



# Computable infinite dimensional filters with applications to discretized diffusion processes.

Mireille Chaleyat-Maurel, Valentine Genon-Catalot

## ► To cite this version:

Mireille Chaleyat-Maurel, Valentine Genon-Catalot. Computable infinite dimensional filters with applications to discretized diffusion processes.. 2005. hal-00004889

**HAL Id: hal-00004889**

**<https://hal.science/hal-00004889>**

Preprint submitted on 9 May 2005

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Computable infinite dimensional filters with applications to discretized diffusion processes.

MIREILLE CHALEYAT-MAUREL<sup>1</sup> and VALENTINE GENON-CATALOT<sup>2</sup>

<sup>1</sup> *Université René Descartes Paris 5, U.F.R. de Mathématiques et Informatique, Laboratoire MAP5 (CNRS UMR 8145) et Laboratoire de Probabilités et Modèles Aléatoires (CNRS-UMR 7599), 45, rue des Saints-Pères, 75270 Paris Cedex 06, France. e-mail: mcm@math-info.univ-paris5.fr*

<sup>2</sup> *Université René Descartes Paris 5, U.F.R. de Mathématiques et Informatique, Laboratoire MAP5 (CNRS-UMR 8145), 45, rue des Saints-Pères, 75270 Paris Cedex 06, France. e-mail: genon@math-info.univ-paris5.fr*

## Abstract

Let us consider a pair signal-observation  $((x_n, y_n), n \geq 0)$  where the unobserved signal  $(x_n)$  is a Markov chain and the observed component is such that, given the whole sequence  $(x_n)$ , the random variables  $(y_n)$  are independent and the conditional distribution of  $y_n$  only depends on the corresponding state variable  $x_n$ . The main problems raised by these observations are the prediction and filtering of  $(x_n)$ . We introduce sufficient conditions allowing to obtain computable filters using mixtures of distributions. The filter system may be finite or infinite dimensional. The method is applied to the case where the signal  $x_n = X_{n\Delta}$  is a discrete sampling of a one dimensional diffusion process: Concrete models are proved to fit in our conditions. Moreover, for these models, exact likelihood inference based on the observation  $(y_0, \dots, y_n)$  is feasible.

**MSC:** primary 93E11, 60G35; secondary 62C10.

**Keywords:** Stochastic filtering, diffusion processes, discrete time observations, hidden Markov models, prior and posterior distributions.

**Running title:** Computable filters.

# 1 Introduction

Let us consider a pair signal-observation  $((x_n, y_n), n \geq 0)$  where the unobserved signal  $(x_n)$  is a Markov chain and the observed component is such that, given the whole sequence  $(x_n)$ , the random variables  $(y_n)$  are independent and the conditional distribution of  $y_n$  only depends on the corresponding state variable  $x_n$ . This is a classical setting in the field of non linear filtering and the process  $(y_n)$  is often called a hidden Markov model.

In this context, a central problem that has been the subject of a huge number of contributions is the study of the exact filter, *i.e.* the sequence of conditional distributions of  $x_n$  given  $y_n, \dots, y_1, y_0$ ,  $n \geq 0$ . On the other hand, statistical inference based on the (non Markovian) observations  $(y_0, \dots, y_n)$  requires the knowledge of the successive conditional distributions of  $y_n$  given  $y_{n-1}, \dots, y_0$ . These are obtained through the prediction filter, *i.e.* the sequence of conditional distributions of  $x_n$  given  $y_{n-1}, \dots, y_1, y_0$ . Although the exact and the prediction filter may both be calculated recursively by an explicit algorithm, iterations become rapidly intractable and exact formulae are difficult to obtain. To overcome this difficulty, authors generally try to find a parametric family  $\mathcal{F}$  of distributions on the state space  $\mathcal{X}$  of  $(x_n)$  (*i.e.* a family of distributions specified by a finite fixed number of real parameters) such that if  $\mathcal{L}(x_0) \in \mathcal{F}$ , then, for all  $n$ ,  $\mathcal{L}(x_n|y_n, \dots, y_1, y_0)$  and  $\mathcal{L}(x_n|y_{n-1}, \dots, y_1, y_0)$  both belong to  $\mathcal{F}$ . In this case, the model  $((x_n, y_n), \mathcal{F})$  is called a finite-dimensional filter system and it is enough to describe each conditional distribution by the parameters that characterize it. This situation is illustrated by the linear Gaussian Kalman filter (see below Section 3). Whenever the initial distribution of the signal is Gaussian, specified by its mean and variance, then, all the successive conditional distributions are Gaussian and there is an explicit recursive algorithm which gives the stochastic process of the conditional means and variances.

Necessary and sufficient conditions for the existence of finite-dimensional filters in discrete time have been given in Sawitzki (1981) (see also Runggaldier and Spizzichino (2001)). The case of continuous time filters was treated in Chaleyat-Maurel and Michel (1984). As a consequence of these papers, it appears that very few finite dimensional filters are available and they are often obtained as the result of an *ad hoc* construction (see *e.g.*, the new constructive approach presented in Ferrante and Vidoni (1998) and the references therein).

In what follows, we propose a method to obtain computable filters that may be finite or infinite dimensional. Our approach is a generalization of the one developed in Genon-Catalot (2003) and Genon-Catalot and Kessler (2004) for a special model. The method is well fitted for the filtering of discretized diffusion processes and illustrated with examples.

More precisely, (Section 2) we consider at first a sequence of parametric families of distributions  $\mathcal{F}^i = \{\nu_\theta^i, \theta \in \Theta\}$ ,  $i \in \mathbb{N}$ , where  $\Theta \subset \mathbb{R}^p$  is a parameter set. Then, we construct an enlarged family by means of mixtures. Let us define the set  $S$  of mixture parameters:

$$S = \{\alpha = (\alpha_i, i \geq 0), \forall i \geq 0, \alpha_i \geq 0, \sum_{i=0}^{\infty} \alpha_i = 1\}. \quad (1)$$

Then, we set:

$$\bar{\mathcal{F}} = \{\nu = \sum_{i \geq 0} \alpha_i \nu_\theta^i = \nu_{\theta, \alpha}, \alpha = (\alpha_i, i \geq 0) \in S, \theta \in \Theta\}. \quad (2)$$

Each distribution  $\nu = \nu_{\theta, \alpha}$  in the above class is thus specified by an usual parameter  $\theta$  and a mixture parameter  $\alpha = (\alpha_i, i \geq 0)$ . We give sufficient conditions on the class  $\mathcal{F} = \cup_{i \in \mathbb{N}} \mathcal{F}^i$  ensuring that if  $\nu = \mathcal{L}(x_0)$  belongs to  $\bar{\mathcal{F}}$ , then, the exact and the prediction filter both evolve within  $\bar{\mathcal{F}}$ . These conditions involve the conditional distribution of  $y_n$  given  $x_n = x$  and the transition operator of the hidden chain  $(x_n)$ . Of course, the most interesting case is when the mixture distributions obtained for the filters have a finite number of components. To this end, we introduce the sub-class  $\bar{\mathcal{F}}_f$  of  $\bar{\mathcal{F}}$  composed of distributions  $\nu = \nu_{\theta, \alpha}$  such that  $\alpha_i = 0$  for  $i$  greater than some integer  $N$ . We give sufficient conditions ensuring that, when  $\mathcal{L}(x_0) \in \bar{\mathcal{F}}_f$ , then, the exact and the prediction filter evolve in  $\bar{\mathcal{F}}_f$  (Conditions (C1)-(C2) and Theorem 2.1). They are thus specified by a finite number of parameters. This finite number may change along the iterations. Nevertheless, the filters are explicitly and exactly computable. Let us note that our method has links with the one developed in Di Masi et al (1983). In this paper, the distributions of the filters are allowed to be finite linear combinations of parametric distributions. The number of terms in the linear combination may also change along the iterations. However, the coefficients in each linear combination are possibly negative. So these distributions are not mixtures of parametric distributions and their interpretation is therefore difficult.

In Section 3, we illustrate our method on the classical Kalman filter with non Gaussian initial distributions. We introduce an appropriate class of non Gaussian distributions and give the corresponding explicit formulae for the filters. In Sections 4-5-6, we consider models satisfying our sufficient conditions and for which the signal  $x_n = X_{n\Delta}$  is a discrete sampling of a one-dimensional diffusion process. In Section 4, we study the observation equation  $y_n = x_n w_n$  where the signal  $(x_n)$  and the noise  $(w_n)$  are independent sequences of positive random variables. The signal  $x_n$  is a discrete sampling of the diffusion process given by the stochastic differential equation

$$dX_t = (\theta X_t + \frac{(\delta - 1)\sigma^2}{2X_t})dt + \sigma d\beta_t, \quad (3)$$

where  $(\beta_t)$  is a standard one-dimensional Brownian motion,  $\theta$  is a real parameter,  $\sigma$  is positive and  $\delta$  is a real number such that  $\delta > 1$ . This process is called the radial Ornstein-Uhlenbeck process. In the case  $\delta = 1$ , we consider the absolute value of a one dimensional Ornstein-Uhlenbeck process. The noises  $w_n$  have a specific distribution to ensure explicit formulae. In Sections 5 and 6, we exploit the results of Section 4 to develop models based on the same signal but different observation equations. In particular, in Section 5, we propose some stochastic volatility type models. Section 7 contains some concluding remarks. Some technical proofs are given in the Appendix.

## 2 Sufficient conditions for computable filters.

### 2.1 The filtering- prediction algorithm.

We introduce our notations in an abstract framework that will become concrete through the examples below.

We denote by  $\mathcal{X}$  the state-space of the unobserved Markov chain  $(x_n)$ , which is equipped

with a sigma-field  $\mathcal{B}$ . We assume that this chain is time-homogeneous and denote by  $P$  its transition operator. The time-homogeneity assumption is here for the sake of simplicity of notations. It is not essential and may be relaxed (see Section 7).

The state-space of the observed component  $(y_n)$  will be denoted by  $\mathcal{Y}$  and its sigma-field by  $\mathcal{C}$ . We assume that the conditional distribution of  $y_n$  given  $x_n = x$  does not depend on  $n$  and that this distribution is given by a density with respect to a common dominating measure  $\mu$  on  $\mathcal{Y}$ :

$$F_x(dy) = \mathcal{L}(y_n|x_n = x) = f_x(y)\mu(dy). \quad (4)$$

As stated above, this is also the conditional distribution of  $y_n$  given the whole sequence  $(x_k)$  when  $x_n = x$ . Again, the time-homogeneity assumption here is not essential. On the contrary, the existence of a common dominating measure  $\mu$  for the family of distributions  $(F_x(dy), x \in \mathcal{X})$  is essential for the filtering-prediction recursive equations. In the statistical vocabulary, it means that this family of distributions is a dominated family when  $x$  is considered as a parameter.

In the concrete models that we investigate in further sections, we shall take  $\mathcal{X}$  equal to  $\mathbb{R}$  or  $(0, +\infty)$ , and  $\mathcal{Y}$  equal to  $\mathbb{R}$ ,  $(0, +\infty)$  or  $\mathbb{N}$ .

