_n^T\mathbf{R}_n^{-1}\mathbf{H}_n\right]^{-1} \quad (20)$$

$$\hat{\mathbf{x}}_{n|n:N} = \mathbf{P}_{n|n:N}\left[\mathbf{P}^{-1}_{n|n+1:N}\hat{\mathbf{x}}_{n|n+1:N} + \hat{w}_{n|n:N}\mathbf{H}_n^T\mathbf{R}_n^{-1}\mathbf{y}_n\right] \quad (21)$$

Let note that (20)-(21) have the form of the measurement-update step in information form, of the backward KF with $\frac{\mathbf{R}_n}{\hat{w}_{n|n:N}}$ as measurement-noise covariance matrix.

On the other hand, following (19) the computation of (13) leads to a Gamma pdf $q(w_n|\mathbf{y}_{n:N}) \sim \Gamma(\frac{\tilde{\alpha}_n}{2}, \frac{\tilde{\beta}_n}{2})$ with

$$\tilde{\alpha}_n = \alpha_n + n_{\mathbf{y}} \quad (22)$$

$$\tilde{\beta}_n = \beta_n + ||\mathbf{y}_n - \mathbf{H}_n\hat{\mathbf{x}}_{n|n:N}||^2_{\mathbf{R}_n^{-1}} + \mathrm{Tr}\left[\mathbf{H}_n^T\mathbf{R}_n^{-1}\mathbf{H}_n\mathbf{P}_{n|n:N}\right], \quad (23)$$

whence the *posterior* mean $\hat{w}_{n|n:N} = \frac{\tilde{\alpha}_n}{\tilde{\beta}_n}$.

However, the quantities of interest $(\hat{\mathbf{x}}_{n|n:N}, \mathbf{P}_{n|n:N})$ and $\tilde{\beta}_n$ (then $\hat{w}_{|n:N}$) are coupled, which makes impossible the exact resolution of (20), (21) and (23). A classical solution may be used to avoid this drawback, consists in fixed-point iteration such that $(\hat{\mathbf{x}}_{n|n:N}, \mathbf{P}_{n|n:N})$ are computed while keeping $\tilde{\beta}_n$ fixed and vice versa. This algorithm, which we shall call *Robust Backward VB Kalman Filter* (RB-VBKF), is summarized in the following proposition.

Proposition 1 (RB-VBKF algorithm) *Assume that we are given (9)-(11). Then*

Time-update step. Approximate the mean of $p(\mathbf{x}_n|\mathbf{y}_{n+1:N})$ by $\hat{\mathbf{x}}_{n|n+1:N}$ given in (16) and its covariance matrix by $\mathbf{P}_{n|n+1:N}$ given in (17).

Measurement-update step. We proceed with iterations.
▷ *Initialization. Let $\hat{w}^{(0)}_{n|n:N} = \frac{\alpha_n}{\beta_n}$.*
▷ *For $i = 0, 1, \cdots, I \gg 1$,*

- *compute $\hat{\mathbf{x}}^{(i)}_{n|n:N}$ and $\mathbf{P}^{(i)}_{n|n:N}$ by (21) and (20) respectively once $\hat{w}_{n|n:N}$ is replaced by $\hat{w}^{(i)}_{n|n:N}$;*
- *compute $\tilde{\beta}^{(i+1)}_n$ (then $\hat{w}^{(i+1)}_{n|n:N} = \frac{\tilde{\alpha}_n}{\tilde{\beta}^{(i+1)}_n}$) by (23) with the use of $\hat{\mathbf{x}}^{(i)}_{n|n:N}$ and $\mathbf{P}^{(i)}_{n|n:N}$.*

Then, the mean and the covariance matrix of $p(\mathbf{x}_n|\mathbf{y}_{n:N})$ are approximated by $\hat{\mathbf{x}}^{(I)}_{n|n:N}$ and $\mathbf{P}^{(I)}_{n|n:N}$ respectively ; the mean of $p(w_n|\mathbf{y}_{n:N})$ is approximated by $\hat{w}^{(I)}_{n|n:N}$.

