Submitted to the Bernoulli

Multiplicative Kalman filtering.

FABIENNE COMTE[1,*] and VALENTINE GENON-CATALOT[1,**] and MATHIEU KESSLER[2]

[1] *Laboratoire MAP 5, Université Paris Descartes, UFR de Mathématiques et Informatique, CNRS UMR 8145, 45 rue des Saints Pères, 75270 Paris Cedex 06, France. E-mail:* *`fabienne.comte@univ-paris5.fr`; **`Valentine.Genon-Catalot@math-info.univ-paris5.fr`

[2] *Departamento de Matemática Aplicada y Estadística, Universidad Politécnica de Cartagena, Paseo Alfonso XIII, 30203 Cartagena, Spain. E-mail:* `mathieu.kessler@upct.es`

We study a non-linear hidden Markov model, where the process of interest is the absolute value of a discretely observed Ornstein-Uhlenbeck diffusion, which is observed after a multiplicative perturbation. We obtain explicit formulae for the recursive relations which link the relevant conditional distributions. As a consequence the predicted, filtered, and smoothed distributions for the hidden process can easily be computed. We illustrate the behaviour of these distributions on simulations.

Keywords: filtering, discrete time observations, hidden Markov models, parametric inference, scale perturbation.

1. Introduction

Consider a real-valued Markov process $(x(t))$ sampled at times $t_1, t_2, \ldots, t_n, \ldots$ $(0 < t_1 < \ldots < t_n \ldots)$ and suppose that the observation at time t_i is of the form

$$Y_i = F(x(t_i), \varepsilon_i), \tag{1}$$

where F is a real-valued known function, the ε_i's are independent random variables (a noise) and the sequence (ε_i) is independent of the process $(x(t))$. This kind of observation process is encountered in many fields of applications such as finance, biology or engineering. The model is known under different names: hidden Markov model, non linear filtering model, state-space model or dynamic model.

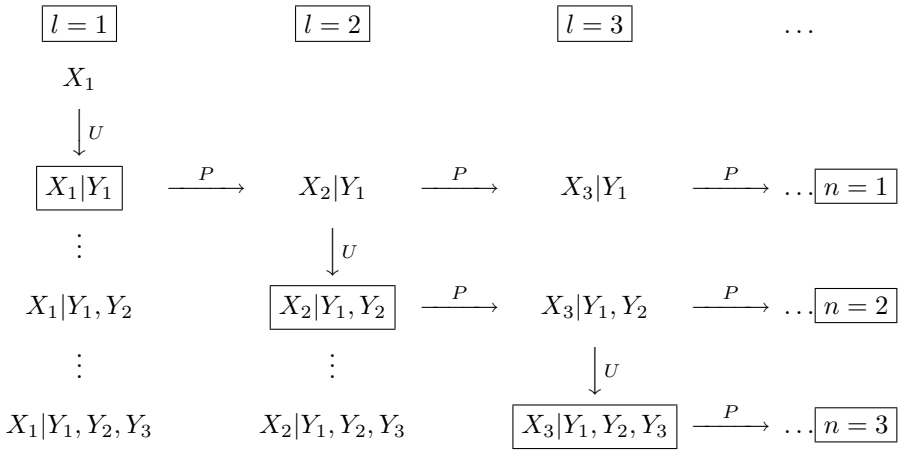
Denote, for $i \in \mathbb{N}$, $X_i = x(t_i)$. For a given l, all the information about the hidden variable X_l is to be drawn from the conditional distribution $\mathcal{L}(X_l|Y_n \ldots, Y_1)$, given that, at time t_n, we have observed $Y_1, \ldots, Y_n$:

- In the case when $l > n$, we use $\mathcal{L}(X_l|Y_n \ldots, Y_1)$ to *predict* the value of the variable X_l.
- In the case when $l = n$, we use $\mathcal{L}(X_l|Y_l \ldots, Y_1)$ to *filter* the value of the variable X_l.
- While, in the case when $l < n$, we use $\mathcal{L}(X_l|Y_n \ldots, Y_1)$ to *smooth* the value of the variable X_l.

These three kind of distributions are related through the following classical operators. Denote by $\nu_{l|n:1}$ the conditional distribution $\mathcal{L}(X_l|Y_n \ldots, Y_1)$,

- *Updating:* U operator, $\nu_{l|l:1}(dx_l) \propto p(y_l|x_l)\nu_{l|l-1:1}(dx_l)$.
- *Prediction:* $\nu_{l+r|l:1} = \nu_{l|l:1}P_{l+r,l}$, where $P_{l+r,l}$ denotes the transition operator from X_l to X_{l+r}. In the case when the chain is assumed to be homogeneous with transition operator P, $P_{l+r,l} = P^r$.
- *Smoothing:*
$$\nu_{l|n:1}(dx_l) \propto p(y_{l+1}, \ldots, y_n|x_l)\nu_{l|l:1}(dx_l).$$

As a consequence, any of the relevant conditional distributions $\nu_{l|n:1}$ is computed stepwise by moving through the following scheme:



A main difficulty associated to the study of this kind of model arises from the fact that these three operators involve high dimensional integrals and admit therefore no closed form expression. There is though one very important exception: the Kalman filtering, where the hidden process $x(t)$ is an Ornstein-Uhlenbeck (O-U) process

$$dx(t) = -\theta x(t)dt + \sigma dW_t,$$

and the noise observation process specification

$$Y_i = x(t_i) + \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{N}(0, \beta^2),$$

implies that the class of Gaussian distributions is closed under the action of the three operators. Therefore, if $\mathcal{L}(X_1)$ is Gaussian, $\nu_{l|n:1}$ are Gaussian and are specified by their means and variances which can be computed recursively.

In this work, we study another example of Hidden Markov model, where the three operators admit a closed form expression. This model consists in a multiplicative perturbation of the absolute value of a discretely observed O-U process. It builds upon a

model previously studied in Genon-Catalot (2003), Genon-Catalot and Kessler (2004) and Chaleyat-Maurel and Genon-Catalot (2006), but presents two significant contributions with respect to these papers: on one hand, the class of noises is extended to a more flexible class which allows in particular to specify the number of finite moments that admits the marginal distribution of Y_i; and on the other hand, we prove that the action of the smoothing operator stays in the same class of distributions as for the prediction and updating operators, and we provide a backward recursion for the computation of the smoothed distribution $\nu_{l|n:1}$.

As a consequence, this work presents a non-linear hidden Markov model, where the hidden chain state space is $\mathbb{R}^+$, for which the explicit computation of any of the conditional distributions $\mathcal{L}(X_l|Y_n \dots, Y_1)$ is possible and computationally easy. It can in particular be useful for testing numerical methods designed for general non-linear hidden Markov models, by allowing to compare their performance to the exact methods.

After introducing the model in Section 2, we describe in detail the three operators in Section 3. Section 4 contains an illustration of the behaviour of the relevant conditional distributions. Concluding remarks are stated in Section 5. Finally, the proofs are collected in an appendix.

2. The model

2.1. The hidden process

The hidden process $(X_n)_{n\geq 1}$ consists in the discrete observation of the absolute value of a continuous time Ornstein-Uhlenbeck process. Concretely,

$$\text{for } n \geq 1, \quad X_n = |\xi_{n\Delta}|, \tag{2}$$

where $\Delta > 0$ is the discretization step and ξ solves

$$d\xi(t) = -\theta\xi(t)dt + \sigma dW_t, \quad \xi(0) = \eta, \tag{3}$$

where $\theta \in \mathbb{R}$, $\sigma > 0$, (W_t) is a standard Wiener process, η a random variable independent of (W_t). Notice that the observation times need not be equidistant, *i.e.* Δ could be Δ_n, though, for the sake of clarity in the exposition, we will stick to the homogeneous case.

It is well known that the transition density $x' \mapsto p_t^{\xi}(x, x'; \theta, \sigma)$ of the process ξ is Gaussian with mean $a(t)x$ and variance $\beta^2(t)$, where

$$a(t) = \exp(-\theta\, t), \qquad \beta(t) = \sigma \left(\frac{1 - \exp\left(-2\theta t\right)}{2\theta} \right)^{1/2}. \tag{4}$$

In the sequel we will use the shorthand notation $a = a(\Delta)$ and $\beta = \beta(\Delta)$.

The process $(X_n)_{n\geq 1}$ is therefore obtained as the absolute value of an $AR(1)$ process with parameters a and β^2.

It is easy to deduce that the hidden process $(X_n)_{n\geq 1}$ is Markov with transition density given by

$$p_\Delta(x,x') = 1_{(x'>0)}\frac{2}{\beta\sqrt{2\pi}}\exp\left(-\frac{x'^2}{2\beta^2}-\frac{a^2x^2}{2\beta^2}\right)\left(\cosh\left(\frac{axx'}{\beta^2}\right)\right), \tag{5}$$

($x \geq 0$).