Let us now briefly recall the filtering-prediction algorithm (see *e.g.* Del Moral and Guionnet (2001)). Given an initial distribution for  $x_0$ , there is a well known algorithm that allows to compute the successive conditional distributions:

$$\begin{aligned} \mathcal{L}(x_0) &\xrightarrow{\text{updating}} \mathcal{L}(x_0|y_0) \xrightarrow{\text{prediction}} \mathcal{L}(x_1|y_0) \\ &\xrightarrow{\text{updating}} \mathcal{L}(x_1|y_1, y_0) \xrightarrow{\text{prediction}} \mathcal{L}(x_2|y_1, y_0) \dots \end{aligned}$$

By the above iterations, we get two kinds of distributions on  $\mathcal{X}$ :

$$\nu_{n|n:0} = \mathcal{L}(x_n|y_n, \dots, y_0), \quad (5)$$

$$\nu_{n+1|n:0} = \mathcal{L}(x_{n+1}|y_n, \dots, y_0), \quad (6)$$

The distribution (5) is called the optimal or exact filter and (6) is the prediction filter.

These are obtained using two steps: the updating and the prediction steps which can be described by introducing the following operators. Let  $\mathcal{P}(\mathcal{X})$  denote the set of probability measures on  $\mathcal{X}$ . For  $\nu \in \mathcal{P}(\mathcal{X})$ , the probability  $\varphi_y(\nu)$  is defined by (see (4)):

$$\varphi_y(\nu)(dx) = \frac{f_x(y)\nu(dx)}{p_\nu(y)}, \quad (7)$$

(with the convention that  $0/0 = 0$ ), where the denominator is equal to

$$p_\nu(y) = \int_{\mathcal{X}} \nu(d\xi) f_\xi(y). \quad (8)$$

The operator  $\varphi_y$  is called the up-dating operator which allows to take into account a new observation.

On the other hand, the prediction operator is as follows. For  $\nu \in \mathcal{P}(\mathcal{X})$ , the probability measure

$$\psi(\nu) = \nu P \quad (9)$$

is obtained by applying the transition operator of the hidden Markov chain. It is defined by:

$$\psi(\nu)(dx') = \nu P(dx') = \int_{\mathcal{X}} \nu(dx) P(x, dx'). \quad (10)$$

We have, for  $n \geq 0$ , (with  $\nu_{0|-1:0} = \nu_0 = \mathcal{L}(x_0)$ )

$$\nu_{n|n:0} = \varphi_{y_n}(\nu_{n|n-1:0}), \quad \nu_{n+1|n:0} = \psi(\nu_{n|n:0}). \quad (11)$$

The optimal filter is obtained via the operator

$$\hat{\Phi}_y = \varphi_y \circ \psi, \quad (12)$$

and

$$\nu_{n+1|n+1:0} = \hat{\Phi}_{y_{n+1}}(\nu_{n|n:0}) = \hat{\Phi}_{y_{n+1}} \circ \dots \circ \hat{\Phi}_{y_1} \circ \varphi_{y_0}(\nu_0). \quad (13)$$

The prediction filter is obtained via the operator

$$\Phi_y = \psi \circ \varphi_y, \quad (14)$$

and

$$\nu_{n+1|n:0} = \Phi_{y_n}(\nu_{n|n-1:0}) = \Phi_{y_n} \circ \dots \circ \Phi_{y_0}(\nu_0). \quad (15)$$

Moreover, the conditional density of  $y_n$  given  $(y_{n-1}, \dots, y_0)$  is obtained as the following marginal density (see (8)):

$$p(y|y_{n-1}, \dots, y_0) = p_{\nu_{n|n-1:0}}(y). \quad (16)$$

And the exact density of  $(y_0, y_1, \dots, y_n)$  is obtained as the product of the successive conditional densities  $p(y_i|y_{i-1}, \dots, y_0)$ .

## 2.2 Sufficient conditions.

The iterations above are rapidly untractable unless both operators (7) and (9) evolve in a parametric family of distributions, *i.e.* distributions specified by a fixed finite number of real parameters. In what follows, this number of parameters will possibly vary along iterations.

More precisely, let us define a class  $\bar{\mathcal{F}}$  of distributions on  $\mathcal{X}$  as follows. First, we start with a parametric class of the form

$$\mathcal{F} = \{\nu_{\theta}^i, \theta \in \Theta, i \in \mathbb{N}\} \quad (17)$$

where  $\Theta \subset \mathbb{R}^p$  is a parameter set. Each distribution in  $\mathcal{F}$  is thus specified by a couple  $(i, \theta) \in \mathbb{N} \times \Theta$  and there is a one-to-one correspondence between  $\mathbb{N} \times \Theta$  and the class  $\mathcal{F}$ .

Then, using the set  $S$  of mixture parameters defined in (1), we build the enlarged class composed of convex combinations of distributions  $\nu_\theta^i$  having the same parameter  $\theta$ :

$$\bar{\mathcal{F}} = \{\nu; \nu = \sum_{i=0}^{\infty} \alpha_i \nu_\theta^i = \nu_{\theta, \alpha}, \theta \in \Theta, \alpha = (\alpha_i, i \geq 0) \in S\}. \quad (18)$$

Now, each distribution  $\nu_{\theta, \alpha}$  on  $\bar{\mathcal{F}}$  is specified by a parameter  $\theta$  and a mixture parameter  $\alpha$ . We stress the fact that all components in a given mixture have the same parameter  $\theta$ . The mixture parameter  $\alpha$  of a  $\nu_{\theta, \alpha}$  may or may not depend on  $\theta$ . For

$$\alpha = \alpha^{(i)} \quad \text{given by} \quad \alpha_i^{(i)} = 1, \quad \alpha_j^{(i)} = 0, j \neq i, \quad (19)$$

we get the distribution  $\nu_\theta^i$ :

$$\nu_\theta^i = \nu_{\theta, \alpha^{(i)}}. \quad (20)$$

Obviously,  $\mathcal{F} \subset \bar{\mathcal{F}}$ . But the resulting extended class may be considerably larger. Of course, the number of components in the mixture can be finite. So, we shall define the length of a mixture parameter  $\alpha$  by

$$l(\alpha) = \sup \{i; \alpha_i > 0\}. \quad (21)$$

We define the sub-class of distributions with finite-length mixture parameter by

$$\bar{\mathcal{F}}_f = \{\nu = \nu_{\theta, \alpha} \in \bar{\mathcal{F}}, l(\alpha) < \infty\}. \quad (22)$$

On the other hand, when  $\alpha$  has infinite length, the series defining an element  $\nu$  in (18) may have an explicit sum, which will be another expression of  $\nu$ .

Now, we want conditions such that, for  $\nu \in \bar{\mathcal{F}}$ ,  $\varphi_y(\nu)$  and  $\psi(\nu)$  both belong to  $\bar{\mathcal{F}}$ . In such a case, it will be enough to express both operators in terms of the couple  $(\theta, \alpha)$  specifying the distributions in  $\bar{\mathcal{F}}$ . Moreover, when the two operators evolve within  $\bar{\mathcal{F}}_f$ , then the exact and the prediction filters are exactly computable even if the number of mixture components varies along the iterations.

Let us consider the following conditions.

- (C1) For all  $y \in \mathcal{Y}$ , for all  $\nu \in \mathcal{F}$ ,  $\varphi_y(\nu) \in \mathcal{F}$  (see (7)). More precisely, for all  $(i, \theta) \in \mathbb{N} \times \Theta$ ,

$$\varphi_y(\nu_\theta^i) = \nu_{T_y(\theta)}^{t_y(i)},$$

with  $T_y(\theta) \in \Theta$ ,  $t_y : \mathbb{N} \rightarrow \mathbb{N}$  a one-to-one mapping and  $(\theta, y) \rightarrow T_y(\theta)$  measurable.

- (C2) For all  $\nu \in \mathcal{F}$ ,  $\psi(\nu) = \nu P \in \bar{\mathcal{F}}$ . More precisely, for all  $(i, \theta)$ ,  $\psi(\nu_\theta^i)$  may be written as

$$\psi(\nu_\theta^i) = \sum_{j \geq 0} \alpha_j^{(i, \theta)} \nu_{\tau(\theta)}^j,$$

where  $\alpha^{(i, \theta)} \in S$ ,  $\tau(\theta) \in \Theta$  and  $\theta \rightarrow (\tau(\theta), \alpha^{(i, \theta)})$  measurable.

- (C2-f) For all  $\nu \in \mathcal{F}$ ,  $\psi(\nu) = \nu P \in \bar{\mathcal{F}}_f$ , with, using the notations of (C2), for all  $(i, \theta)$ ,  $l(\alpha^{(i, \theta)}) = L(i) < \infty$  and the mapping  $i \rightarrow L(i)$  is non decreasing.

- (C3) For all  $x \in \mathcal{X}$ ,  $P(x, dx')$  belongs to the class  $\bar{\mathcal{F}}$ , and may be written as

$$P(x, dx') = \sum_{i \geq 0} \alpha_i^0(x) \nu_{\theta_0}^i(dx'),$$

where  $\alpha^0(x) \in S$ ,  $\theta_0 \in \Theta$  and  $x \rightarrow \alpha^0(x)$  is measurable.

Let us make some comments about these conditions. Condition (C1) concerns only  $\varphi_y$  and the class  $\mathcal{F}$ . The up-dating operator  $\varphi_y$  becomes the following mapping from  $\mathcal{F}$  to  $\mathcal{F}$ :

$$(i, \theta) \rightarrow (t_y(i), T_y(\theta)). \quad (23)$$

Note that, in condition (C1), the function  $t_y(\cdot)$  must not depend on  $\theta$ . Thus, the class  $\mathcal{F}$  is a conjugate class of distributions for the parametric family  $F_x(dy) = f_x(y)d\mu(y)$  (in the sense of Bayesian estimation). Conditions (C2)-(C2-f)-(C3) concern the transition operator  $P$  and the class  $\mathcal{F}$ . Condition (C3) implies that, when the signal starts at a fixed  $x_0 = x$ , then, the distribution of  $x_1$  belongs to the enlarged class  $\bar{\mathcal{F}}$ . Therefore, we can consider that Dirac measures belong to the enlarged class or directly add all Dirac measures to this class. Conditions (C2-f) and (C3) may appear contradictory. Actually, this is not the case because a distribution  $\nu$  in  $\bar{\mathcal{F}}$  may have two different representations, *i.e.* the equality  $\nu_{\theta, \alpha} = \nu_{\theta', \alpha'}$  does not imply  $(\theta, \alpha) = (\theta', \alpha')$ . Moreover, one representation may be finite and the other infinite. We discuss this point in Section 4.

We have the following result.

**Theorem 2.1.** 1. Assume (C1)-(C2). If  $\nu$  belongs to  $\bar{\mathcal{F}}$  (see (18)), then,  $\varphi_y(\nu)$  and  $\psi(\nu)$  both belong to  $\bar{\mathcal{F}}$ . More precisely, if  $\nu = \nu_{\theta, \alpha}$  then,

$$\varphi_y(\nu) = \nu_{T_y(\theta), a_y(\theta, \alpha)} \quad \text{and} \quad \psi(\nu) = \nu_{\tau(\theta), b(\theta, \alpha)}$$

where  $T_y(\theta), \tau(\theta)$  are defined in Conditions (C1)-(C2) and the mixture parameters  $a_y(\theta, \alpha)$  and  $b(\theta, \alpha)$  are given in formulae (29) and (31) or (33).