4. A ROBUST BACKWARD-FORWARD VB KALMAN FIXED-INTERVAL SMOOTHER

Based on the same idea as above, we use VB approach to propagate an approximation of $p(\mathbf{z}_n|\mathbf{y}_{0:N})$ in the forward direction. Let $p(\mathbf{z}_{0:N}|\mathbf{y}_{0:N}) \approx q(\mathbf{x}_{0:N}|\mathbf{y}_{0:N})q(w_{0:N}|\mathbf{y}_{0:N})$. Then, $q(\mathbf{x}_{0:N}|\mathbf{y}_{0:N})$ and $q(w_{0:N}|\mathbf{y}_{0:N})$ can be expressed by (12) and (13) respectively, once $\mathbf{x}_n$ (resp. w_n) is replaced by $\mathbf{x}_{0:N}$ (resp. $w_{0:N}$). These pdf's involve $p(\mathbf{z}_{0:N}, \mathbf{y}_{0:N})$, which, in turn, following (4)-(5), can be factorized as

$$p(\mathbf{z}_{0:N}, \mathbf{y}_{0:N}) = p(\mathbf{x}_N)\prod_{n=0}^{N-1} p(\mathbf{x}_n|\mathbf{x}_{n+1})\prod_{n=0}^{N} p(\mathbf{y}_n|\mathbf{z}_n)p(w_n). \quad (24)$$

Such factorization is, indeed, a key computational tool for the development of recursive algorithms. Now, one can verify that $p(w_{0:N}|\mathbf{y}_{0:N}) = \prod_{n=0}^{N} p(w_n|\mathbf{y}_{0:N})$, and thus for each $n = 0, \cdots, N$, $p(w_n|\mathbf{y}_{0:N}) \sim \Gamma(\frac{\overline{\alpha}_n}{2}, \frac{\overline{\beta}_n}{2})$ with $\overline{\alpha}_n$ and $\overline{\beta}_n$ are given by (22) and (23) respectively, once $(\hat{\mathbf{x}}_{n|n:N}, \mathbf{P}_{n|n:N})$ are replaced by $(\hat{\mathbf{x}}_{n|0:N}, \mathbf{P}_{n|0:N})$. Finally, as above, a fixed point iteration procedure have to be performed for updating $\overline{\beta}_n$ for fixed $(\hat{\mathbf{x}}_{n|0:N}, \mathbf{P}_{n|0:N})$ and vice versa.

Similarly, by injecting (24) in (12), once $\mathbf{x}_n$ (resp. w_n) is replaced by $\mathbf{x}_{0:N}$ (resp. $w_{0:N}$), and using (9)-(11), one can see that the marginal smoothing pdf $q(\mathbf{x}_n|\mathbf{y}_{0:N})$ is Gaussian whose parameters are propagated in the forward direction as :

$$\mathbf{K}_{n|0:N} = \mathbf{P}_{n|n:N}\tilde{\mathbf{F}}_n\mathbf{P}^{-1}_{n-1|n:N} \quad (25)$$

$$\hat{\mathbf{x}}_{n|0:N} = \hat{\mathbf{x}}_{n|n:N} + \mathbf{K}_{n|0:N}[\hat{\mathbf{x}}_{n-1|0:N} - \hat{\mathbf{x}}_{n-1|n:N}] \quad (26)$$

$$\mathbf{P}_{n|0:N} = \mathbf{P}_{n|n:N} - \mathbf{K}_{n|0:N}[\mathbf{P}_{n-1|n:N} - \mathbf{P}_{n-1|0:N}]\mathbf{K}^T_{n|0:N}. \quad (27)$$

Let note that (25)-(27) are of the same form as the forward step of the backward-forward fixed-interval smoother previously introduced in the backward LGHMC framework [17] [16]. Furthermore, (25)-(27) involve $(\hat{\mathbf{x}}_{n|n:N}, \mathbf{P}_{n|n:N})$, such parameters have been already computed in the backward direction by RB-VBKF algorithm (Prop. 1). The proposed algorithm, which we shall call *Robust Backward-Forward VB Kalman Smoother* (RBF-VBKS), is summarized in the following proposition.

Proposition 2 (RBF-VBKS algorithm) *Assume that we are given (9)-(11). Then for each iteration $i = 0, \cdots, I$,* *Backward step. For $n = N, \cdots, 0$, compute $(\hat{\mathbf{x}}^{(i)}_{n|n:N}, \mathbf{P}^{(i)}_{n|n:N})$ and $(\hat{\mathbf{x}}^{(i)}_{n|n+1:N}, \mathbf{P}^{(i)}_{n|n+1:N})$ by using RB-VBKF formulas.*

Forward step. For $n = 0, \cdots, N$, compute $(\hat{\mathbf{x}}^{(i)}_{n|0:N}, \mathbf{P}^{(i)}_{n|0:N})$ by using (25)-(27), then $\overline{\beta}^{(i)}_n$ by (23) once $(\hat{\mathbf{x}}_{n|n:N}, \mathbf{P}_{n|n:N})$ are replaced by $(\hat{\mathbf{x}}^{(i)}_{n|0:N}, \mathbf{P}^{(i)}_{n|0:N})$.