Using the series expansion of the cosh, this transition density (5) can be written as a mixture of distributions which will turn out to be relevant for the description of the three conditional operators described in the introduction:

$$p_\Delta(x,x') = \sum_{i\geq 0} w_i(a^2x^2/\beta^2)g_{i,\beta}(x'), \tag{6}$$

where, for $i \geq 0$, the weights w_i are given by

$$u \in \mathbb{R}, \quad w_i(u) = \exp(-u/2)\ (u/2)^i/i! \tag{7}$$

and the function

$$g_{i,\sigma}(x) = 1_{(x>0)}\frac{2}{\sigma\sqrt{2\pi}}\exp\left(-\frac{x^2}{2\sigma^2}\right)\frac{x^{2i}}{C_{2i}\sigma^{2i}}, \tag{8}$$

is the density of the square root of a Gamma distribution with scale parameter σ and location parameter $(i+1/2)$. The normalising constant C_{2i}, is simply the moment of order $2i$ of a Gaussian standard variable, i.e. for $i \in \mathbb{N}$,

$$C_{2i} = \frac{(2i)!}{2^i i!} = (2i-1)(2i-3)\ldots 1.$$

In a slight abuse of notation, we will use the notation $g_{i,\sigma}$ to denote both the density and its associated distribution.

The transition operator $P = P_\Delta$ of (X_n) acting on functions is given, for any function f for which the integral makes sense, by:

$$Pf(x) = \int p_\Delta(x,x')f(x')dx'. \tag{9}$$

Acting on measures, it is given by:

$$\nu P(dx') = \left(\int \nu(dx)p_\Delta(x,x')\right)dx'. \tag{10}$$

We denote by gP the density of νP when $\nu(dx) = g(x)dx$, *i.e.* the function

$$gP(x') = \int g(x)p_\Delta(x,x')dx. \tag{11}$$

Note that the r-th iterate of P is simply that $P^r = P_{r\Delta}$ and corresponds to the parameters $a(r\Delta), \beta(r\Delta)$.

If, for some $l \in \mathbb{N}$, g is the density of the law of X_l, $x' \mapsto gP(x')$ is the density of the law of X_{l+1}. In particular, if, given $Y_1, \ldots, Y_l$, $\nu_{l|l:1}$ is the law of X_l, the distribution $\nu_{l+1|l:1} = \mathcal{L}(X_{l+1}|Y_1, \ldots, Y_1)$ is obtained as $\nu_{l|l:1}P$.

A crucial feature for the understanding of the prediction step can be easily deduced from the series expansion (6):

Property 1. *Let $l \in \mathbb{N}$,*

$$g_{l,\sigma}P(x') = \sum_{i\geq 0} \left(\int w_i(a^2x^2/\beta^2) g_{l,\sigma}(x)dx \right) g_{i,\sigma}(x'),$$

i.e the action of P on a density $g_{l,\sigma}$ of the form (8) yields a mixture of $g_{j,\sigma}$ distributions. This mixture has infinite length, though we will see in proposition 3.1 that, by considering a different scale parameter, it admits an equivalent mixture representation of finite length.

An additional feature which strengthen the role of the distributions $g_{i,\sigma}$ is described in the following property:

Property 2. *If $\theta > 0$ or equivalently $a < 1$, the unique invariant distribution of the chain $(X_n)_{n\geq 1}$ admits the density g_{0,σ_s}, see (8), with*

$$\sigma_s^2 = \frac{\sigma^2}{2\theta} = \frac{\beta^2}{1-a^2}. \tag{12}$$

These two properties make it natural to consider using the distributions $g_{i,\sigma},\ i \geq 0,\ \sigma > 0$ as building components of mixtures representations of distributions.

2.2. The noise

We focus on obtaining exact computations as for the standard Kalman filter. The perturbation is assumed to act in an multiplicative way:

$$Y_n = \psi_n\, X_n, \tag{13}$$

where $(\psi_n)_{n\geq 1}$ is a sequence of *i.i.d* random variables, independent from the process ξ.

In Genon-Catalot and Kessler (2004) and Chaleyat-Maurel and Genon-Catalot (2006), it is assumed that the r.v.'s ψ_n have distribution $\Gamma^{-1/2}$ where Γ has exponential distribution with parameter λ. This leads to explicit computations. Here, we consider a more flexible class of distributions for the noise (ψ_n) that includes the previous ones.

We assume that ψ_1 has the law $(\Gamma(k,\lambda))^{-1/2}$, for some $k \in \mathbb{N}^*$ and $\lambda > 0$, *i.e.* admits the density

$$f(\psi) = 1_{\psi>0} \frac{2\lambda^k}{\Gamma(k)\psi^{2k+1}} \exp\left(-\frac{\lambda}{\psi^2}\right). \tag{14}$$

Let us check now how, given this specification of the noise structure, the updating operator behaves.

To do so, we need to consider the conditional distribution $\mathcal{L}(Y_i|X_i = x)$, which, from (14), can be seen to admit the density,

$$p_x(y) = 1_{y>0}\ \frac{2\lambda^k x^{2k}}{\Gamma(k)y^{2k+1}} \exp\,(-\frac{\lambda x^2}{y^2}), \quad \text{for} \quad x > 0. \tag{15}$$

Assume that X_l has density g, by the Bayes rule, the conditional distribution $\mathcal{L}(X_l|Y_l = y)$ admits a density proportional to $p_x(y)g(x)$. In particular, if g is of the form $g_{i,\sigma}$, see (8), we deduce from (15), that the density of $\mathcal{L}(X_l|Y_l = y)$ is proportional to

$$\propto x^{2k+2i} \exp\,[-(\frac{1}{\sigma^2} + \frac{2\lambda}{y^2})x^2/2]dx,$$

which is $g_{i+k,T_y(\sigma)}$, once we have set

$$\frac{1}{T_y^2(\sigma)} = \frac{1}{\sigma^2} + \frac{2\lambda}{y^2}.$$

To sum up we have checked the following joint property of the noise specification and of the distributions $g_{i,\sigma}$:

Property 3. *If $X_l \sim g_{i,\sigma}$, $\mathcal{L}(X_l|Y_l = y_l) = g_{i+k,T_{y_l}(\sigma)}$.*

Once again, the family $g_{i,\sigma}$ of distributions appears to play a crucial role in the description of the fundamental operators associated to the filtering process.

We finish this subsection by a last property involving the noise specification and the distributions $g_{i,\sigma}$.

Property 4. *If $X_l \sim g_{i,\sigma}$, the marginal density of Y_l is*

$$p_{g_{i,\sigma}}(y) = 2\ 1_{(y>0)} \frac{\lambda^k \sigma^{2k}}{\Gamma(k)} \frac{C_{2(i+k)}}{C_{2i}} \frac{y^{2i}}{(y^2 + 2\lambda\sigma^2)^{i+k+1/2}}. \tag{16}$$

For $\sigma = 0$, the marginal distribution of Y is δ_0.

This property is easily proved combining the expressions (8) and (15). Notice in particular that, in the case when X_n is in its stationary state, (16) is the marginal density of Y if we set $i = 0$ and $\sigma = \sigma_s$ defined in (12).

2.3. The class $\mathcal{F}$

From the basic properties of the prediction and updating operators described in the two previous subsections, it is natural to introduce the following class of distributions, built as mixtures of $g_{i,\sigma}$ components.

$$\mathcal{F} = \{\nu_{\alpha,\sigma} = \sum_{i\geq 0} \alpha_i g_{i,\sigma},\ \sigma > 0, \alpha = (\alpha_i, i \geq 0) \in \mathcal{S}\}, \tag{17}$$

where $\mathcal{S}$ is the set of finite-length mixture parameters:

$$\mathcal{S} = \{\alpha = (\alpha_i, i \geq 0) : \forall i \geq 0, \alpha_i \geq 0, \sum_{i=0}^{\infty} \alpha_i = 1 \text{ and } \exists L \in \mathbb{N}, \forall i > L, \alpha_i = 0\}. \quad (18)$$

(If $\sigma = 0$, we set $\nu_{\alpha,0} = \delta_0$ for all α).

Given α in $\mathcal{S}$, we will also use the notation α to denote the mapping defined on $\mathbb{N}$ by, for $i \geq 0$, $\alpha(i) = \alpha_i$.

Consider a mixture parameter $\alpha \in \mathcal{S}$, we denote by $l(\alpha) = \sup\{i; \alpha_i > 0\} < +\infty$ the length of the mixture parameter α. Note that the class $\mathcal{F}$ involves mixtures of $g_{i,\sigma}$ sharing the same scale parameter σ.

2.4. Summing up

To sum up, our model consists in the observation of

$$Y_n = \psi_n \, X_n,$$

where

- $X_n = |\xi_{n\Delta}|$, where ξ is a OU process with drit parameter $-\theta$ and diffusion parameter σ, see (3).
- $(\psi_n)_{n\geq 1}$ is a sequence of i.i.d r.v, independent from the process ξ, and $\psi_1 \sim (\Gamma(k, \lambda))^{-1/2}$, see (14).
- We take $\mathcal{L}(X_1)$ to belong to $\mathcal{F}$, see (17).

2.5. Remarks

1. In hidden Markov models, the iterations of the prediction and updating steps along the scheme described in the introduction become rapidly intractable unless both the updating and prediction operators evolve within a parametric family of distributions. In this case, the model is called a finite-dimensional filter system since each conditional distribution is specified by a finite number of statistics (the stochastic process of parameters at stage n). This is the case of the model associated to the standard Kalman filter. However, this notion is very restrictive (see *e.g.* Sawitzki, 1981; Runggaldier and Spizzichino, 2001): very few finite-dimensional filter system are available and they are often obtained as the result of an *ad hoc* construction. In Chaleyat-Maurel and Genon-Catalot (2006), the notion of computable infinite dimensional filters is introduced. This notion has links with the one proposed in Di Masi et al. (1983). The conditional distributions are allowed to evolve in much larger classes built using mixtures. These distributions are specified by a number of parameters which is finite but may vary along the iterations. The model described in subsection 2.4 is a computable infinite dimensional filter with respect to the class of distributions $\mathcal{F}$.