2. Assume (C1)-(C2-f). If  $\nu$  belongs to  $\bar{\mathcal{F}}_f$  (see (22)), then,  $\varphi_y(\nu)$  and  $\psi(\nu)$  both belong to  $\bar{\mathcal{F}}_f$ .

*Proof.* Consider first the up-dating operator  $\varphi_y$  (see (4)-(7)-(8)). Let  $\nu = \nu_{\theta, \alpha} = \sum_{i \geq 0} \alpha_i \nu_{\theta}^i$ . Then (see (8)),

$$p_{\nu}(y) = \sum_{i \geq 0} \alpha_i p_{\nu_{\theta}^i}(y), \quad (24)$$

with, for all  $i$ ,

$$p_{\nu_{\theta}^i}(y) = \int_{\mathcal{X}} \nu_{\theta}^i(dx) f_x(y). \quad (25)$$

We have

$$f_x(y) \nu(dx) = \sum_{i \geq 0} \alpha_i \nu_{\theta}^i(dx) f_x(y). \quad (26)$$

Using (C1), since the mapping  $t_y$  is one-to-one, we get

$$f_x(y) \nu(dx) = \sum_{i \geq 0} \alpha_i p_{\nu_{\theta}^i}(y) \nu_{T_y(\theta)}^{t_y(i)}(dx) = \sum_{j \geq 0} \alpha_{t_y^{-1}(j)} p_{\nu_{\theta}^{t_y^{-1}(j)}}(y) \nu_{T_y(\theta)}^j(dx) \quad (27)$$



Hence,

$$\varphi_y(\nu_{\theta,\alpha}) = \nu_{T_y(\theta), a_y(\theta,\alpha)}, \quad (28)$$

where the parameter  $T_y(\theta)$  is defined in (C1) and the mixture coefficient  $a_y(\theta, \alpha)$  is given by (see (24)):

$$a_y(\theta, \alpha)_j = \frac{\alpha_{t_y^{-1}(j)} p_{\nu_{\theta}^{t_y^{-1}(j)}}(y)}{p_{\nu}(y)}. \quad (29)$$

The operator  $\varphi_y$  on  $\tilde{\mathcal{F}}$  can be expressed in terms of the parameters by the mapping:

$$(\theta, \alpha) \in \Theta \times S \rightarrow (T_y(\theta), a_y(\theta, \alpha)) \in \Theta \times S. \quad (30)$$

By (29), the mixture parameter  $a_y(\theta, \alpha)$  has finite length when  $\alpha$  has finite length.

Consider now the prediction operator  $\psi$  (see (9)). By linearity and (C2), we get

$$\psi(\nu) = \sum_{i \geq 0} \alpha_i \psi(\nu_{\theta}^i) = \sum_{j \geq 0} b(\theta, \alpha)_j \nu_{\tau(\theta)}^j = \nu_{\tau(\theta), b(\theta, \alpha)},$$

where  $\tau(\theta)$  defined in (C2) and the new mixture parameter is obtained by interchanging sums and is given by

$$b(\theta, \alpha)_j = \sum_{i \geq 0} \alpha_i \alpha_j^{(i, \theta)}. \quad (31)$$

The prediction operator  $\psi$  is therefore now given defined by the mapping:

$$(\theta, \alpha) \rightarrow (\tau(\theta), b(\theta, \alpha)). \quad (32)$$

Now, if  $l(\alpha) = p$ , and (C2-f) holds, then

$$b(\theta, \alpha)_j = \sum_{i=0}^p \alpha_i \alpha_j^{(i, \theta)} 1_{(j \leq L(i))} = \sum_{i \geq L^{-1}(j), i \leq p} \alpha_i \alpha_j^{(i, \theta)} \quad (33)$$

where  $L^{-1}(j) = \inf\{i; L(i) \geq j\}$ . Since  $\alpha_i = 0$  for  $i > p$ ,  $b(\theta, \alpha)_j = 0$  as soon as  $L^{-1}(j) > p$ .  $\square$

**Remark.**

1. It is worth noting that our conditions imply that the parameters  $T_y(\theta), \tau(\theta)$  only depend on  $\theta$  whereas  $a_y(\theta, \alpha), b(\theta, \alpha)$  depend on  $(\theta, \alpha)$ .
2. For  $x \in \mathcal{X}$ , it is immediate to check that  $\varphi_y(\delta_x) = \delta_x$ . By (C3),  $\psi(\delta_x) = \nu_{\theta_0, \alpha^0(x)}$ . So the algorithm starting with a deterministic initial condition evolves in  $\tilde{\mathcal{F}}$ .

### 3 The Kalman filter with non Gaussian initial condition.

Our first example is based on the classical and simplest standard one-dimensional Kalman filter. It is well known (see *e.g.* Makowski (1986)) that, whatever the initial distribution for the Kalman filter, it is possible to compute the prediction and exact filters. We illustrate this property through a special family of initial distributions. Let us recall the model. The observation equation is given by

$$y_n = h x_n + \gamma w_n, \quad (34)$$

with  $h, \gamma$  constants ( $\gamma > 0$ ),  $(w_n)$  a standard one-dimensional Gaussian white noise. And for the signal

$$x_n = a x_{n-1} + \beta \eta_n, \quad (35)$$

with  $a, \beta$  constants ( $\beta > 0$ ),  $(\eta_n)$  a standard one-dimensional Gaussian white noise. The sequences  $(x_n)$  and  $(w_n)$  are assumed to be independent. Now, the conditional distribution of  $y_n$  given  $x_n = x$  is

$$f_x(y)dy = \mathcal{N}(hx, \gamma^2). \quad (36)$$

And the transition operator of  $(x_n)$  is

$$P(x, dx') = p(x, x')dx' = \mathcal{N}(ax, \beta^2). \quad (37)$$

We introduce below a class of non Gaussian distributions and show that our conditions (C1)-(C3) hold for this class. Therefore, (5)-(6) can be explicitly computed. Before doing this, we recall the classical case.

#### 3.1 The standard Kalman filter.

It is well-known that if the initial distribution is Gaussian (or deterministic) then, for all  $n$ , the distributions (5) and (6) are Gaussian. Let us denote by  $\mathcal{G} = \{\mathcal{N}(m, \sigma^2), m \in \mathbb{R}, \sigma^2 > 0\}$  the class of Gaussian distributions. The up-dating and prediction operators are from  $\mathcal{G}$  onto  $\mathcal{G}$ . And, some classical and elementary computations yield:

- The up-dating step is:  $\nu = \mathcal{N}(m, \sigma^2) \rightarrow \varphi_y(\nu) = \mathcal{N}(\hat{m}, \hat{\sigma}^2)$ , with

$$\hat{m} = \frac{m\gamma^2 + hy\sigma^2}{\gamma^2 + h^2\sigma^2}, \quad \hat{\sigma}^2 = \frac{\sigma^2\gamma^2}{\gamma^2 + h^2\sigma^2}. \quad (38)$$

- The marginal distribution is:  $p_\nu(y)dy = \mathcal{N}(hm, \gamma^2 + h^2\sigma^2)$
- The prediction step is:  $\nu = \mathcal{N}(m, \sigma^2) \rightarrow \psi(\nu) = \mathcal{N}(\bar{m}, \bar{\sigma}^2)$  with

$$\bar{m} = a m, \quad \bar{\sigma}^2 = \beta^2 + a^2\sigma^2. \quad (39)$$

The formulae above hold true when  $\sigma = 0$  allowing to include the case of Dirac measures. Note that, for all  $x$ ,  $P(x, dx') = \mathcal{N}(ax, \beta^2)$  also belongs to  $\mathcal{G}$ .

### 3.2 An extended Kalman filter.

Now, we enlarge the class of Gaussian distributions using new distributions and mixtures. Consider three parameters  $\mu, m, \sigma^2$  with  $\mu, m \in \mathbb{R}$  and  $\sigma^2 > 0$ . For  $i = 0$ , set

$$\nu_{(0,m,\sigma^2)}^0(dx) = \mathcal{N}(m, \sigma^2). \quad (40)$$

and  $\mathcal{F}^0 = \mathcal{G}$ . For  $i \geq 1$ , set

$$\nu_{(\mu,m,\sigma^2)}^i(dx) = \frac{(x + \mu)^{2i}}{C_{2i}(m + \mu; \sigma^2)} \nu_{(0,m,\sigma^2)}^0(dx), \quad (41)$$

where the normalizing constant is given by

$$C_{2i}(m + \mu; \sigma^2) = \mathbb{E}((\sigma X + \mu + m)^{2i}), \quad (42)$$

for  $X$  a standard Gaussian random variable. Let denote by

$$C_{2i} = \mathbb{E}(X^{2i}) = \frac{(2i)!}{2^i i!}, \quad (43)$$

the  $2i$ -th moment of  $X$ . Then, for  $0 \leq k \leq i$ , some elementary computations yield:

$$\binom{2i}{2k} C_{2(i-k)} = \frac{C_{2i}}{C_{2k}} \binom{i}{k}. \quad (44)$$

We deduce

$$C_{2i}(m + \mu; \sigma^2) = \sum_{k=0}^i (m + \mu)^{2k} \sigma^{2(i-k)} \frac{C_{2i}}{C_{2k}} \binom{i}{k}. \quad (45)$$

Set  $\mathcal{F}^i = \{\nu_{(\mu,m,\sigma^2)}^i, (\mu, m, \sigma^2) \in \mathbb{R} \times \mathbb{R} \times (0, +\infty)\}$ . Define  $\mathcal{F} = \bigcup_{i \geq 0} \mathcal{F}^i$ .

In the Appendix, we study some elementary properties of these distributions.

The class  $\bar{\mathcal{F}}$  is defined as in (18). All distributions in  $\bar{\mathcal{F}}$  have density with respect to a Gaussian law.

Now, we check conditions (C1)-(C2-f)-(C3). Condition (C3) evidently holds (see (37)) since  $\bar{\mathcal{F}}$  contains all Gaussian distributions. By the following proposition, condition (C2-f) holds.

**Proposition 3.1.** *Consider the model given by (34)-(35)-(36)-(37).*

1. For  $i = 0$ , and  $\nu = \mathcal{N}(m, \sigma^2)$ ,  $\varphi_y(\nu) = \mathcal{N}(\hat{m}, \hat{\sigma}^2)$  with  $\hat{m}, \hat{\sigma}^2$  given in (38).
2. For  $i \geq 1$ , and  $\nu = \nu_{(\mu,m,\sigma^2)}^i(dx)$  in  $\mathcal{F}^i$ ,  $\varphi_y(\nu) = \nu_{(\mu,\hat{m},\hat{\sigma}^2)}^i$ .

Condition (C1) holds with  $t_y(i) = i$ , for all  $i \geq 0$  and  $T_y(\mu, m, \sigma^2) = (\mu, \hat{m}, \hat{\sigma}^2)$  (where for  $i = 0$ ,  $\mu = 0$ ).