Finally, the mean and the covariance matrix of $p(\mathbf{x}_n|\mathbf{y}_{0:N})$ are approximated by $\hat{\mathbf{x}}^{(I)}_{n|0:N}$ and $\mathbf{P}^{(I)}_{n|0:N}$ respectively; the mean of $p(w_n|\mathbf{y}_{0:N})$ is approximated by $\hat{w}^{(I)}_{n|0:N} = \frac{\alpha_n + n_{\mathbf{y}}}{\overline{\beta}^{(I)}_n}$.

5. NUMERICAL EXAMPLE

Let us finally provide a numerical example. We consider a model with fixed parameters : $\mathbf{F} = \begin{bmatrix} 0.9 & 0.01 \\ 0.17 & 0.5 \end{bmatrix}$, $\mathbf{H} = [0.45 \;\; 0.6]$, $\mathbf{Q} = 0.1 \times \mathbf{I}_2$ and $R = 0.02$. Let generate $N = 200$ observations containing outliers. Outliers are generated in such a way that the weight $w_n = 1$ (then $r_n = \frac{R}{w_n} = 0.02$) with 70% probability, and $w_n = 0.01$ (then $r_n = 2$) with 30% probability. Fig. 1(a) shows that the generated data is effectively infested by outliers. We aim to assess the performance of the proposed algorithms RB-VBKF and RBF-VBKS, as well as the standard Backward KF (BKF) which uses R rather than $\frac{R}{w_n}$. The proposed algorithms use $\alpha_n = \beta_n = 4$, $\forall n$, and perform with $I = 100$ iterations.

For this purpose, we first estimate the unknown states $\mathbf{x}_n$ and weights w_n, $\forall n$; the observed data samples $\mathbf{y}_n$ are then reconstructed based on these estimates. Results are plotted in Fig. 1(b). As expected, BKF is very sensitive to outliers, while RB-VBKF and RBF-VBKS are more robust to outliers. Fig. 2 shows the theoretic mean square error (MSE) of the estimation of $\mathbf{x}_n$, averaged on 10 independent realizations, *i.e.*, the trace of $\mathbf{P}_{n|n:N}$ when filtering is performed or that of $\mathbf{P}_{n|0:N}$ when smoothing is considered. As expected, RBF-VBKS which uses the whole data set outperforms RB-VBKF, which, in turn, outperforms BKF.

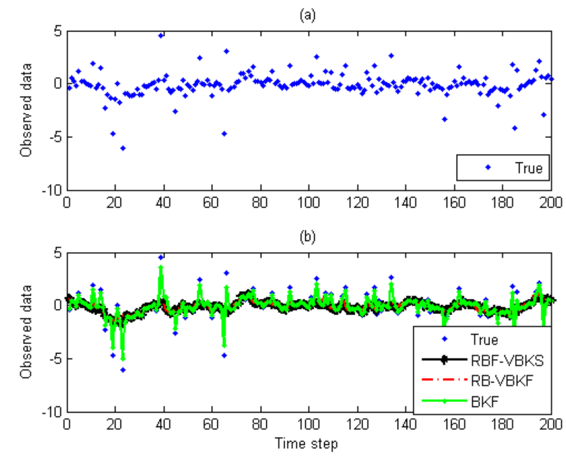

Fig. 1. Observed data and its reconstruction by RBF-VBKS, RB-VBKF and BKF.

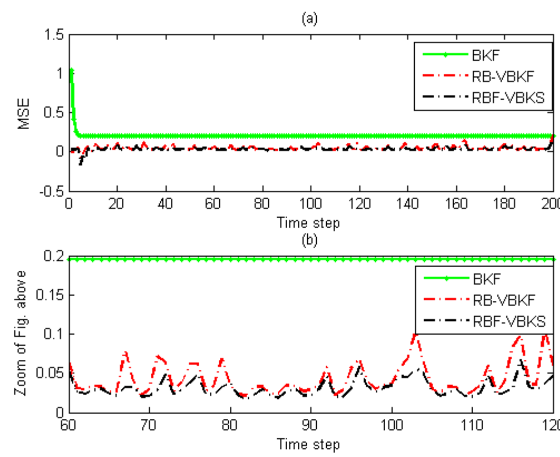

Fig. 2. MSE of the tracking of $\mathbf{x}_n$ by RBF-VBKS, RB-VBKF and BKF. Sub-Fig (b) is a zoom of (a).