2. Our model can be transformed into an additive perturbation framework by a logarithmic transformation:

$$\log(Y_n) = \log(X_n) + \log(\psi_n) \tag{19}$$

In section 3, we obtain a closed form description of the relevant conditional distributions for the untransformed multiplicative model. It is straightforward to obtain the corresponding conditional distributions for model (19).

It is worth noticing that model (19) is an example of additive perturbation model with a non Gaussian hidden process, non Gaussian noise, but where the explicit computation of the filtering, prediction, and smoothing distributions is possible.

3. An explicit formula for the moments of a distribution $\nu_{\alpha,\sigma}$ in $\mathcal{F}$ can readily be obtained:

$$\text{For any } r > 0, \quad \int x^r \nu_{\alpha,\sigma}(dx) = 2^{r/2}\sigma^r \sum_{i\geq 0} \alpha_i \frac{\Gamma(i+1/2+r/2)}{\Gamma(i+1/2)}. \tag{20}$$

Notice that an alternative expression in terms of the moments of a standard Gaussian variable can be obtained if we take into account that

$$\Gamma(b) = \frac{\sqrt{2\pi}}{2^b} C_{2b-1}, \ b > 0,$$

where, for $\alpha > -1/2$, $C_\alpha = \mathbb{E}[|Z|^\alpha]$, $Z \sim \mathcal{N}(0,1)$.

3. Description of the updating, prediction and smoothing operators

The key joint properties of the distributions $g_{i,\sigma}$, the noise ψ_n and and transition operator of the chain X_n emphasised in the previous sections allow to obtain closed form recursive expressions for the relevant filtering distributions $\nu_{l|n:1} = \mathcal{L}(X_l|Y_1,\ldots,Y_n)$. By convention, we set $\nu_{1|0:1} = \mathcal{L}(X_1)$.

Proposition 3.1. *Assume that $\nu_{1|0:1}$ belongs to the class $\mathcal{F}$. Then, for all $n, l \geq 1$, $\nu_{l|n:1}$ belongs to $\mathcal{F}$. The corresponding scale and mixture parameters, $\sigma_{l|n:1}$ and $\alpha_{l|n:1} = (\alpha_{l|n:1}(i), 0 \leq i \leq l(\alpha_{l|n:1}))$ can be computed recursively as follows*

I. Updating step *For $l \geq 1$,*

$$\frac{1}{\sigma^2_{l|l:1}} = \frac{1}{\sigma^2_{l|l-1:1}} + \frac{2\lambda}{y_l^2}. \tag{21}$$

and, for $j \geq 0$,

$$\alpha_{l|l:1}(j) \propto 1_{(j\geq k)}(2j-1)(2j-3)\ldots(2(j-k)+1)\left(\frac{\sigma_{l|l:1}}{\sigma_{l|l-1:1}}\right)^{2(j-k)} \alpha_{l|l-1:1}(j-k). \tag{22}$$

On the other hand, $l(\alpha_{l|l:1}) = l(\alpha_{l|l-1:1}) + k$. *If* $y_l = 0$, *then,* $\sigma_{l|l:1} = 0$, $\alpha_{l|l:1}(k) = 1$ *and* $\nu_{l|l:1} = \delta_0$. *Hence, formulae (21) and (22) still hold in this degenerate case.*

II. Prediction, $\nu_{l+r|l:1} = \mathcal{L}(X_{l+r}|Y_1, \ldots, Y_l)$. *For* $r \geq 1$, *the predictive conditional distribution* $\nu_{l+r|l:1}$ *belongs to the class* $\mathcal{F}$. *The corresponding scale and mixture parameters are given by*

$$\sigma^2_{l+r|l:1} = \beta^2(r\Delta) + a^2(r\Delta)\sigma^2_{l|l:1}, \tag{23}$$

where $a^2(t)$ *and* $\beta^2(t)$ *are given in (4). On the other hand, for* $j \geq 0$,

$$\alpha_{l+r|l:1}(j) = \sum_{i \geq j} \alpha_{l|l:1}(i)\kappa^{(i)}_j, \tag{24}$$

with, for $i \geq j$,

$$\kappa^{(i)}_j = \binom{i}{j} \left(1 - \frac{\beta^2(r\Delta)}{\sigma^2_{l+r|l:1}}\right)^j \left(\frac{\beta^2(r\Delta)}{\sigma^2_{l+r|l:1}}\right)^{i-j}. \tag{25}$$

III. Smoothing.

1. *For* $l < n$, *we have*

$$\nu_{l|n:1}(dx) = \frac{\sum_{j=0}^{(n-l)k} d^{l,n}_j \; h_{j,\Phi_{l,n}}(x)}{\prod_{i=l+1}^n p(y_i|y_{i-1}, \ldots, y_1)} \; \nu_{l|l:1}(dx), \tag{26}$$

where $h_{i,\sigma}(x) = x^{2i} e^{-x^2/(2\sigma^2)}$,

$$\Phi_{l,n} = \Phi_{l,n}(y_{l+1}, \ldots, y_n), \tag{27}$$

and

$$d^{l,n}_j = d^{l,n}_j(y_{l+1}, \ldots, y_n), \tag{28}$$

depend on the observations $(y_{l+1}, \ldots, y_n)$.
Recall that $a = a(\Delta)$ *and* $\beta = \beta(\Delta)$, *see (4). Introduce*

$$\Phi^2(\sigma) = \frac{\beta^2 + \sigma^2}{a^2}, \tag{29}$$

and, for $0 \leq l \leq i$,

$$c^{i,\sigma}_l = \frac{C_{2i}}{C_{2l}} \binom{i}{l} \left(\frac{\sigma^2}{\beta^2 + \sigma^2}\right)^{i+l+(1/2)} a^{2l} \; \beta^{2(i-l)}. \tag{30}$$

The coefficients $\Phi_{l,n}$ *and* $d^{l,n}_\cdot$ *can be computed recursively as follows. On one hand,*

$$\Phi_{n-1,n} = \Phi(\frac{y_n}{\sqrt{2\lambda}}) \qquad d^{n-1,n}_j = \frac{2\lambda^k}{\Gamma(k)\; y_n^{2k+1}} c_j^{k,\frac{y_n}{\sqrt{2\lambda}}}. \tag{31}$$

On the other hand, we have a downward recursion. For $l+1<n$,

$$\Phi_{l,n}=\Phi(\gamma_{l,n}) \quad with \quad \frac{1}{\gamma_{l,n}^2}=\frac{1}{\Phi_{l+1,n}^2}+\frac{2\lambda}{y_{l+1}^2}. \tag{32}$$

For $s=0,\dots,(n-l)k$,

$$d_s^{l,n}=\frac{2\lambda^k}{\Gamma(k)\ y_{l+1}^{2k+1}}\sum_{j=(s-k)^+}^{(n-l-1)k} c_s^{j+k,\gamma_{l,n}} d_j^{l+1,n}. \tag{33}$$

2. *Moreover* $\nu_{l|n:1}$ *is of the class* $\mathcal{F}$ *and we have*

$$\nu_{l|n:1}(dx)=\sum_{s=0}^{nk}\alpha_{l|n:1}(s)g_{s,\sigma_{l|n:1}}(x)dx, \tag{34}$$

where

$$\frac{1}{\sigma_{l|n:1}^2}=\frac{1}{\sigma_{l|l:1}^2}+\frac{1}{\Phi_{l,n}^2}, \tag{35}$$

and

$$\begin{aligned}\alpha_{l|n:1}(s)=&\frac{\sigma_{l|n:1}^{2s+1}}{\prod_{i=l+1}^n p(y_i|y_{i-1},\dots,y_1)}\times\\ &\sum_{u\in\mathcal{U}_s}(2s-1)(2s-3)\dots(2u+1)\frac{1}{\sigma_{l|l:1}^{2u+1}}d_{s-u}^{l,n}\alpha_{l|l:1}(u),\end{aligned} \tag{36}$$

where the sum is taken over the set

$$\mathcal{U}_s=\{u:[s-(n-l)k]^+\le u\le\min(s,lk)\}.$$

Proof in appendix.