The proof is obtained in the same way as for the up-dating step for the classical Kalman filter (see the Appendix). Now, looking at formulae (8) and (10), since  $f_x(y)$  and  $p(x, x')$  in this model are both Gaussian kernels, the computation of marginal distributions and the checking of (C2-f) are identical up to a change of notations. The results are given in the following proposition.

**Proposition 3.2.** *Let  $\nu = \nu_{(\mu, m, \sigma^2)}^i(dx)$  belong to  $\mathcal{F}^i$ .*

1. *Then*

$$\psi(\nu) = \sum_{k=0}^i \bar{\alpha}_k^{(i)} \nu_{(\bar{\mu}, \bar{m}, \bar{\sigma}^2)}^k, \quad (46)$$

*with  $\bar{m}, \bar{\sigma}^2$  given in (39),*

$$\bar{\mu} = \frac{m\beta^2 + \mu\bar{\sigma}^2}{a\sigma^2}, \quad (47)$$

*and for  $k = 0, \dots, i$ ,*

$$\bar{\alpha}_k^{(i)} = \binom{i}{k} \frac{\beta^{2(i-k)}}{B_i} \sum_{j=0}^k \binom{k}{j} \frac{(\mu + m)^{2j} a^{2(k-j)} \sigma^{2(k-j)}}{C_{2j} \sigma^{2j} \bar{\sigma}^{2(i-j)}}, \quad (48)$$

*with*

$$B_i = \sum_{k=0}^i \binom{i}{k} \frac{(\mu + m)^{2k}}{C_{2k} \sigma^{2k}} \quad (49)$$

*( $\sum_{k=0}^i \bar{\alpha}_k^{(i)} = 1$ ). Hence, Condition (C2-f) holds with  $\tau(\mu, m, \sigma^2) = (\bar{\mu}, \bar{m}, \bar{\sigma}^2)$ , for  $k = 0, \dots, i$ ,  $\alpha_k^{(i, \mu, m, \sigma^2)} = \bar{\alpha}_k^{(i)}$  and the length of the mixture parameter of  $\psi(\nu)$  for  $\nu \in \mathcal{F}^i$  is  $L(i) = i$ .*

2. *The marginal distribution  $p_\nu(y)dy$  is given by the same formula as (46) with  $(a, \beta^2)$  everywhere replaced by  $(h, \gamma^2)$ .*

The proof is given in the Appendix. Note that, using  $\frac{a^2 \sigma^2}{\bar{\sigma}^2} = 1 - \frac{\beta^2}{\bar{\sigma}^2}$ , we have

$$\bar{m} + \bar{\mu} = a(m + \mu) \left(1 - \frac{\beta^2}{\bar{\sigma}^2}\right)^{-1}.$$

**Remark.** It is worth noting that, in this model, the number of parameters remains fixed along iterations: If the initial condition is specified by parameters  $(\mu, m, \sigma, \alpha)$  with  $l(\alpha) = p$ , *i.e.*  $3 + p + 1$  parameters, then the length of the mixture parameter will always be equal to  $p$  and the number of parameters will remained fixed equal to  $3 + p + 1$ . This is not surprising since the Kalman filter is a finite-dimensional filter, even when the initial condition is non Gaussian (see *e.g.* Makowski (1986)).

## 4 Scale perturbation of a radial Ornstein-Uhlenbeck process.

In this section, we consider multiplicative perturbation models of the form

$$y_n = x_n w_n \quad (50)$$

where  $(w_n)$  is a sequence of i.i.d. positive random variables and  $(x_n)$  is also a positive signal independent of the sequence  $(w_n)$ . The multiplicative structure comes from the field of Finance with the so-called stochastic volatility models. However, in stochastic volatility models, the noises are standard Gaussian variables (see the next section). The advantage of positive signal and noise is that we can interpret the model as a scale perturbation of a positive signal.

Now, we consider a signal which is a discretization of the radial Ornstein-Uhlenbeck process. And, for the noise, we consider positive random variables with a specific distribution and build a class  $\tilde{\mathcal{F}}$  such that conditions (C1)-(C2-f)-(C3) hold.

## 4.1 The signal

We assume that

$$x_n = X_{n\Delta} \quad (51)$$

is a discretization of a continuous time diffusion  $(X_t)$  equal to a radial Ornstein-Uhlenbeck process. We recall its definition and properties.

### 4.1.1 The one-dimensional radial Ornstein-Uhlenbeck process.

Consider the one-dimensional Ornstein-Uhlenbeck process given by:

$$\xi_t = \xi_0 + \int_0^t \theta \xi_s ds + \sigma W_t, \quad (52)$$

where  $(W_t)$  is a standard Brownian motion and the initial variable  $\xi_0$  is independent of  $(W_t)$ . Then,

$$\xi_t = \xi_0 e^{\theta t} + \sigma \int_0^t e^{\theta(t-s)} dW_s. \quad (53)$$

Let  $X_t = |\xi_t|$ . Then, a simple computation shows that the conditional distribution of  $X_t$  given  $\xi_0 = \xi$  only depends on  $x = |\xi|$  so that  $(X_t)$  is a Markov process. We may call it the one-dimensional radial Ornstein-Uhlenbeck process. Let us now give the conditional density of  $x_n = X_{n\Delta}$  given  $x_{n-1} = X_{(n-1)\Delta}$ , *i.e.* the transition density of  $(x_n)$  (see (51)). For  $\Delta > 0$ , we set

$$U_n = \xi_{n\Delta} \quad \text{and} \quad a = e^{\theta\Delta}, \quad \beta^2 = \sigma^2 \frac{e^{2\theta\Delta} - 1}{2\theta}. \quad (54)$$

Then, as can be easily deduced from (53),  $(U_n)$  is a standard AR(1)-process satisfying

$$U_n = aU_{n-1} + \beta\eta_n, \quad n \geq 1, \quad U_0 = \xi_0 \quad (55)$$

where  $(\eta_n, n \geq 1)$  is a sequence of i.i.d. random variables having distribution  $\mathcal{N}(0, 1)$ . Now,  $x_n = |U_n| = X_{n\Delta}$  is a Markov chain having transition density (for positive  $x$ )

$$p^{(1)}(x, x') = (p(x, x') + p(x, -x'))1_{(x' > 0)}, \quad (56)$$

where  $p(u, u')$  is the transition density of  $(U_n)$ , *i.e.*

$$p(u, u') = \frac{1}{\beta\sqrt{2\pi}} \exp\left(-\frac{(u' - au)^2}{2\beta^2}\right). \quad (57)$$

A simple computation shows that (with  $x > 0$ )

$$p^{(1)}(x, x') = 1_{(x' > 0)} \frac{2}{\beta\sqrt{2\pi}} \exp\left(-\frac{x'^2}{2\beta^2}\right) \exp\left(-\frac{a^2 x^2}{2\beta^2}\right) \left(\cosh\left(\frac{axx'}{\beta^2}\right)\right). \quad (58)$$

Now, using the series expansion of  $\cosh$ , we obtain a representation of this transition density as the following mixture of distributions:

$$p^{(1)}(x, x') = 1_{(x' > 0)} \frac{2}{\beta\sqrt{2\pi}} \exp\left(-\frac{x'^2}{2\beta^2}\right) \sum_{i \geq 0} \alpha_i(x) \frac{x'^{2i}}{C_{2i}\beta^{2i}}, \quad (59)$$

where  $C_{2i} = \frac{(2i)!}{2^i i!}$  is the  $2i$ -th moment of a standard Gaussian variable (see (43)), and for  $i \geq 0$ ,

$$\alpha_i(x) = \exp\left(-\frac{a^2 x^2}{2\beta^2}\right) \left(\frac{a^2 x^2}{2\beta^2}\right)^i \frac{1}{i!}. \quad (60)$$

For  $\theta < 0$  ( $0 < a < 1$ ), the process  $(x_n)$  has a stationary distribution given by

$$\pi^{(1)}(dx) = 1_{(x > 0)} \frac{2}{\rho\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\rho^2}\right) dx \quad (61)$$

with (see (54))

$$\rho^2 = \frac{\beta^2}{1 - a^2} = \frac{\sigma^2}{2|\theta|}. \quad (62)$$

These formulae will be the useful tool for the construction of the class  $\bar{\mathcal{F}}$  below.

#### 4.1.2 The $\delta$ -dimensional radial Ornstein-Uhlenbeck process.

For  $\delta > 1$ , consider the stochastic differential equation

$$dX_t = (\theta X_t + \frac{\sigma^2(\delta - 1)}{2X_t})dt + \sigma d\beta_t, \quad X_0 = \eta. \quad (63)$$

where  $(\beta_t)$  is a standard Brownian motion and  $\eta$  is a random variable independent of  $(\beta_t)$ . The values  $\theta, \sigma, \delta$  are constant parameters. This process is called a radial Ornstein-Uhlenbeck process. This is due to the fact that, when  $\delta$  is an integer greater than 1, then  $X_t$  is the Euclidian norm of a  $\delta$ -dimensional vector  $(\xi_t^1, \dots, \xi_t^\delta)$  whose components are i.i.d. Ornstein-Uhlenbeck processes satisfying:

$$d\xi_t^j = \theta \xi_t^j dt + \sigma dW_t^j.$$

Moreover, when  $(X_t)$  is given by (63), the process  $R_t = X_t^2$  is the classical Cox-Ingersoll-Ross diffusion model given by:

$$dR_t = (2\theta R_t + \delta\sigma^2)dt + 2\sigma R_t^{1/2} d\beta_t. \quad (64)$$

The processes  $(X_t)$  and  $(R_t)$  have explicit transition probabilities with densities with respect to the Lebesgue measure on  $(0, +\infty)$ . There are closed-form formulae for the transition densities when  $\delta$  is an odd integer. Otherwise, they depend on Bessel functions and have explicit developments as sums of series. As above, we set  $x_n = X_{n\Delta}$  and give the expression of the transition operator of this Markov chain using the notations (54). For  $\alpha \geq 0$ , let us set

$$C_\alpha = \mathbb{E}|X|^\alpha, \quad (65)$$

for  $X$  a standard Gaussian random variable.

**Proposition 4.1.** *1. Assume  $\delta > 1$ . Then the transition density of the  $\delta$ -dimensional radial Ornstein-Uhlenbeck process (see (63)) is equal to (with  $x > 0$ ):*

$$p^{(\delta)}(x, x') = 1_{(x' > 0)} \frac{2}{\beta\sqrt{2\pi}} \exp\left(-\frac{x'^2}{2\beta^2}\right) \sum_{k \geq 0} \alpha_k(x) \frac{x'^{\delta-1+2k}}{C_{\delta-1+2k} \beta^{\delta-1+2k}}, \quad (66)$$

where the mixture coefficients are given by (60) and the couple  $(a, \beta^2)$  is linked with the original parameters  $(\theta, \sigma^2)$  through relations (54).