6. RELATION TO PREVIOUS WORK

Our work focuses on processing sequential data corrupted with outliers. Recent studies in [9] [10] [14] [15] have been done in linear HMC with heavy-tailed measurement noise framework, and thereby proposed approximate robust Kalman-like forward filtering and forward-backward fixed interval smoothing algorithms based on VB approach. In our work we exploit the backward Markovianity of the model in order to address two new problems, such as, the backward filtering and the backward-forward fixed-interval smoothing. For this purpose, new approximate robust Kalman-like algorithms have been developed. On the other hand, the use of the backward property of dynamical models is not entirely new, and as such, a wide range of KF-like algorithms have been introduced in the LGHMC framework [16] [17]. However, in our best knowledge, the use of backward models in presence of outliers context is original.

7. CONCLUSION

In this paper we have addressed the backward filtering problem and the backward-forward fixed-interval smoothing problem in situations for which the observations are corrupted with outliers. We exploited the backward Markovianity of the model in order to derive robust Kalman-like algorithms based on the variational Bayesian approach. The potential of the proposed algorithms for processing sequential data corrupted with outliers, in particular to outperform the classical Kalman filter, has been shown through simulations.

8. REFERENCES

[1] Y Bar-Shalom, X. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*, New York : Wiley, 2001.

[2] H. Durrant-Whyte and T. Bailey, "Simultaneous Localisation and Mapping (SLAM) : Part I The Essential Algorithms," *IEEE Robot. Autom. Mag.*, vol. 13, pp. 99–110, 2006.

[3] P. M. Djuric, J. Zhang, T. Ghirmai, Y. Huang, and J. H. Kotecha, "Applications of particle filtering to communications : A review," in *Proceedings of the EUropean SIgnal Processing COnference (EUSIPCO)*, Toulouse, France, September 2002.

[4] P. E. Caines, *Linear Stochastic Systems*, Wiley series in Probability and Mathematical Statistics. Wiley, New York, 1988.

[5] B. Ait-El-Fquih and F. Desbouvries, "Kalman filtering in triplet Markov chains," *IEEE Tr. on Signal Processing*, vol. 54, no. 8, pp. 2957–63, Aug. 2006.

[6] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng., Trans. ASME, Series D*, vol. 82, no. 1, pp. 35–45, 1960.

[7] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Prentice Hall, Englewood Cliffs, New Jersey, 1979.

[8] D.S. Moore and G. P. McCabe, *Introduction to the Practice of Statistics*, W.H. Freeman & Company, March 1999.

[9] J. Ting, E. Theodorou, and S. Schaal, "A Kalman filter for robust outlier detection," in *Proceedings of IEEE Int. Conf. Intell. Robots Syst.*, 2007.

[10] J. Ting, E. Theodorou, and S. Schaal, "Learning an outlier-robust Kalman filter," in *Proceedings of 2007 Eur. Conf. Mach. Learn.*, 2007.

[11] R. Peché, S. Särkkä, and J. Hartikainen, "Recursive Outlier-Robust Filtering and Smoothing for Nonlinear Systems using the Multivariate Student-t Distribution," in *Proceedings of 2012 Eur. Conf. Mach. Learn. for Sig. Proc.*, 2012.

[12] T.S. Jaakkola, "Tutorial on variational approximation methods," in *Advanced Mean Field Methods - Theory and Practice*, M. Opper and D. Saad, Eds., pp. 129–59. MIT Press, 2001.

[13] V. Smidl and A. Quinn, *The Variational Bayes Method in Signal Processing*, Springer, 2006.

[14] G. Agamennoni and J. Nieto, "An outlier-robust Kalman filter," in *Proceedings of Int. Conf. Robot. Autom.*, 2011.

[15] G. Agamennoni, J. Nieto, and E. Nebot, "Approximate Inference in State-Space Models With Heavy-Tailed Noise," *IEEE Trans. on Signal Processing*, vol. 60, no. 10, pp. 5024–37, Oct. 2012.

[16] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear estimation*, Prentice Hall Information and System Sciences Series. Prentice Hall, Upper Saddle River, NJ, 2000.

[17] B. Ait-El-Fquih and F. Desbouvries, "On Bayesian Fixed-Interval Smoothing Algorithms," *IEEE Transactions on Automatic Conrol*, vol. 53, no. 10, pp. 2437–42, 2008.