Remarks

- The denominator in (26) can be computed explicitly. Indeed, we have

$$p(y_l|y_{l-1},\dots,y_1)=\int p_{x_l}(y_l)\nu_{l|l-1:1}(dx_l).$$

Since $\nu_{l|l-1:1}=\sum_i\alpha_{l|l-1:1}(i)g_{i,\sigma_{l|l-1:1}}$, we use formula (16) in Property 4 to deduce that

$$\begin{aligned}p(y_l|y_{l-1},\dots,y_1)=1_{(y_l>0)}\sum_{i\ge0}&\alpha_{l|l-1:1}(i)2\ \frac{\lambda^k{\sigma_{l|l-1:1}}^{2k}}{\Gamma(k)}\\ &\times\frac{C_{2(i+k)}}{C_{2i}}\frac{{y_l}^{2i}}{({y_l}^2+2\lambda{\sigma_{l|l-1:1}}^2)^{i+k+1/2}}.\end{aligned}$$

- The length $l(\alpha)$ of α in $\mathcal{S}$ was introduced in subsection 2.3.
  We deduce from (22) and (24) that $l(\alpha_{l|l:1}) = l(\alpha_{l|l-1:1}) + k$ while $l(\alpha_{l+r|l:1}) = l(\alpha_{l|l:1})$. Starting from $\mathcal{L}(X_1) = \nu_{\alpha_1,\sigma_1}$ with $l(\alpha_1) = l_1$, we have that $l(\alpha_{l|l:1})$ is $l_1 + kl$. The length of the mixture parameter increases therefore linearly, however, as we will see on simulations, the number of significant coefficients remains very small, see Table 1. This is due to the fact, that, as soon as the observed y_l is close to zero, $\sigma_{l|l:1}$ is close to zero (see (21)), which implies that $\sigma^2_{l+1|l:1}$ is close to β^2, see (23). As a consequence, we deduce from (25), that $\alpha_{l+1|l:1}$ is close to $(1, 0)$, which resets the length to 1. In the case when the chain (X_n) is ergodic, we deduce from the density of the marginal distribution of Y_n in (16) that it is pretty likely that Y_n takes values close to zero.
- We have chosen to restrict the elements of $\mathcal{F}$ to have finite length mixture coefficients. However the computations of the prediction, updating and smoothing steps do not depend on the length of the mixture coefficients and are therefore valid also if $\mathcal{L}(X_1)$ is $\nu_{\alpha,\sigma}$ with $l(\alpha) = +\infty$. This could be the case for example, if X_1 is assumed to be generated from a deterministic $X_0 = x_0$. Indeed, its law is in this case $\nu^{x_0}(dx) = p_\Delta(x_0, x)dx$, which as seen from (6), admits a infinite length representation as a mixture of distributions $g_{i,\beta}$.
  It is however worth noticing that, if ergodicity is assumed, the distribution of X_1 loses its importance asymptotically.
- The derivation of the updating, prediction and smoothing steps do not rely on the assumption that $a < 1$ (ergodic case for X_n).
- Equation (21) can be interpreted in terms of gain of observing y_l: the inverse of $\sigma^2_{l|l:1}$, the square of the scale parameter after observing y_l, is larger than the inverse of $\sigma^2_{l|l-1:1}$, the square of the scale parameter before observing y_l. The increment is $2\lambda/y_l^2$. A similar interpretation holds for formula (35).

4. Numerical study

4.1. Typical trajectories for the hidden and the observed processes

Recall that the observed process Y_n is $\psi_n X_n$, and that the noise ψ admits the density (14). The noise level can therefore be calibrated through the choice of the pair (λ, k).

First notice that $\mathbb{E}(\psi_1^\alpha) < \infty$ for $\alpha < 2k$ and that $\mathbb{E}(\psi_1^\alpha) = \lambda^{\alpha/2} \frac{\Gamma(k-(\alpha/2))}{\Gamma(k)}$. In particular, if $k = 1$, ψ_1 has no second order moment. As a consequence, the same holds for Y_i. Therefore, our model may be used to model heavy-tailed data. To calibrate the noise, we may choose the pair (λ, k) such that $\mathbb{E}\psi_1 = 1$, *i.e.*

$$\lambda = \lambda_1(k) = \left(\frac{\Gamma(k)}{\Gamma(k - 1/2)} \right)^2. \tag{37}$$

Using the Stirling formula, we get that $\lambda_1(k) \sim k - 3/2$ as k tends to infinity. Since, ψ_1 has distribution $(2\lambda_1(k))^{1/2}/\chi(2k)$, we obtain that $\psi_1 \to 1$ in probability as k tends to infinity.

Another way of calibrating the noise is to choose

$$\lambda = \lambda_2(k) = k + 1/2. \tag{38}$$

In this case, ψ_1 has distribution $\sqrt{2k+1}/\chi(2k)$ which also converges to 1. Moreover, the density $f(\psi)$ of ψ_1 has a unique mode attained at $\hat{\psi} = 1$.

In both cases, by increasing the value of k, we can reduce the influence of the noise.

In Figure 1 below, trajectories of the observed process and the hidden process are plotted for two different noise levels, $k = 2$ and $k = 20$. In both cases the parameter λ in the noise density was computed using formula (37). The hidden process was simulated with $\theta = 0.5$, $\sigma = 0.2$, see (3). The time step Δ was chosen to be 0.5. The distribution of the initial variable X_1 was chosen to be the stationary density g_{0,σ_s}, see (12).

On the other hand, the marginal densities of Y_n and X_n in the stationary state are jointly represented in Figure 2, for the same set of parameters, but for $k = 1$ in the noise specification.

4.2. Typical evolution of the scale and mixture parameters for the filtering and prediction distributions

Table 1 illustrates a typical run of the filters described in Proposition 3.1, corresponding to one simulated trajectory of the observed process Y. The time step was chosen to be $\Delta = 0.5$, while the parameters values were set to $\theta = 0.5$, $\sigma = 0.2$ for the hidden process X, see (2). In particular, the hidden chain is ergodic and X_1 was simulated from the invariant distribution. The noise specification was $k = 2$ in (14) and λ was set to satisfy (37).

From this table, one checks that, as predicted from the relations described in Proposition 3.1, the order of magnitude of the observation y_l directly influences the scale and mixture parameters of the filtered distributions: when y_l is close to zero, $\sigma_{l|l:1}$ is close to zero, see (21), and as a consequence, $\sigma_{l+1|l:1}$ is close to β, see (23) (in our case, $\beta = 0.1254$). As for the mixture parameters, we deduce from (24) and (25) that, when $\sigma_{l+1|l:1}$ is close to β, $\alpha_{l+1|l:1}(j) \simeq 0$ for $j \geq 1$. To sum up, when y_l is close to zero, $\nu_{l+1|l:1}$ is reinitialised to have mixture parameter $\alpha_{l+1|l:1} \simeq (1, 0)$ and scale parameter $\sigma_{l+1|l:1} \simeq \beta$, which corresponds to the distribution of the absolute value of a Gaussian variable $\mathcal{N}(0, \beta^2)$.

In particular, it explains why the length of the mixture parameter remains small in practise, while it theoretically increases linearly with l. In the implementation of the filtering algorithm, given a mixture parameter $\alpha = (\alpha_i)_{0 \leq i \leq l(\alpha)}$, we considered a threshold on its components: we only worked with its first L components, being L the

Observation												
y_1 :	0.007	$\sigma_{1\|1}$	0.005	$\alpha_{1\|1}$	0	0	1					
		$\sigma_{2\|1}$	0.126	$\alpha_{2\|1}$	0.998	0.002	0					
y_2 :	0.059	$\sigma_{2\|2:1}$	0.035	$\alpha_{2\|2:1}$	0	0	0.999	0.001	0			
		$\sigma_{3\|2:1}$	0.128	$\alpha_{3\|2:1}$	0.91	0.088	0.002	0	0			
y_3 :	0.028	$\sigma_{3\|3:1}$	0.018	$\alpha_{3\|3:1}$	0	0	0.991	0.009	0	0	0	
		$\sigma_{4\|3:1}$	0.126	$\alpha_{4\|3:1}$	0.976	0.024	0	0	0	0	0	
y_4 :	0.236	$\sigma_{4\|4:1}$	0.096	$\alpha_{4\|4:1}$	0	0	0.934	0.065	0.001	0	0	0
		$\sigma_{5\|4:1}$	0.146	$\alpha_{5\|4:1}$	0.535	0.39	0.074	0.001	0	0	0	0
y_5 :	0.109	$\sigma_{5\|5:1}$	0.062	$\alpha_{5\|5:1}$	0	0	0.585	0.384	0.031	0	0	0
		$\sigma_{6\|5:1}$	0.134	$\alpha_{6\|5:1}$	0.715	0.255	0.029	0.001	0	0	0	0
y_6 :	0.148	$\sigma_{6\|6:1}$	0.076	$\alpha_{6\|6:1}$	0	0	0.615	0.354	0.03	0.001	0	0
		$\sigma_{7\|6:1}$	0.139	$\alpha_{7\|6:1}$	0.616	0.327	0.054	0.003	0	0	0	0
y_7 :	0.123	$\sigma_{7\|7:1}$	0.067	$\alpha_{7\|7:1}$	0	0	0.594	0.371	0.034	0.001	0	0
		$\sigma_{8\|7:1}$	0.136	$\alpha_{8\|7:1}$	0.677	0.284	0.038	0.002	0	0	0	0
y_8 :	0.032	$\sigma_{8\|8:1}$	0.020	$\alpha_{8\|8:1}$	0	0	0.957	0.042	0	0	0	0
		$\sigma_{9\|8:1}$	0.126	$\alpha_{9\|8:1}$	0.97	0.03	0	0	0	0	0	0
y_9:	0.186	$\sigma_{9\|9:1}$	0.086	$\alpha_{9\|9:1}$	0	0	0.933	0.066	0.001	0	0	0
		$\sigma_{10\|9:1}$	0.142	$\alpha_{10\|9:1}$	0.598	0.348	0.053	0.001	0	0	0	0
y_{10}:	0.024	$\sigma_{10\|10:1}$	0.015	$\alpha_{10\|10:1}$	0	0	0.968	0.032	0	0	0	0

Table 1. Typical evolution, as described in Proposition 3.1, for one trajectory of the observed process Y_n, of the scale and mixture parameters of the filtered density $\nu_{l|l:1} = \mathcal{L}(X_l|Y_l, \dots Y_1)$ and of the predicted density $\nu_{l+1|l:1} = \mathcal{L}(X_{l+1}|Y_l, \dots Y_1)$.

smallest integer with $\sum_{i>L} \alpha_i \leq 10^{-9}$. The remaining $l(\alpha) - L$ coefficients were then set to zero.