2. Assume that  $\delta = 2n + 1$  with  $n \geq 1$  an integer. Define the operator  $\mathcal{T}$ , acting on functions  $f \in C^1((0, +\infty), \mathbb{R})$ , by  $\mathcal{T}(f)(x) = \frac{f'(x)}{x}$ . Then, the transition density of  $(x_n = X_{n\Delta})$ , where  $(X_t)$  is the  $2n + 1$ -dimensional radial Ornstein-Uhlenbeck process is equal to (with  $x > 0$ )

$$p^{(2n+1)}(x, x') = 1_{(x' > 0)} \frac{2}{\beta\sqrt{2\pi}} \exp\left(-\frac{x'^2}{2\beta^2}\right) \exp\left(-\frac{a^2 x^2}{2\beta^2}\right) \frac{x'^{2n}}{\beta^{2n}} [\mathcal{T}^n(\cosh)(z)]_{z=\frac{axx'}{\beta^2}} \quad (67)$$

where  $\mathcal{T}^n = \mathcal{T} \circ \mathcal{T} \dots \circ \mathcal{T}$  is the  $n$ -th iterate of  $\mathcal{T}$ .

Moreover, the above formulae also hold for  $\delta = 1$  ( $n = 0$ ) (see (58)-(59)) with the convention that  $\mathcal{T}^0(f) = f$ .

When  $\theta < 0$  ( $0 < a < 1$ ), the process  $(X_t)$  and the Markov chain  $(x_n)$  have a stationary distribution equal to (see (54)-(62))

$$\pi^{(\delta)}(dx) = 1_{(x > 0)} \frac{2}{\rho\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\rho^2}\right) \frac{x^{\delta-1}}{\rho^{\delta-1} C_{\delta-1}} dx. \quad (68)$$

Details are given in the Appendix.

## 4.2 Distribution of the noise

Assume that, for all  $n$ ,  $w_n$  has the distribution of  $\Gamma^{-1/2}$  where  $\Gamma$  has an exponential distribution with parameter  $\lambda > 0$ . Then, for all positive  $x$ , the distribution of  $Y = x w_1$  is given by:

$$F_x(dy) = f_x(y) dy, \quad \text{with} \quad f_x(y) = \frac{2\lambda x^2}{y^3} \exp\left(-\frac{\lambda x^2}{y^2}\right) 1_{(0, \infty)}(y). \quad (69)$$

It is worth noting that the distribution of  $w_1$  ( $F_1(dy)$ ) satisfies:

$$\mathbb{E}(|\log w_1|) < \infty \quad \text{and} \quad \mathbb{E}(w_1^r) < \infty \quad \text{if and only if} \quad r < 2. \quad (70)$$

Instead of an exponential distribution, we could take a Gamma distribution with integer index.

### 4.3 The class of distributions.

First, for  $i \geq 0$ , we set, for  $\delta \geq 1$ ,

$$g_i^{(\delta)}(x) = 1_{(x>0)} \frac{2}{(2\pi)^{1/2}} \frac{x^{\delta-1+2i}}{C_{\delta-1+2i}} \exp\left(-\frac{x^2}{2}\right), \quad (71)$$

where  $C_\alpha$  is defined in (65). Thus,  $g_i^{(\delta)}$  is a probability density on  $(0, +\infty)$ . Then, for  $\sigma > 0$ , we set

$$\nu_\sigma^{i,(\delta)}(dx) = \frac{1}{\sigma} g_i^{(\delta)}\left(\frac{x}{\sigma}\right) dx = 1_{(x>0)} \frac{2}{\sigma(2\pi)^{1/2}} \frac{x^{\delta-1+2i}}{C_{\delta-1+2i} \sigma^{\delta-1+2i}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx. \quad (72)$$

For each  $i$ , the distribution  $\nu_\sigma^{i,(\delta)}$  is equal to the distribution of  $\sqrt{G_i^{(\delta)}}$  where  $G_i^{(\delta)}$  is Gamma with parameters  $(i + \delta/2, 1/2\sigma^2)$ . This Gamma distribution is identical to a  $\sigma^2 \chi^2(\delta + 2i)$  (with non integer parameter). Hence, as  $i$  increases, the distributions  $(\nu_\sigma^{i,(\delta)}, i \geq 0)$  are stochastically increasing. Let us now define

$$\mathcal{F}^{i,(\delta)} = \{\nu_\sigma^{i,(\delta)}; \sigma > 0\}, \quad \mathcal{F}^{(\delta)} = \cup_{i \geq 0} \mathcal{F}^{i,(\delta)}. \quad (73)$$

And,

$$\bar{\mathcal{F}}^{(\delta)} = \{\nu = \nu_{\sigma, \alpha} = \sum_{i \geq 0} \alpha_i \nu_\sigma^{i,(\delta)}, \alpha = (\alpha_i) \in S, \nu_\sigma^{i,(\delta)} \in \mathcal{F}^{i,(\delta)}, i \geq 0\}. \quad (74)$$

The law of  $|X|$  for  $X$  a Gaussian variable with mean  $m$  and variance  $\sigma^2$  has the density

$$g(x) = 1_{(x>0)} \frac{2}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \exp\left(-\frac{m^2}{2\sigma^2}\right) \left(\cosh\left(\frac{mx}{\sigma^2}\right)\right). \quad (75)$$

A Taylor series development of the cosh yields a distribution of the class  $\bar{\mathcal{F}}^{(1)}$  with mixture parameter

$$\alpha_i(m, \sigma) = \exp\left(-\frac{m^2}{2\sigma^2}\right) \frac{1}{i!} \left(\frac{m^2}{2\sigma^2}\right)^i, \quad i \geq 0. \quad (76)$$

All distributions in  $\bar{\mathcal{F}}^{(\delta)}$  have a density with respect to a  $\nu_\sigma^{0,(\delta)}(dx)$  for some positive  $\sigma$  (see (72)). And this density is expressed as an entire series of even powers. The transition density (58) and the stationary density (61) belong to  $\bar{\mathcal{F}}^{(1)}$ . The transition density (66) and the stationary density (68) belong to  $\bar{\mathcal{F}}^{(\delta)}$ . Thus, our condition (C3) holds for the model defined by (51) and (63).

### 4.4 Up-dating, marginal and prediction operators.

In this section, we show that the filtering and prediction algorithms evolve in the class  $\bar{\mathcal{F}}^{(\delta)}$  when the signal is a discrete regular sampling of the  $\delta$ -dimensional radial Ornstein-Uhlenbeck process. For this, it is enough to check conditions (C1)-(C2)-(C3).



We have already noted that condition (C3) holds. Actually, the form of the transition density (66) and of the stationary distribution (68) (when it exists) indicates how to define the class  $\mathcal{F}^{(\delta)}$ .

**Proposition 4.2.** *Let  $y > 0$ , and consider the up-dating operator  $\varphi_y$  (see (7)) corresponding to  $f_x(y)$  given in (69). For  $i \geq 0$  and  $\sigma > 0$ , with  $\nu_\sigma^{i,(\delta)}$  defined in (72),*

$$\varphi(\nu_\sigma^{i,(\delta)}) = \nu_{T_y(\sigma)}^{i+1,(\delta)}, \quad \text{with} \quad T_y(\sigma) = \frac{\sigma y}{(y^2 + 2\lambda\sigma^2)^{1/2}}.$$

Thus,  $t_y(i) = t(i) = i + 1$  and (C1) holds for  $(\varphi_y, \mathcal{F}^{(\delta)})$ .

*Proof.* We have

$$f_x(y)g_i^{(\delta)}(x/\sigma) \propto x^{\delta-1+2(i+1)} \exp\left(-\left(\frac{2\lambda}{y^2} + \frac{1}{\sigma^2}\right)\frac{x^2}{2}\right).$$

Hence, we may define

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

Moreover, we have  $t_y(i) = t(i) = i + 1$ : if  $\nu \in \mathcal{F}^{i,(\delta)}$ , then,  $\varphi_y(\nu) \in \mathcal{F}^{i+1,(\delta)}$ . So, we get the result.  $\square$

We also give the marginal distribution.

**Proposition 4.3.** *For  $i \geq 0$  and  $\sigma > 0$ , with  $\nu_\sigma^{i,(\delta)}$  defined in (72), the marginal density (see (8)) is equal to*

$$p_{\nu_\sigma^{i,(\delta)}}(y) = 1_{y>0} \frac{2}{\lambda^{1/2}\sigma} p_i^{(\delta)}\left(\frac{y}{\lambda^{1/2}\sigma}\right) \quad (77)$$

with

$$p_i^{(\delta)}(y) = \frac{(\delta + 2i)y^{\delta-1+2i}}{(y^2 + 2)^{i+1+\delta/2}}. \quad (78)$$

*Proof.* Using  $C_{\delta-1+2(i+1)} = (\delta + 2i)C_{\delta-1+2i}$ , we get

$$\int_{\mathbb{R}^+} \frac{1}{\sigma} g_i^{(\delta)}\left(\frac{x}{\sigma}\right) f_x(y) dx = \frac{\lambda\sigma^2(\delta + 2i)y^{\delta-1+2i}}{(y^2 + 2\lambda\sigma^2)^{i+1+\delta/2}} = \lambda^{-1/2}\sigma^{-1} p_i^{(\delta)}\left(\frac{y}{\lambda^{1/2}\sigma}\right).$$

$\square$

**Remark.** Proposition 4.3 allows to obtain the density of  $(y_0, y_1, \dots, y_n)$ , i.e. the exact likelihood based on this observation. Indeed, this joint density is obtained as the product of the conditional densities of  $y_i$  given  $y_{i-1}, \dots, y_1$ . These are computed as marginal densities (see (16)).

**Proposition 4.4.** Consider the transition operator  $P^{(\delta)}$  with transition density (66). Let  $\nu = \nu_\sigma^{i,(\delta)}$  be given by (72) with  $i \geq 0$ , i.e.  $\nu_\sigma^{i,(\delta)} \in \mathcal{F}^{i,(\delta)}$ . Set  $\nu P^{(\delta)} = \psi^{(\delta)}(\nu)$ . Then,

$$\psi^{(\delta)}(\nu_\sigma^{i,(\delta)}) = \sum_{k=0}^i \alpha_k^{(i,\sigma)} \nu_{\tau(\sigma)}^{k,(\delta)} \quad (79)$$

with

$$\tau^2(\sigma) = \beta^2 + a^2 \sigma^2, \quad (80)$$

and for  $k = 0, 1, \dots, i$ ,

$$\alpha_k^{(i,\sigma)} = \binom{i}{k} \left(1 - \frac{\beta^2}{\tau^2(\sigma)}\right)^k \left(\frac{\beta^2}{\tau^2(\sigma)}\right)^{i-k} \quad (81)$$

Thus,  $L(i) = i$  and  $\psi^{(\delta)}(\nu_\sigma^{i,(\delta)})$  belongs to  $\bar{\mathcal{F}}^{(\delta)}$ . Condition (C2-f) holds for  $(P^{(\delta)}, \bar{\mathcal{F}}^{(\delta)})$ .

*Proof.* We have to compute

$$A = \int_0^\infty (1/\sigma) g_i^{(\delta)}(x/\sigma) p^{(\delta)}(x, x') dx, \quad (82)$$

with  $g_i^{(\delta)}$  given in (71) and  $p^{(\delta)}(x, x')$  given in (66). Let us define  $s^2$  by

$$\frac{1}{s^2} = \frac{a^2}{\beta^2} + \frac{1}{\sigma^2} = \frac{\tau^2(\sigma)}{\beta^2 \sigma^2}. \quad (83)$$

Hence, for all  $k \geq 0$ ,

$$\int_0^\infty 2 x^{2(i+k)+\delta-1} \exp\left(-\frac{a^2 x^2}{2\beta^2}\right) \exp\left(-\frac{x^2}{2\sigma^2}\right) \frac{dx}{(2\pi)^{1/2}} = C_{2(i+k)+\delta-1} s^{2(i+k)+\delta}. \quad (84)$$

Now, using  $s/\beta\sigma = 1/\tau(\sigma)$ ,  $A$  is given as the following expression

$$A = \frac{2}{\tau(\sigma)(2\pi)^{1/2}} \exp\left(-\frac{x'^2}{2\beta^2}\right) \frac{x'^{\delta-1}}{\tau^{\delta-1}(\sigma)} \Sigma, \quad (85)$$

with

$$\Sigma = \sum_{k=0}^\infty \frac{x'^{2k}}{\beta^{2k}} \left(\frac{a^2}{2\beta^2}\right)^k \frac{C_{2(i+k)+\delta-1}}{k! C_{2k+\delta-1} C_{2i+\delta-1}} \frac{s^{2(i+k)}}{\sigma^{2i}}. \quad (86)$$

Using (83) and some computations, we get, for all  $k \geq 0$ ,

$$\left(\frac{a^2}{\beta^2}\right)^k \frac{s^{2(i+k)}}{\sigma^{2i}} = \left(\frac{\beta^2}{\tau^2(\sigma)}\right)^i \left(1 - \frac{\beta^2}{\tau^2(\sigma)}\right)^k. \quad (87)$$

Now, we set

$$\frac{1}{c^2} = \frac{1}{\beta^2} \left(1 - \frac{\beta^2}{\tau^2(\sigma)}\right). \quad (88)$$

This yields

$$\Sigma = \sum_{k=0}^\infty \frac{1}{k!} \left(\frac{x'^2}{2c^2}\right)^k \left(\frac{\beta^2}{\tau^2(\sigma)}\right)^i \frac{C_{2(i+k)+\delta-1}}{C_{2k+\delta-1} C_{2i+\delta-1}}. \quad (89)$$

Now, we use the following lemma whose proof is given in the Appendix.

**Lemma 4.1.** For all integer  $k \geq 0$ , and all  $i \geq n$

$$\begin{aligned} \frac{C'_{2(i+k)+\delta-1}}{C_{2k+\delta-1}C_{2i+\delta-1}} &= \sum_{j=0}^i k(k-1)\dots(k-j+1)a_j \\ &= a_0 + ka_1 + k(k-1)a_2 + \dots + k(k-1)\dots(k-i+1)a_i. \end{aligned} \quad (90)$$

with, for  $j = 0, 1, \dots, i$ ,

$$a_j = \binom{i}{j} \frac{2^j}{C_{2j+\delta-1}},$$

and the coefficient of  $a_0$  is equal to 1.

Now, we transform expression (89) into

$$\Sigma = \sum_{j=0}^i \Sigma_j, \quad (91)$$

where

$$\Sigma_j = a_j \left( \frac{\beta^2}{\tau^2(\sigma)} \right)^i \sum_{k=0}^{\infty} \frac{1}{k!} \left( \frac{x'^2}{2c^2} \right)^k k(k-1)\dots(k-j+1). \quad (92)$$

But,  $k(k-1)\dots(k-j+1) = 0$  for  $k = 0, 1, \dots, j-1$ . So, we get

$$\begin{aligned} \Sigma_j &= a_j \left( \frac{\beta^2}{\tau^2(\sigma)} \right)^i \sum_{k \geq j} \left( \frac{x'^2}{2c^2} \right)^k \frac{1}{(k-j)!} \\ &= \frac{1}{C_{2j+\delta-1}} \left( \frac{\beta^2}{\tau^2(\sigma)} \right)^i \binom{i}{j} \left( \frac{x'^2}{c^2} \right)^j \exp\left(\frac{x'^2}{2c^2}\right). \end{aligned} \quad (93)$$

Now, we compute  $A$  from (85). Using (88), we add the exponents of the exponential terms and after some elementary computations, we obtain

$$A = \frac{2 \mathbf{1}_{(x'>0)}}{\tau(\sigma)(2\pi)^{1/2}} \exp\left(-\frac{x'^2}{2\tau^2(\sigma)}\right) \Sigma' \quad (94)$$

with

$$\Sigma' = \sum_{j=0}^i \Sigma'_j, \quad (95)$$

and

$$\Sigma'_j = \binom{i}{j} \frac{1}{C_{2j+\delta-1}} \left( \frac{x'^{2j+\delta-1}}{\tau^{2j+\delta-1}(\sigma)} \right) \left( \frac{\beta^2}{\tau^2(\sigma)} \right)^{i-j} \left( 1 - \frac{\beta^2}{\tau^2(\sigma)} \right)^j. \quad (96)$$

So the proof is complete.  $\square$

**Remarks.**

1. Here, we have two representations of  $\psi^{(\delta)}(\nu_{\sigma}^{i,(\delta)})$ . One has scale parameter  $\beta$  and a mixture parameter with infinite length. The second has scale parameter  $\tau(\sigma)$  and a finite length mixture parameter. The latter appears as a minimal representation of this distribution in a sense that we try to clarify (work in progress).

2. By Propositions 4.2 and 4.4, we see that both the exact filter and the prediction filter evolve in the extended class  $\bar{\mathcal{F}}^{(\delta)}$ . If the initial distribution of the signal is in the subclass  $\mathcal{F}_f^{(\delta)}$  (*e.g.* if the signal is in stationary regime), it has a mixture coefficient with finite length (see (74)). Then, the number of mixture components grows of a unit at each iteration but remains finite. However, the numerical simulations that we have done in Genon-Catalot and Kessler (2004) show that there are only two or three significantly non nul mixture coefficients. Some stability results are also obtained that may be extended to the model investigated here.

## 5 Stochastic volatility type models.

We first draw some immediate consequences of the previous section. Then, we introduce some new type of stochastic volatility models.

### 5.1 Scale perturbation of a Cox-Ingersoll-Ross diffusion process.

Consider now the model obtained by taking squares of the previous one. Set

$$z_n = y_n^2 = r_n v_n, \quad (97)$$

with  $r_n = x_n^2$ ,  $v_n = w_n^2$  and  $(x_n, w_n)$  as in the previous section. Then  $r_n = R_{n\Delta}$  is a discrete sampling of the Cox-Ingersoll-Ross diffusion model (64). It is a Markov chain with transition (see (66))

$$\begin{aligned} q^{(\delta)}(r, r') &= (1/2)r'^{-\frac{1}{2}} p^{(\delta)}(r^{1/2}, r'^{1/2}) \\ &= 1_{(r'>0)} \frac{1}{\beta\sqrt{2\pi}} \exp\left(-\frac{r'}{2\beta^2}\right) \sum_{k \geq 0} \alpha_k(r^{\frac{1}{2}}) \frac{(r')^{k-1+\frac{\delta}{2}}}{C_{2k+\delta-1}\beta^{2k+\delta-1}}. \end{aligned} \quad (98)$$

(see (60) for  $(\alpha_k(x))$ ). This is now a mixture of Gamma distributions with parameters  $(k + \frac{\delta}{2}, \frac{1}{2\beta^2})$ .

The distribution of the noise  $(v_n)$  is now inverse exponential. And the class of distributions is composed with mixtures of Gamma distributions with parameters  $(i + \frac{\delta}{2}, \frac{1}{2\sigma^2})$ , for  $i \geq 0$ . The filtering and prediction algorithm can be explicitly expressed with the same formulae after some simple changes  $z = y^2$  for the observations and the change of variables  $x = r^{\frac{1}{2}}$  for the distributions.

Note that another computable filter is obtained by setting

$$z'_n = \frac{1}{z_n} = \frac{1}{r_n} v'_n, \quad v'_n = \frac{1}{v_n}. \quad (99)$$

### 5.2 Stochastic volatility type models.

The above considerations lead to some new type of stochastic volatility models. Indeed, stochastic volatility models usually postulate that the observed price process  $(S_n)$  of an

asset is such that

$$Z_n = \log \frac{S_{n+1}}{S_n} = \sqrt{V_n} \varepsilon_n, \quad (100)$$

where  $(V_n)$  is a positive Markov chain (the unobserved volatility),  $(\varepsilon_n)$  is a sequence of i.i.d. standard Gaussian variables, the two sequences being independent. We do not know explicit filters for such stochastic volatility models when the signal is a discrete sampling of a diffusion process.

Now, taking squares in (100), we get that  $Z_n^2 = V_n \varepsilon_n^2$ , where  $\varepsilon_n^2$  is distributed as a  $\chi^2(1) = G(1/2, 1/2)$ . Our previous study suggests to replace the  $G(1/2, 1/2)$  distribution by a  $G(1, \lambda)$  (possibly a  $G(k, \lambda)$  with  $k$  integer). More precisely, the following stochastic volatility type models will provide explicit filters through a symetrization device. Consider

$$Z'_n = \sqrt{r_n} \varepsilon'_n, \quad \text{or} \quad Z''_n = \frac{1}{\sqrt{r_n}} \varepsilon''_n, \quad (101)$$

with  $r_n = R_{n\Delta}$  a discrete sampling of a Cox-Ingersoll-Ross diffusion. For the noises, consider a symmetric Bernoulli variable  $\varepsilon \pm 1$  with probability  $1/2$ , independent of a random variable  $\Gamma$  having distribution  $G(1, \lambda)$  (exponential distribution). Then, assume that  $\varepsilon'_n$  is distributed as  $\frac{\varepsilon}{\sqrt{\Gamma}}$  and that  $\varepsilon''_n$  is distributed as  $\varepsilon \sqrt{\Gamma}$ . Then, the two models of filtering given in (101) can be solved explicitly.