On the other hand, in Figure 3, a trajectory of the process $\sigma_{l|l:1}$, the scale parameter of the filtered density $\nu_{l|l:1}$, is represented. The parameters were set to the same values as for Table 1. In Genon-Catalot and Kessler (2004), the existence of a stationary distribution for $\nu_{l|l:1}$ was proved in the case when the hidden process is ergodic. The plotted trajectory is coherent with this stationarity property.

4.3. Comparison of the prediction, filtered and smoothed distributions

In Figure 4 a visual comparison of the prediction density $\mathcal{L}(X_{10}|Y_9, \dots, Y_1)$, the filtered density $\mathcal{L}(X_{10}|Y_{10}, \dots, Y_1)$ and the smoothing density $\mathcal{L}(X_{10}|Y_{11}, \dots, Y_1)$ is presented for one trajectory of the observed process. It illustrates the improvement of our inference about the unobserved value X_{10}, as more observations of Y are collected. The value of the parameters used for the simulation are the same as in the previous subsection. The true unobserved value of X_{10} is 0.1041.

Notice that $Y_9 = 0.3400$, which influences the centre of the predicted density $\mathcal{L}(X_{10}|Y_9, \dots, Y_1)$.

Smoothed conditional distributions when more observations of Y are available, e.g $\mathcal{L}(X_{10}|Y_{12}, \dots, Y_1)$, $\mathcal{L}(X_{10}|Y_{13}, \dots, Y_1)$, etc..., can be computed using the formulae in Proposition 3.1. However, the new observations Y_{12}, Y_{13}, ... do not contain much information about X_{10} and the smoothed densities are very similar to $\mathcal{L}(X_{10}|Y_{11}, \dots, Y_1)$.

Finally, it is worth mentioning that the scale parameters of the plotted distributions are

$$\sigma_{10|9:1} = 0.1585, \qquad \sigma_{10|10:1} = 0.0466, \qquad \sigma_{10|11:1} = 0.0448.$$

On the other hand, and for the same trajectory, the evolution of the conditional mean $E[X_{10}|Y_n, \dots, Y_1]$ as n varies from 1 to 11 is plotted in Figure 5. The top horizontal line corresponds to the mean of ν_{0,σ_s}, see (12), the stationary distribution for X, which corresponds to the best guess for the value of X_{10}, when no observations of Y are available. The bottom line corresponds to the true unobserved value of X_{10}. We can see that, for n up to 7, the conditional mean $E[X_{10}|Y_n, \dots, Y_1]$ is close to $\mathbb{E}[X_{10}]$, due to the fact that the observations $Y_1, \dots Y_7$ contain little information about X_{10}. From $n = 8$, the value of the conditional mean is influenced by the values of Y, getting close to the unobserved X_{10} for $n = 10$ and even better for $n = 11$. From $n = 12$, $E[X_{10}|Y_n, \dots, Y_1]$ does not vary significantly with respect to $E[X_{10}|Y_{11}, \dots, Y_1]$.

4.4. L^2 error of prediction

Assume that the observation up to time n is available, and consider the prediction of the unobserved value X_l, through the mean of the conditional distribution $\mathcal{L}(X_l|Y_n, \dots, Y_1) = \nu_{l|n:1}$, which is the optimal L^2 predictor. Since $\nu_{l|n:1}$ belongs to $\mathcal{F}$ a closed form expression for its moments is available, see (20), therefore the optimal L^2 predictor admits a closed form expression and so does its associated conditional L^2 square-error:

$$err^2(l; Y_n, \dots, Y_1) = \mathbb{E}\left[(X_l - \mathbb{E}[X_l|Y_n, \dots, Y_1])^2 | Y_n, \dots, Y_1\right].$$

In order to compare the goodness of the prediction for different values of n, a Monte-Carlo approximation of the unconditional L^2 square-error

$$err^2(l; n) = \mathbb{E}[err^2(l; Y_n, \dots, Y_1)]$$

was performed, based on the simulation of 10000 trajectories of the observed process. The time step was chosen to be $\Delta = 0.5$, while the parameters values were set to $\theta = 0.5$, $\sigma = 0.2$ for the hidden process X, see (2). In particular, the hidden chain is ergodic and X_1 was simulated from the invariant distribution. The noise specification was $k = 2$ in (14) and λ was set to satisfy (37). In Table 2, the mean $\widehat{err^2(10; n)}$ and the associated 95% margin error computed from 10000 values of $err^2(10; Y_n, \dots, Y_1)$ when n ranges from 9 to 12 is reported.

As expected, while there is a significant improvement in the L^2 square error, $err(10; n)^2$ when n varies from 9 to 11, there is no significant differences between $n = 11$ and $n = 12$ (and higher values of n not reported here): Y_{12} contains little relevant information about X_{10}.

n	9	10	11	12
$\widehat{err^2(10;n)}$	0.01101	0.00316	0.00280	0.00277
	(± 8.98e-05)	(± 6.23e-05)	(± 5.26e-05)	(± 5.16e-05)

Table 2. Monte-Carlo approximation of the unconditional prediction square error $\mathbb{E}[(X_{10} - \mathbb{E}[X_{10}|Y_n, \ldots, Y_1])^2]$, for different values of n. The means together with the associated 95% margin error (within brackets) of the conditional square error $\mathbb{E}[(X_{10} - \mathbb{E}[X_{10}|Y_n, \ldots, Y_1])^2|Y_n \ldots, Y_1]$ corresponding to 10000 simulated trajectories are reported.

5. Concluding remarks

In this work, we consider a nonlinear hidden Markov model which shares interesting features with the standard Kalman filter:

1. If the initial distribution of the hidden process is chosen to belong to a given family $\mathcal{F}$ of distributions, the predicting, filtering and smoothing operators admit closed forms.
2. The action of these operators are expressed in terms of the parameters which characterize a distribution of the family $\mathcal{F}$. A crucial difference with the Kalman filter, though, is due to the fact that $\mathcal{F}$ is not a family parameterized by a fixed number of parameters, for related issues see Chaleyat-Maurel and Genon-Catalot (2006). However, we have seen that, even if the length of the mixture parameter increases linearly theoretically, in practise, only a reasonnable number of coefficients are significant. As a consequence, the implementation of the recursions is easy and not computationnaly prohibitive.
3. A closed form expression for the likelihood is available, which allows to carry out likelihood-based inference.
4. The space of the hidden process is the whole positive line and therefore not compact.

We therefore feel that the detailed obtention of any conditional distribution $\mathcal{L}(X_l|Y_n \ldots, Y_1)$ and the illustration of its behaviour on simulations that we have presented in this work make this model a particularly attractive candidate to serve as a benchmark, together with the standard Kalman filter model, in the investigation of Hidden Markov models methods: numerical procedures like particle filter approximations to the relevant conditional distributions, can be tested on this model and on the other hand, asymptotic results can be investigated.

Let us mention finally that the extension of the stated results to non equispaced observation times is straightforward.

References

Olivier Cappé, Eric Moulines, and Tobias Rydén. *Inference in hidden Markov models*. Springer Series in Statistics. Springer, New York, 2005.

Mireille Chaleyat-Maurel and Valentine Genon-Catalot. Computable infinite-dimensional filters with applications to discretized diffusion processes. *Stochastic Process. Appl.*, 116(10):1447–1467, 2006.

G. B. Di Masi, W. J. Runggaldier, and B. Barozzi. Generalized finite-dimensional filters in discrete time. In *Nonlinear stochastic problems (Algarve, 1982)*, volume 104 of *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.*, pages 267–277. Reidel, Dordrecht, 1983.

Valentine Genon-Catalot. A non-linear explicit filter. *Statist. Probab. Lett.*, 61(2):145–154, 2003.

Valentine Genon-Catalot and Mathieu Kessler. Random scale perturbation of an AR(1) process and its properties as a nonlinear explicit filter. *Bernoulli*, 10(4):701–720, 2004.

Wolfgang J. Runggaldier and Fabio Spizzichino. Sufficient conditions for finite dimensionality of filters in discrete time: a Laplace transform-based approach. *Bernoulli*, 7 (2):211–221, 2001.

Günther Sawitzki. Finite-dimensional filter systems in discrete time. *Stochastics*, 5(1-2): 107–114, 1981.

Appendix

Proof of Proposition 3.1.

Proof of Part 1, updating step

Assume $\mathcal{L}(X_l)$ is $\nu_{\alpha,\sigma}$ in $\mathcal{F}$. The law $\mathcal{L}(X_l|Y_l)$ is, by the Bayes rule, proportional to $\sum \alpha_i g_{i,\sigma}(x)p_x(y_l)$. Now, from (15),

$$g_{i,\sigma}(x)p_x(y_l) = \frac{C_{2(i+k)}}{C_{2i}} \frac{(T_{y_l}(\sigma))^{2(i+k)+1}}{\sigma^{2i+1}} g_{i+k,T_{y_l}(\sigma)}(x) \times C,$$

where T_{y_l} is defined as

$$\frac{1}{T^2_{y_l}(\sigma)} = \frac{1}{\sigma^2} + \frac{2\lambda}{y_l^2},$$

and C is a constant that does not depend on i or on y_l. Some simple manipulations lead to

$$g_{i,\sigma}(.)p_.(y_l) \propto (2j-1)(2j-3)\dots(2(j-k)+1)\left(\frac{T_{y_l}(\sigma)}{\sigma}\right)^{2(j-k)} g_{i+k,T_{y_l}(\sigma)}(.),$$

where the proportionnality constant does not depend on i. We easily deduce (21) and (22).