## 6 A discretized Cox-Ingersoll-Ross diffusion and conditionally Poisson observations.

Models which have no representation as  $y_n = H(x_n, w_n)$ , for some simple function  $H$ , are also of interest. These models are completely specified by the conditional distribution (4) and the transition operator of the hidden Markov chain. We investigate below such an example. Suppose that the couple (signal, observation) is defined as follows. The signal is the process  $(r_n)$  obtained as above from a discrete sampling of the square-root model (64). Now, the observation  $y_n$  is such that, given  $r_n = r$ ,  $y_n$  has a Poisson distribution with parameter  $\lambda r$ , *i.e.*

$$\mathbb{P}(y_n = y | r_n = r) = f_r(y) = \exp(-\lambda r) \frac{(\lambda r)^y}{y!}, \quad y \in \mathbb{N}. \quad (102)$$

We can check our conditions with the class of distributions fitted with the signal  $(r_n)$ , *i.e.* the class of mixtures of Gamma distributions  $G(i + \frac{\delta}{2}, \frac{1}{2\sigma^2})$  with parameters  $(i + \frac{\delta}{2}, \frac{1}{2\sigma^2})$ , for  $i \geq 0$ . Only (C1) needs to be checked. Let us set for this Gamma density

$$\gamma_\sigma^{i,(\delta)}(r) = 1_{(r>0)} \frac{1}{\beta \sqrt{2\pi}} \exp\left(-\frac{r}{2\sigma^2}\right) \frac{r^{i-1+\frac{\delta}{2}}}{C_{2i+\delta-1} \sigma^{2i+\delta-1}}. \quad (103)$$

Now,

$$f_r(y) \gamma_\sigma^i(r) \propto \exp\left[-\left(\lambda + \frac{1}{2\sigma^2}\right)r\right] r^{(y+i-1+\frac{\delta}{2})}. \quad (104)$$

This is again a Gamma distribution of the same type. And we get

$$t_y(i) = y + i, \quad T_y(\sigma) = \left(2\lambda + \frac{1}{\sigma^2}\right)^{-\frac{1}{2}}. \quad (105)$$

The filtering and prediction algorithm will evolve in the family of mixtures of Gamma distributions with tail index  $i + (\delta/2)$ ,  $i \geq 0$ . The marginal distributions can be explicitly computed.

## 7 Concluding remarks.

This work has to be completed by numerical simulations. In Genon-Catalot and Kessler (2004), the model corresponding to a one-dimensional radial Ornstein-Uhlenbeck process is studied and implemented. The numerical results show that the number of significantly non nul mixture coefficients is less than 2 or 3. Theoretical properties linked with the stability of the filters are established in this paper which may be extended to the models of Section 4.

The above results may be extended to the case of a non time-homogeneous signal. For instance, it is possible to consider a non regular discrete sampling of the underlying diffusion model (*i.e.* to consider  $x_n = X_{t_n}$  with  $0 < t_1 < \dots < t_n < \dots$ ). It is also possible to consider non time-homogeneous conditional distributions of the observation given the signal. Note also that the signal may or may not be ergodic.

**Acknowledgments.** The authors wish to thank Wolfgang Runggaldier and Pavel Chigansky for helpful discussions and references.

## References

- [1] Chaleyat-Maurel M. and Michel D. (1984). Des résultats de non existence de filtre de dimension finie. *Stochastics* **13** (1-2), 83-102.
- [2] Del Moral P. and Guionnet A. (2001). On the stability of interacting processes with applications to filtering and genetic algorithms. *Ann. Inst. H. Poincaré, Probab. et Stat.* **37**, (2), 155-194.
- [3] Di Masi G.B., Runggaldier W.J. and Barozzi B. (1983). Generalized finite-dimensional filters in discrete time. *R.S. Bucy and J.M.F. Moura (eds), Nonlinear Stochastic Problems*, 267-277.
- [4] Douc P. and Matias L. (2001). Asymptotics of the maximum likelihood estimator for general hidden Markov models. *Bernoulli* **7**(3), 381-420.
- [5] Ferrante M. and Vidoni P. (1998). Finite dimensional filters for non linear stochastic difference equations with multiplicative noises. *Stoch. Proc. Applic.* **77**, 69-81.
- [6] Genon-Catalot V. (2003). A non linear explicit filter. *Statist. and Prob. letters* **61**, 145-154.
- [7] Genon-Catalot V., Jeantheau T. and Larédo C. (2000). Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli* **6** (6), 1051-1079.

- [8] Genon-Catalot V. and Kessler M. (2004). Random scale perturbation of an AR(1) process and its properties as a nonlinear explicit filter. *Bernoulli* **(10)** (4), 701-720.
- [9] Karlin S. and Taylor H.M. (1981). *A Second Course in Stochastic Processes*. Academic Press.
- [10] Makowski A.M. (1986). Filtering for partially observed linear systems with non-Gaussian initial conditions. *Stochastics* **16**, 1-24.
- [11] Nikiforov A. and Ouvarov V. (1983). *Fonctions spéciales de la physique mathématique*. Editions Mir, Moscou.
- [12] Runggaldier W. and Spizzichino F. (2001). Sufficient conditions for finite dimensionality of filters in discrete time: A Laplace transform-based approach. *Bernoulli* **7** (2), 211-221.
- [13] Sawitzki G. (1981). Finite dimensional filter systems in discrete time. *Stochastics*, vol. **5**, 107-114.

## 8 Appendix

### 8.1 The extended Kalman filter

*The class  $\mathcal{F}^i$ .* We compute the Laplace transform of a distribution  $\nu_{(\mu, m, \sigma^2)}^i$  (see (41)). For  $X$  a random variable having the previous distribution, elementary computations using (42)-(45) yield, for  $\lambda \in \mathbb{C}$ ,

$$E(\exp \lambda X) = \frac{C_{2i}(m + \mu + \lambda \sigma^2; \sigma^2)}{C_{2i}(m + \mu; \sigma^2)} \exp(\lambda m + \frac{\lambda^2}{2\sigma^2}). \quad (106)$$

All parameters  $(i, m, \mu, \sigma^2)$  are identifiable. From this formula, we can prove that, for all  $m, \mu$  and all  $i$ , as  $\sigma$  tends to 0,  $\nu_{(\mu, m, \sigma^2)}^i$  weakly converges to the Dirac measure  $\delta_m$ . Moreover, for any mixture coefficient  $\alpha$ ,  $\sum_{i \geq 0} \alpha_i \nu_{(\mu, m, \sigma^2)}^i$  weakly converges also to  $\delta_m$ .

*Proof of Proposition 3.1.* Let us consider a random variable  $X$  with distribution  $\nu = \nu_{(\mu, m, \sigma^2)}^i$  and let  $Y = hX + \varepsilon$  with  $X$  and  $\varepsilon$  independent, and  $\varepsilon$  having distribution  $\mathcal{N}(0, \gamma^2)$ . Then,  $\varphi_y(\nu)$  is exactly the conditional distribution of  $X$  given  $Y = y$ . Its density is proportional to:

$$x \rightarrow (x + \mu)^i \exp[-(\frac{(y - hx)^2}{2\gamma^2} + \frac{(x - m)^2}{2\sigma^2})] \quad (107)$$

We compute the exponent of the exponential above and obtain:

$$\frac{(y - hx)^2}{2\gamma^2} + \frac{(x - m)^2}{2\sigma^2} = \frac{(x - \hat{m}(y))^2}{2\hat{\sigma}^2}, \quad (108)$$

with:

$$\hat{m}(y) = \frac{m\gamma^2 + hy\sigma^2}{\bar{\sigma}^2}, \quad \bar{\sigma}^2 = \gamma^2 + h^2\sigma^2, \quad (109)$$

and

$$\hat{\sigma}^2 = \frac{\sigma^2 \gamma^2}{\bar{\sigma}^2}. \quad (110)$$

This implies that  $\varphi_y(\nu) = \nu_{(\mu, \hat{m}(y), \hat{\sigma}^2)}^i$ . So, we get the proposition. Note that this result contains the standard case where  $i = 0$  and  $\mu = 0$ .

*Proof of Proposition 3.2.* As noted in the text above, in this model, the transition kernel  $p(x, x')$  and the conditional kernel  $f_x(y)$  are of the same form. Therefore, the computations of  $\psi(\nu)$  and of the marginal density  $p_\nu(y)$  of  $Y$  are identical up to a change of notations  $((a, \beta^2)$  for  $\psi(\nu)$ , and  $(h, \gamma^2)$  for  $p_\nu(y)$ ). Because of the previous proof, it is more convenient here to compute the marginal density of  $Y$  when  $Y = hX + \varepsilon$  and  $(X, \varepsilon)$  are as in the previous proof. We shall use the same notations as in the statement of Proposition 3.2, but the formulae will be given with  $(h, \gamma^2)$ . We have to integrate  $f_x(y) \nu(dx)$  (with  $\nu = \nu_{(\mu, m, \sigma^2)}^i$ ) with respect to  $x$ . After some elementary computations, we obtain:

$$p_\nu(y) = A_{2i} \exp \left[ -\frac{(y - \bar{m})^2}{2\bar{\sigma}^2} \right], \quad \bar{m} = hm, \quad (111)$$

where (see (109)-(110))

$$A_{2i} = \frac{C_{2i}(\mu + \hat{m}(y); \hat{\sigma}^2)}{C_{2i}(\mu + m; \sigma^2)}. \quad (112)$$

Let us set (see (49))

$$B_i = \sum_{k=0}^i \binom{i}{k} \frac{(\mu + m)^{2k}}{C_{2k} \sigma^{2k}}, \quad (113)$$

$$\bar{\mu} = \frac{m\gamma^2 + \mu\bar{\sigma}^2}{h\sigma^2}. \quad (114)$$

Thus,

$$\mu + \hat{m}(y) = \frac{h\sigma^2}{\bar{\sigma}^2} (y + \bar{\mu}), \quad \frac{\bar{\mu} + \bar{m}}{\bar{\sigma}} = \frac{\mu + m}{\sigma} \frac{\bar{\sigma}}{h\sigma}. \quad (115)$$

After some computations, we obtain

$$A_{2i} = \frac{1}{B_i} \sum_{k=0}^i \binom{i}{k} \left( \frac{h\sigma^2}{\bar{\sigma}^2} \right)^{2k} \frac{(y + \bar{\mu})^{2k}}{C_{2k} \hat{\sigma}^{2k}}. \quad (116)$$

Now, we set

$$a_{k,i} = \binom{i}{k} \left( \frac{h^2\sigma^2}{\gamma^2} \right)^{2k} \sum_{j=0}^k \binom{k}{j} \frac{(\bar{\mu} + \bar{m})^{2j}}{C_{2j} \bar{\sigma}^{2j}}, \quad (117)$$

and

$$\bar{\alpha}_k^{(i)} = \frac{\gamma^{2i}}{B_i \bar{\sigma}^{2i}} a_{k,i}. \quad (118)$$

Finally, for  $k = 0, \dots, i$ , we obtain the following mixture coefficients:

$$\bar{\alpha}_k^{(i)} = \binom{i}{k} \frac{\gamma^{2(i-k)}}{B_i} \sum_{j=0}^k \binom{k}{j} \frac{(\mu + m)^{2j}}{C_{2j} \sigma^{2j}} \frac{h^{2(k-j)} \sigma^{2(k-j)}}{\bar{\sigma}^{2(i-j)}}. \quad (119)$$

And

$$p_\nu(y) dy = \sum_{k=0}^i \bar{\alpha}_k^{(i)} \nu_{(\bar{\mu}, \bar{m}, \bar{\sigma}^2)}^k(dy). \quad (120)$$

So the proof is complete.



## 8.2 The radial Ornstein-Uhlenbeck process.

### 8.2.1 Gaussian moments, Gamma function.

Let us set, for  $\alpha \geq 0$ , and  $X$  a standard Gaussian variable,

$$C_\alpha = \mathbb{E}(|X|^\alpha) \quad (121)$$

And recall the definition of the usual Gamma function

$$\Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x} dx, a > 0. \quad (122)$$

The following relations are obtained by elementary computations.

$$C_{\alpha+1} = \alpha C_{\alpha-1}, \alpha \geq 1, \quad C_\alpha = \frac{\Gamma(\frac{\alpha+1}{2})}{\sqrt{2\pi} 2^{\frac{\alpha+1}{2}}}, \alpha \geq 0, \quad \Gamma(a) = \frac{\sqrt{2\pi}}{2^a} C_{2a-1}, a \geq 1/2. \quad (123)$$

Thus, when  $\alpha = 2i$ ,  $i \in \mathbb{N}$ , *i.e.*  $\alpha$  is an even integer, we obtain,  $C_0 = 1$  and for  $i \geq 1$ ,

$$C_{2i} = (2i-1)C_{2(i-1)} = (2i-1)(2i-3)\dots 5.3.1 = \frac{(2i)!}{2^i i!}. \quad (124)$$

### 8.2.2 Transition densities.

For an integer  $\delta > 1$ , consider processes  $(\xi_t^1, \dots, \xi_t^\delta)$  satisfying for all  $j$ :

$$d\xi_t^j = \theta \xi_t^j dt + \sigma dW_t^j$$

where  $(W_t^j)$  are independent Wiener processes. Let us set  $R_t = \sum_{j=1}^\delta (\xi_t^j)^2$ ,  $X_t = R_t^{1/2}$ . By the Ito formula, we obtain  $dR_t = \sum_{j=1}^\delta 2\xi_t^j d\xi_t^j + \delta\sigma^2 dt$ . By Lévy's characterization, the process defined by

$$\beta_t = \int_0^t \frac{\sum_{j=1}^\delta \xi_t^j dW_t^j}{X_t}$$

is a standard Brownian motion. And,

$$dR_t = (2\theta R_t + \delta\sigma^2)dt + 2\sigma R_t^{1/2} d\beta_t. \quad (125)$$

Therefore, the process  $(R_t)$  is the classical Cox-Ingersoll-Ross diffusion process. Another application of the Ito formula gives the stochastic differential of  $(X_t)$ :

$$dX_t = (\theta X_t + \frac{(\delta-1)\sigma^2}{2X_t})dt + \sigma d\beta_t. \quad (126)$$

Now, we do not assume any more that  $\delta$  is an integer. We assume in the stochastic differential equations (125) and (126) that  $\delta$  is a real parameter satisfying  $\delta > 1$  and define the index  $\nu = (\delta/2) - 1$ . When  $\theta = 0$  and  $\sigma = 1$ ,  $(X_t)$  is the standard Bessel process with index  $\nu$ . The scale and speed densities of  $(X_t)$  are obtained by the classical formulae for one-dimensional diffusion processes. The scale density is given by

$$s(x) = \exp\left(-\frac{2}{\sigma^2} \int^x (\theta u + \frac{(\delta-1)\sigma^2}{2u}) du\right) \propto x^{-(\delta-1)} \exp\left(-\frac{\theta x^2}{\sigma^2}\right). \quad (127)$$

The speed density is  $m(x) = s^{-1}(x)$ . The diffusion process (126) is positive recurrent on  $(0, +\infty)$  for  $\theta < 0$ . In this case, its stationary density is obtained by normalizing  $m$  into a probability density. Setting

$$\rho = \frac{\sigma}{(2|\theta|)^{1/2}},$$

we obtain the stationary density

$$\pi^{(\delta)}(x) = 1_{x>0} \frac{2}{\rho\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\rho^2}\right) \left(\frac{x}{\rho}\right)^{\delta-1} \frac{1}{C_{\delta-1}}. \quad (128)$$

This is the distribution of  $\Gamma^{1/2}$  with  $\Gamma$  having Gamma distribution  $G(\delta/2, 1/2\rho^2)$ .

The processes  $(X_t)$  and  $(R_t)$  have explicit transition probabilities with densities with respect to the Lebesgue measure on  $(0, +\infty)$ . For these, we refer *e.g.* to Karlin and Taylor (p.333-334). For the properties of Bessel functions that we use, we refer *e.g.* to Nikiforov and Ouvarov (1983). The conditional density of  $X_\Delta$  given  $X_0 = x$  is as follows:

$$\begin{aligned} p(\Delta, x, x') &= p^{(\delta)}(x, x') = 2 \times 1_{(x'>0)} (x')^{\delta-1} \exp\left(\frac{\theta x'^2}{\sigma^2}\right) \times \\ &\exp\left(-\frac{\theta}{\sigma^2} \left(\frac{e^{2\theta\Delta}}{e^{2\theta\Delta}-1}\right)(x^2 + x'^2)\right) \left(\frac{\theta}{\sigma^2(e^{2\theta\Delta}-1)}\right) (xx'e^{\theta\Delta})^{-\nu} \times \\ &I_\nu(xx'e^{\theta\Delta} \frac{2\theta}{\sigma^2(e^{2\theta\Delta}-1)}) \end{aligned}$$

where  $I_\nu(z)$  is the Bessel function with index  $\nu$ . This function is given by the following series development

$$I_\nu(z) = \left(\frac{z}{2}\right)^\nu \sum_{k \geq 0} \left(\frac{z}{2}\right)^{2k} \frac{1}{k! \Gamma(k + \nu + 1)}, \quad (129)$$

where  $\Gamma$  is the usual Gamma function. Now, we use the notations (54) and the relations (123) to transform  $p^{(\delta)}(x, x')$  and obtain:

$$p^{(\delta)}(x, x') = 1_{(x'>0)} \frac{2}{\beta\sqrt{2\pi}} \exp\left(-\frac{x'^2}{2\beta^2}\right) \sum_{k \geq 0} \alpha_k(x) \frac{x'^{\delta-1+2k}}{C_{\delta-1+2k} \beta^{\delta-1+2k}}, \quad (130)$$

where the mixture coefficients are given by (see (60))

$$\alpha_k(x) = \exp\left(-\frac{a^2 x^2}{2\beta^2}\right) \frac{1}{k!} \left(\frac{a^2 x^2}{2\beta^2}\right)^k$$

Now, when  $\delta = 2n + 1$ , the index is  $\nu = n - \frac{1}{2}$  a half integer. Then, the Bessel function is explicit and equal to:

$$I_{n-\frac{1}{2}}(z) = \left(\frac{2}{\pi z}\right)^{1/2} z^n \mathcal{T}^n(\cosh(z)) \quad (131)$$

where  $\mathcal{T}^n = \mathcal{T} \circ \dots \mathcal{T}$  is the  $n$ -th iterate of the operator  $\mathcal{T}(f)(x) = f'(x)/x$ . And we obtain (66).

### 8.2.3 Technical lemma.

*Proof of Lemma 4.1*

Let us set

$$\varphi(k) = \frac{C_{2(i+k)+\delta-1}}{C_{2k+\delta-1}C_{2i+\delta-1}} = \frac{1}{C_{2i+\delta-1}}(\delta-1+2k+2i-1)(\delta-1+2k+2i-3)\dots(\delta-1+2k+1). \quad (132)$$

Hence,  $\varphi(x)$  is a polynomial of degree  $i$  which admits a unique representation as a sum of the elementary polynomials  $1, x, x(x-1), \dots, x(x-1)\dots(x-i+1)$ , say

$$\varphi(x) = a'_0 + a'_1x + a'_2x(x-1) + \dots + a'_ix(x-1)\dots(x-i+1). \quad (133)$$

Let us set

$$\psi(x) = a_0 + a_1x + a_2x(x-1) + \dots + a_ix(x-1)\dots(x-i+1) \quad (134)$$

with the coefficients  $a_j$  given in the statement of Lemma 4.1. We will prove that  $\varphi$  and  $\psi$  are identical. For this, it is enough to check that

$$\varphi(j) = \psi(j) = a_0 + ja_1 + j(j-1)a_2 + \dots + j!a_j \quad \text{for all } j = 0, 1, \dots, i. \quad (135)$$

Computing the constant and the higher degree terms, it is easy to see that

$$\varphi(0) = a_0 = \frac{1}{C_{\delta-1}} \quad \text{and} \quad a'_i = a_i = \frac{2^i}{C_{\delta-1+2i}}.$$

Now, let us fix  $0 < j < i$ . Then,  $\varphi(j)$  and  $\psi(j)$  have the following expressions:

$$\begin{aligned} \varphi(j) &= \frac{C_{\delta-1+2(i+j)}}{C_{\delta-1+2j}C_{\delta-1+2i}} \\ &= \frac{1}{C_{\delta-1+2j}}(\delta-1+2j+2i-1)(\delta-1+2j+2i-3)\dots(\delta-1+2i+1) \\ &= P_{\delta-1}^{(j)}(i), \end{aligned} \quad (136)$$

$$\begin{aligned} \psi(j) &= \frac{1}{C_{\delta-1}} + j\frac{2i}{C_{\delta-1+2}} + j(j-1)\frac{2^2i(i-1)}{2C_{\delta-1+4}} \\ &\quad + j(j-1)(j-2)\frac{2^3i(i-1)(i-2)}{3!C_{\delta-1+6}} + \dots \\ &\quad + j!\frac{2^ji(i-1)\dots(i-j+1)}{j!C_{\delta-1+2j}} = Q_{\delta-1}^{(j)}(i). \end{aligned} \quad (137)$$

Hence, both quantities are polynomials of degree  $j$  as functions of the variable  $i$ .

We now prove that, for all  $j$ , all  $\delta$  and all  $y$ ,

$$P_{\delta-1}^{(j)}(x) = Q_{\delta-1}^{(j)}(y), \quad (138)$$

with

$$P_{\delta-1}^{(0)}(y) = Q_{\delta-1}^{(0)}(y) = \frac{1}{C_{\delta-1}}, \quad (139)$$

and

$$P_{\delta-1}^{(j)}(y) = \frac{1}{C_{\delta-1+2j}}(2y + \delta - 1 + 2j - 1)(2y + \delta - 1 + 2j - 3) \dots (2y + \delta - 1 + 1) \quad (140)$$

$$Q_{\delta-1}^{(j)}(y) = \sum_{k=0}^j \binom{j}{k} \frac{2^k y(y-1) \dots (y-k+1)}{C_{\delta-1+2k}}. \quad (141)$$

Let us first look at  $P_{\delta-1}^{(j)}(y)$ . Using  $C_{\delta-1+2j} = (\delta - 1 + 2j - 1)C_{\delta-1+2(j-1)}$ , we get

$$P_{\delta-1}^{(j)}(y) = P_{\delta-1}^{(j-1)}(y) + 2yP_{\delta+1}^{(j-1)}(y-1) \quad (142)$$

Now, we look at  $Q_{\delta-1}^{(j)}(y)$ . Using the relation

$$\binom{j}{k} = \binom{j-1}{k} + \binom{j-1}{k-1}, \quad (143)$$

we obtain

$$Q_{\delta-1}^{(j)}(y) = Q_{\delta-1}^{(j-1)}(y) + 2yQ_{\delta+1}^{(j-1)}(y-1). \quad (144)$$

Therefore, both families of polynomials satisfy the same relation (142). Since (139) holds, we get (138). So, the proof of the Lemma is now complete.