Proof of Part 2, prediction step

We first begin by describing the action of P on a distribution $g_{i,\sigma}$:

Lemma 5.1. *Consider $g_{i,\sigma}$ and P, see (8) and (9),*

$$g_{i,\sigma}P = \sum_{j=0}^{i} \alpha_j^{(i,\sigma)} g_{j,\tau(\sigma)} \tag{39}$$

with

$$\tau^2(\sigma) = \beta^2 + a^2\sigma^2, \tag{40}$$

and for $j = 0, 1, \ldots, i$,

$$\alpha_j^{(i,\sigma)} = \binom{i}{j} \left(1 - \frac{\beta^2}{\tau^2(\sigma)}\right)^j \left(\frac{\beta^2}{\tau^2(\sigma)}\right)^{i-j} \tag{41}$$

The result holds for $\sigma = 0$.

This lemma is essentially proved in Chaleyat-Maurel and Genon-Catalot (2006), Proposition 3.4 and Genon-Catalot and Kessler (2004).
The result of the above lemma is easily extended to the case of a mixture of distributions $g_{i,\sigma}$, and we deduce (23)to (25) after noticing that $\nu_{l+r|l:1}$ is simply $\nu_{l|l:1}P^r$ and that $P^r = P_\Delta^r = P_{r\Delta}$ is associated with $a(r\Delta)$ and $\beta(r\Delta)$.

Proof of Part 3, smoothing

We use the following classical result (see *e.g.* Cappé et al., 2005, p. 63),

$$\text{For } l \leq n, \quad \nu_{l|n:1}(dx_l) = \frac{p_{l,n}(y_{l+1}, \ldots, y_n; x_l)}{\prod_{i=l+1}^{n} p(y_i|y_{i-1}, \ldots, y_1)} \nu_{l|l:1}(dx_l), \tag{42}$$

where $p_{l,n}(y_{l+1}, \ldots, y_n; x_l)$ denotes the conditional density of $(Y_{l+1}, \ldots, Y_n)$ given $X_l = x_l$. By convention, we set $p_{n,n}(\emptyset; x) = 1$.

The proof consists therefore in obtaining explicit recursive relations to compute $p_{l,n}(y_{l+1}, \ldots, y_n; x_l)$.

We begin by describing a backward recursion for these quantities:
Recall that, for any function h for which the integral below is well defined,

$$Ph(x) = \int_0^{+\infty} p_\Delta(x, x')h(x')dx'. \tag{43}$$

We have

Lemma 5.2. *On one hand, for all n,*

$$p_{n-1,n}(y_n; x) = P(p_\cdot(y_n))(x).s \tag{44}$$

On the other hand, for $l + 1 < n$,

$$p_{l,n}(y_{l+1}, \ldots, y_n; x) = P(p_\cdot(y_{l+1})p_{l+1,n}(y_{l+2}, \ldots, y_n; \cdot))(x) \tag{45}$$

Proof. Given $X_{n-1}=x$, (X_n,Y_n) has distribution $p_\Delta(x,x_n)p_{x_n}(y_n)dx_ndy_n$. We deduce (44). Then, for $n\geq l+2$,

$$\begin{aligned}p_{l,n}(y_{l+1},\ldots,y_n;x) &= \int_0^{+\infty} p_\Delta(x,x_{l+1})p_{x_{l+1}}(y_{l+1}) \\ &\quad \times \prod_{i=l+2}^{n} p_\Delta(x_{i-1},x_i)p_{x_i}(y_i)dx_{l+1}\ldots dx_n \\ &= \int_0^{+\infty} p_\Delta(x,x_{l+1})p_{x_{l+1}}(y_{l+1})p_{l+1,n}(y_{l+2},\ldots,y_n;x_{l+1})dx_{l+1},\end{aligned}$$

which gives (45) . ∎

We now search for the concrete algorithm to compute (45) and start with a lemma.

Lemma 5.3. *For $i\geq 0,\sigma>0$, let us set*

$$h_{i,\sigma}(x)=x^{2i}\exp\left(-\frac{x^2}{2\sigma^2}\right).$$

Then, (see (43)),

$$Ph_{i,\sigma}=\sum_{l=0}^{i} c_l^{i,\sigma}h_{l,\Phi(\sigma)},$$

where

$$\frac{1}{\Phi^2(\sigma)}=\frac{a^2}{\beta^2+\sigma^2}, \tag{46}$$

and

$$c_l^{i,\sigma}=\frac{C_{2i}}{C_{2l}}\binom{i}{l}\left(\frac{\sigma^2}{\beta^2+\sigma^2}\right)^{i+l+(1/2)} a^{2l}\ \beta^{2(i-l)}. \tag{47}$$

Proof. We check from (5), that the transition density $p_\Delta(x,x')$ is an even function of x'. Therefore

$$Ph_{i,\sigma}=\frac{1}{2}(I(a)+I(-a)), \tag{48}$$

where

$$I(a)=\int_{\mathbb{R}} h_{i,\sigma}(x')\exp(-\frac{(x'-ax)^2}{2\beta^2})\beta^{-1}(2\pi)^{-1/2}dx'. \tag{49}$$

We have

$$\frac{x'^2}{\sigma^2}+\frac{(x'-ax)^2}{\beta^2}=\frac{1}{T^2}(x'-\mu x)^2+\frac{x^2}{\Phi^2(\sigma)}, \tag{50}$$

where

$$\frac{1}{T^2}=\frac{1}{\sigma^2}+\frac{1}{\beta^2},\quad \text{equivalently}\quad T^2=\frac{\sigma^2\beta^2}{\sigma^2+\beta^2}, \tag{51}$$

$$\mu = \frac{a/\beta^2}{1/T^2} = a\frac{\sigma^2}{\sigma^2+\beta^2}, \tag{52}$$

and

$$\frac{1}{\Phi^2(\sigma)} = \frac{a^2}{\beta^2}\left(-1+\frac{1/\beta^2}{1/T^2}\right) = -\frac{a^2}{\beta^2+\sigma^2}. \tag{53}$$

The following lemma is useful.

Lemma 5.1. *For all $\mu \in R$ and $T > 0$, we have*

$$\int_R u^{2i}\exp(-\frac{(u-\mu)^2}{2T^2})(2\pi)^{-1/2}T^{-1}du = C_{2i}\sum_{k=0}^{i}\frac{1}{C_{2k}}\binom{i}{k}\mu^{2k}T^{2(i-k)}$$

Proof. We have:

$$\int_R (v+\mu)^{2i}\exp(-\frac{v^2}{2T^2})(2\pi)^{-1/2}T^{-1}dv = \sum_{k=0}^{i} C_{2(i-k)}\binom{2i}{2k}\mu^{2k}T^{2(i-k)}.$$

Noting that

$$\binom{2i}{2k}C_{2(i-k)} = \binom{i}{k}\frac{C_{2i}}{C_{2k}}, \tag{54}$$

we get the result . ∎

We apply this lemma to obtain

$$I(a) = \sum_{l=0}^{i} c_l^{i,\sigma} h_{l,\Phi(\sigma)}, \tag{55}$$

with $\Phi(\sigma)$ and the $c_l^{i,\sigma}$ given by

$$\frac{1}{\Phi^2(\sigma)} = \frac{a^2}{\beta^2+\sigma^2}, \tag{56}$$

$$c_l^{i,\sigma} = \frac{C_{2i}}{C_{2l}}\binom{i}{l}\left(\frac{\sigma^2}{\beta^2+\sigma^2}\right)^{i+l+(1/2)} a^{2l}\beta^{2(i-l)} \tag{57}$$

Since $I(a) = I(-a)$, we obtain that $Ph_{i,\sigma} = I(a)$. . ∎

To conclude the proof of point a) of the part III of Proposition 3.1, we, on one hand, apply Lemma 5.3 to obtain 31 and, on the other hand, assuming that

$$p_{l',n}(y_{l'+1},\ldots,y_n;x) = \sum_{j=0}^{(n-l')k} d_j^{l',n}\, h_{j,\Phi_{l',n}}(x), \tag{58}$$

holds for $l' = l+1$, we deduce it for $l' = l$ by checking that

$$
\begin{aligned}
p_\cdot(y_{l+1})p_{l+1,n}(y_{l+2},\dots,y_n;\cdot) &= \\
&\frac{2\lambda^k}{\Gamma(k)\ y_{l+1}^{2k+1}} h_{k,y_{l+1}/\sqrt{2\lambda}}(.) \sum_{j=0}^{(n-l-1)k} d_j^{l+1,n} h_{j,\Phi_{l+1,n}}(.) \\
&= \frac{2\lambda^k}{\Gamma(k)\ y_{l+1}^{2k+1}} \sum_{j=0}^{(n-l-1)k} d_j^{l+1,n} h_{k+j,\gamma_{l,n}}(.),
\end{aligned}
$$

with $\gamma_{l,n}$ given by (32).
We then apply Lemma 5.3 to compute $P(h_{j+k,\gamma_{l,n}})$ and get

$$
P(h_{j+k,\gamma_{l,n}}) = \sum_{t=0}^{j+k} c_t^{j+k,\gamma_{l,n}} h_{t,\Phi(\gamma_{l,n})}. \tag{59}
$$

It remains to interchange the sums $\sum_{j=0}^{(n-l-1)k} \sum_{t=0}^{j+k}$ and join all terms. This gives (32)-(33) . ∎

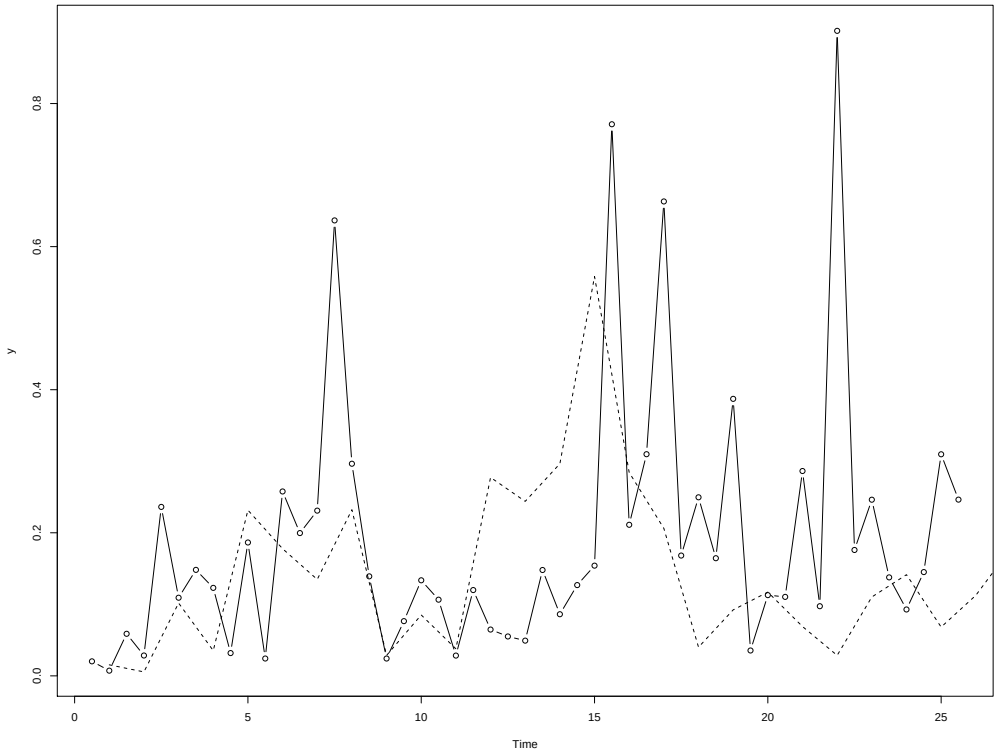
Proof of part III, smoothing b). Using the same notation for the measure and its density, we write $\nu_{l|l:1} = \sum \alpha_{l|l:1}(i) g_{i,\sigma_{l|l:1}}$. We deduce from (58) for $l' = l$ that

$$
\begin{aligned}
\nu_{l|n:1}(x_l) &\propto \sum_{i,j} \alpha_{l|l:1}(i) d_j^{l,n} \frac{1}{\sigma_{l:l:1}^{2i+1}} \frac{2}{\sqrt{2\pi} C_{2i}} h_{i,\sigma_{l|l:1}}(x_l) \\
&\propto \frac{1}{\prod_{i=l+1}^n p(y_i|y_{i-1},\dots,y_1)} \sum_{i,j} \alpha_{l|l:1}(i) d_j^{l,n} \frac{C_{2(i+j)}}{C_{2i}} \frac{\tau^{2(i+j)+1}}{\sigma_{l:l:1}^{2i+1}} g_{i+j,\tau}(x_l),
\end{aligned}
$$

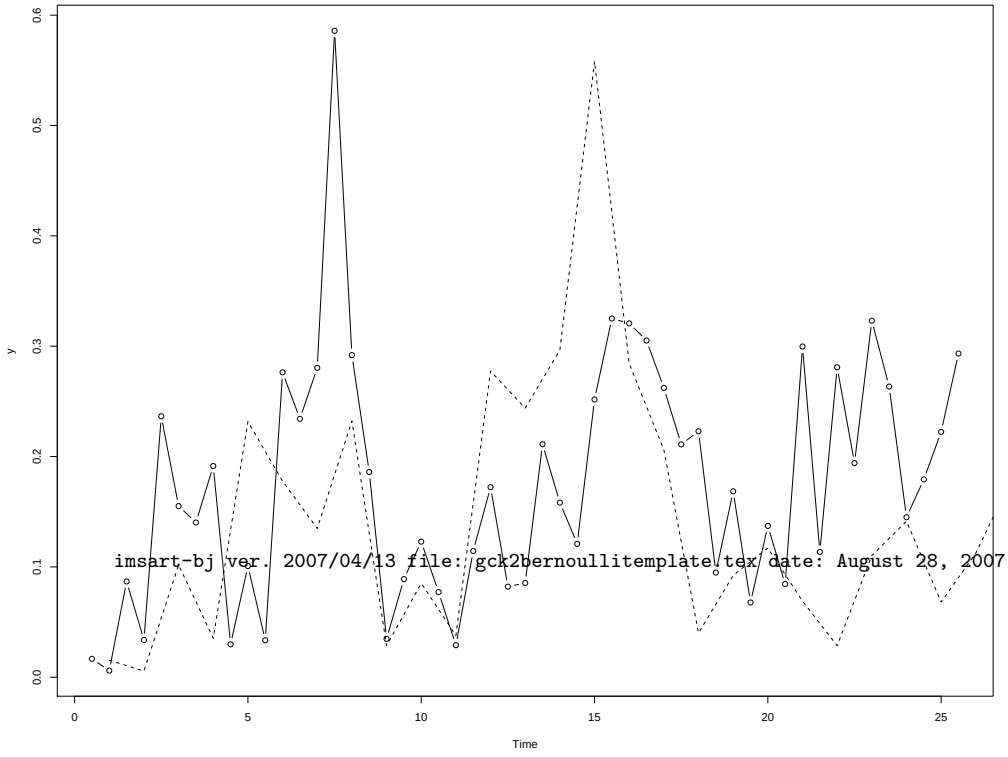
where $\tau = \sigma_{l|n:1}$ satisfies

$$
\frac{1}{\tau^2} = \frac{1}{\sigma_{l|l:1}^2} + \frac{1}{\phi_{l,n}^2}.
$$

The relation (36) is obtained rearranging terms in the above sum, after carrying out the change of indexes: $(i,j) \to (i, t = i+j)$.



a)



b)

Figure 1. Two trajectories for different noise levels. Solid line: the observed process $(Y_n)_{1\leq n\leq 50}$, dashed line: the hidden process $(X_n)_{1\leq n\leq 50}$. For both a) and b), the noise level was calibrated by choosing λ

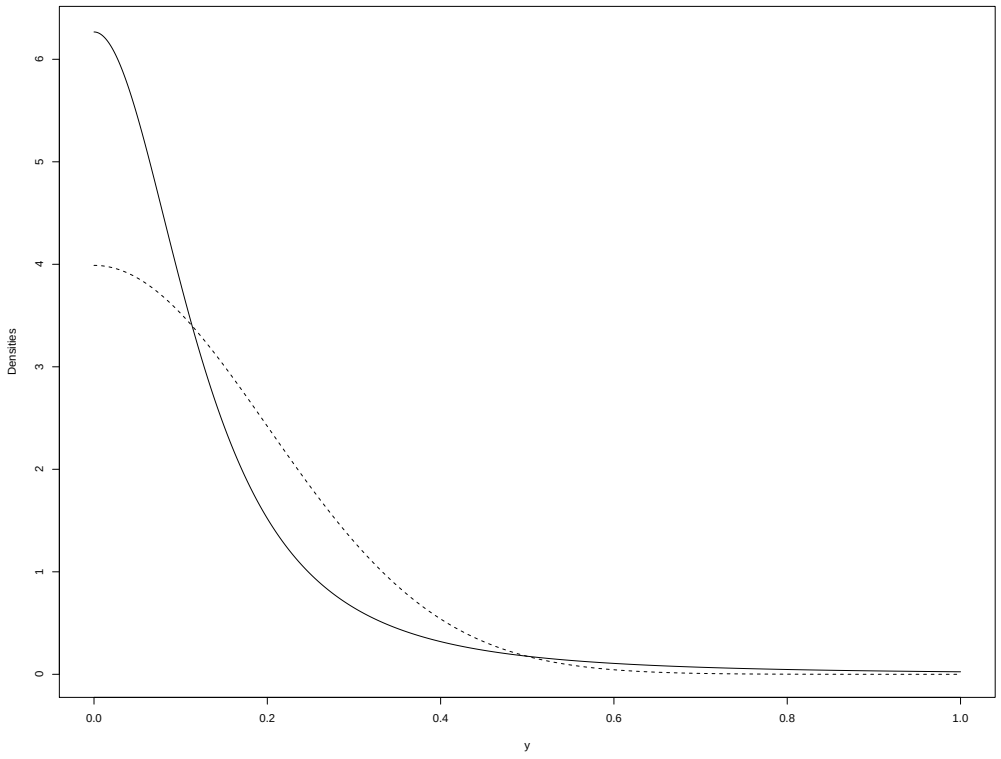


Figure 2. Marginal densities of the observed process Y_n (solid line) and the hidden process X_n (dashed line) in the stationary state. The parameter k in the noise specification was chosen to be $k = 1$.

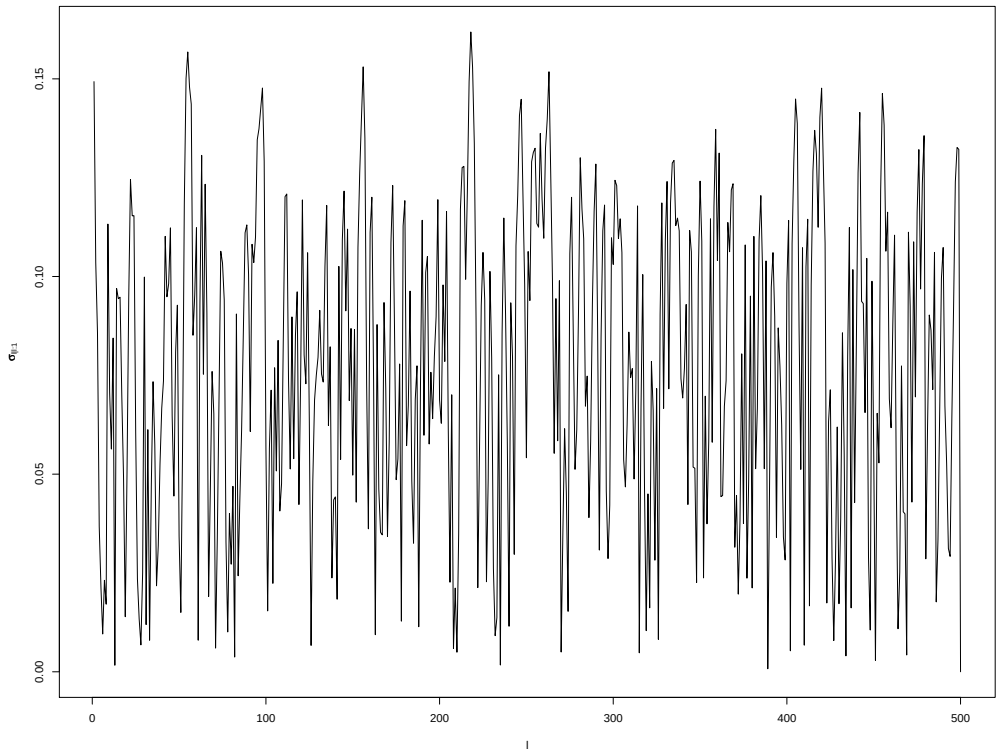


Figure 3. Evolution of the scale parameter $\sigma_{l|l:1}$ of the conditional distribution $\mathcal{L}(X_l|Y_l,\ldots,Y_1)$ as l increases. The hidden process (X_n) is specified to be ergodic.

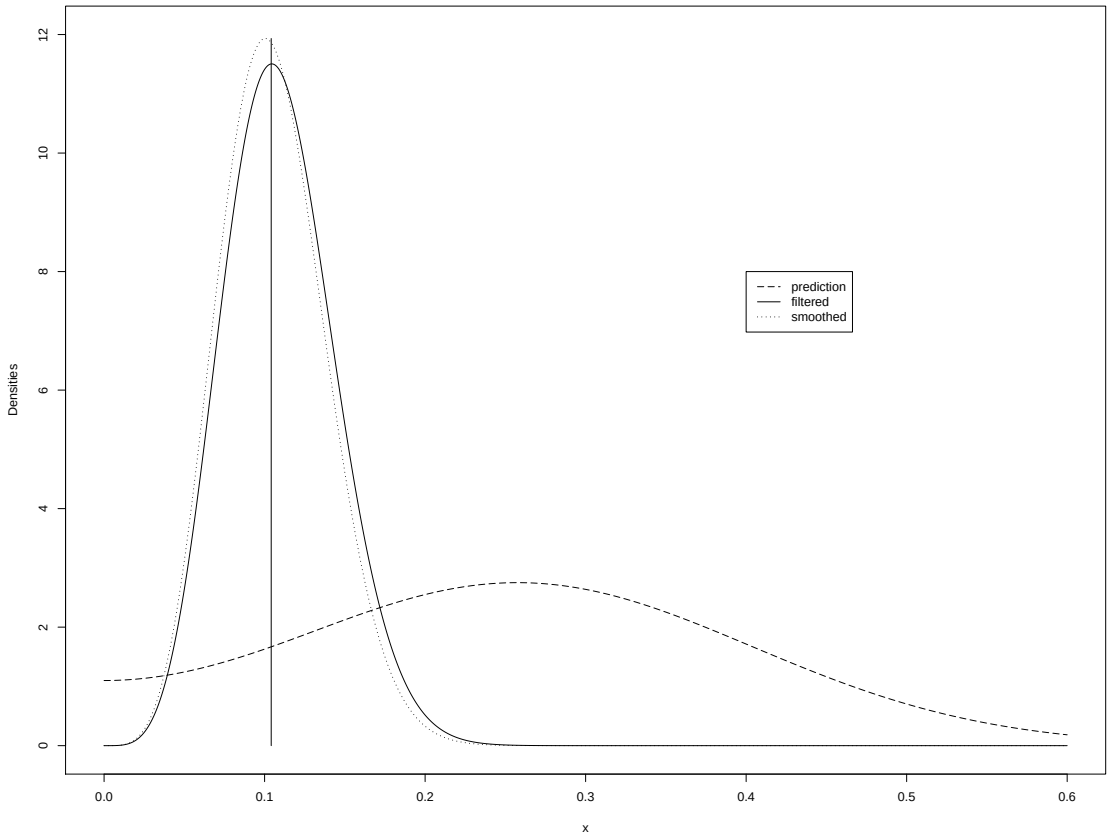


Figure 4. Comparison, for one trajectory of Y, of the three conditional densities: $\mathcal{L}(X_l|Y_n,\ldots,Y_1)$ for $l = 10$ and $n = 9$ (predicted, dotted), $n = 10$ (filtered, solid), and $n = 11$ (smoothed, dashed). The vertical line corresponds to the value of the unobserved X_l.

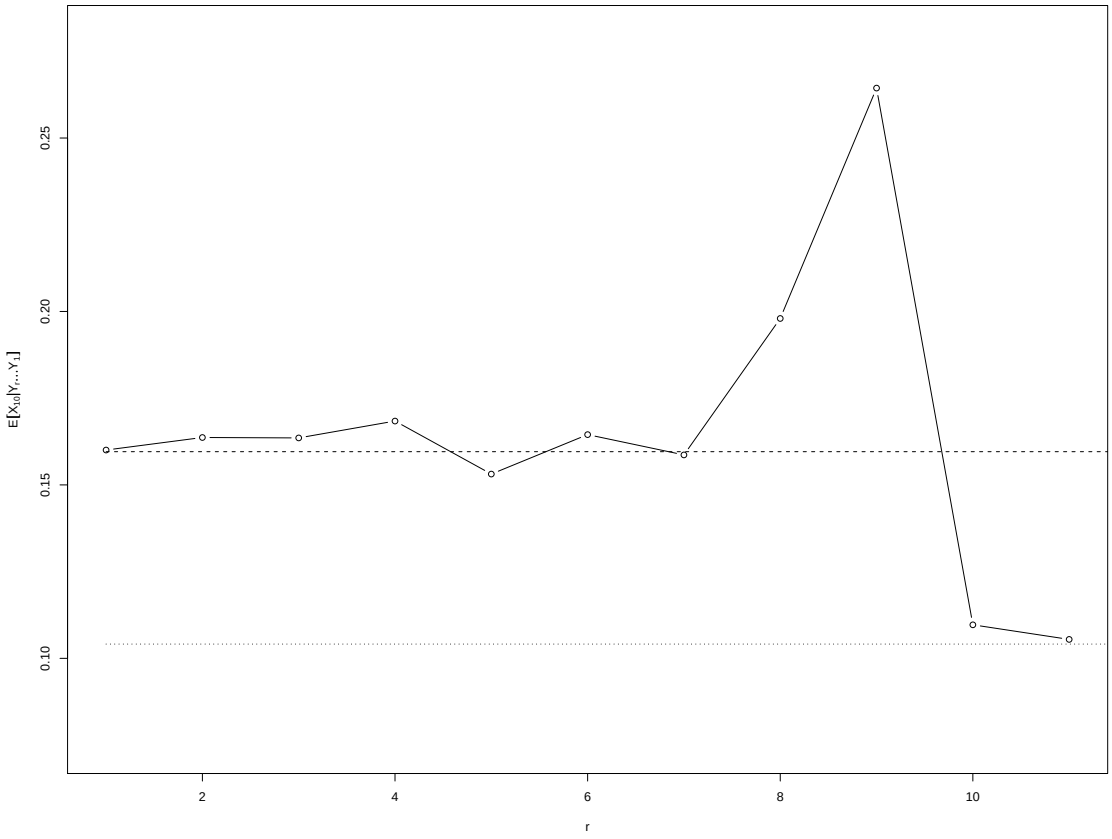


Figure 5. Evolution, for one trajectory of Y, of the conditional mean $E[X_{10}|Y_n, \ldots, Y_1]$ as n varies from 1 to 11. The top horizontal line (dashed) corresponds to the mean of the invariant distribution of X_{10}, while the bottom horizontal line (dotted) corresponds to the true unobserved value of X_{10}.