



**HAL**  
open science

# Particle filtering approximations for a Gaussian-generalized inverse Gaussian model

Marco Ferrante, Nadia Frigo

► **To cite this version:**

Marco Ferrante, Nadia Frigo. Particle filtering approximations for a Gaussian-generalized inverse Gaussian model. *Statistics and Probability Letters*, 2009, 79 (4), pp.442. <10.1016/j.spl.2008.09.017>. <hal-00505539>

**HAL Id: hal-00505539**

**<https://hal.science/hal-00505539v1>**

Submitted on 24 Jul 2010

**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.



HAL Authorization

## Accepted Manuscript

Particle filtering approximations for a Gaussian-generalized inverse Gaussian model

Marco Ferrante, Nadia Frigo

PII: S0167-7152(08)00441-0  
DOI: [10.1016/j.spl.2008.09.017](https://doi.org/10.1016/j.spl.2008.09.017)  
Reference: STAPRO 5215

To appear in: *Statistics and Probability Letters*

Received date: 5 March 2008  
Revised date: 16 September 2008  
Accepted date: 17 September 2008

Please cite this article as: Ferrante, M., Frigo, N., Particle filtering approximations for a Gaussian-generalized inverse Gaussian model. *Statistics and Probability Letters* (2008), doi:[10.1016/j.spl.2008.09.017](https://doi.org/10.1016/j.spl.2008.09.017)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



\* Manuscript

[Click here to view linked References](#)

# Particle filtering approximations for a Gaussian-generalized inverse Gaussian model

Marco Ferrante <sup>a,\*</sup>, Nadia Frigo <sup>b</sup>

<sup>a</sup>*Dipartimento di Matematica Pura e Applicata, Università degli Studi di Padova, via Trieste 63, 35121 Padova, Italy*

<sup>b</sup>*Dipartimento di Scienze Statistiche, Università degli Studi di Padova, via Cesare Battisti 241, 35121 Padova, Italy*

---

## Abstract

We consider the filtering problem for a class of discrete-time partially observable stochastic processes. Under strong conditions on the parameters involved and on the initial condition, we are able to prove that it admits a finite dimensional filter. Relaxing these assumptions, we use a Rao Blackwellisation procedure to perform a Particle filtering approximation of the filtering distribution, then we prove its convergence and extend this study to a jump Markov model.

*Key words:* Optimal filter, Finite dimensional filter, Particle filtering  
*1991 MSC:* 60G35, 62M20, 93E11

---

## 1 Introduction

The filtering problem in discrete-time can be described as follows: Let us consider a stochastic process  $\{(U_t, Y_t), t \in \mathbb{N}\}$ , where  $U_t \in \mathbb{R}^{d_U}$ ,  $d_U \in \mathbb{N}$ , and  $Y_t \in \mathbb{R}^{d_Y}$ ,  $d_Y \in \mathbb{N}$ , for any  $t \in \mathbb{N}$ . We assume that the process  $\{U_t, t \in \mathbb{N}\}$  is not directly observable, while so it is the process  $\{Y_t, t \in \mathbb{N}\}$ , which is related to  $\{U_t, t \in \mathbb{N}\}$ . A standard set of assumptions is as follows:

- (M1)  $\{U_t, t \in \mathbb{N}\}$  is a Markov chain with initial density  $a_0(\cdot)$  and transition density  $a_t(x_{t-1}, x)$ , with respect to Lebesgue measure on  $\mathbb{R}^{d_U}$ ;

---

\* Corresponding author.

*Email addresses:* [ferrante@math.unipd.it](mailto:ferrante@math.unipd.it) (Marco Ferrante),  
[nadia@stat.unipd.it](mailto:nadia@stat.unipd.it) (Nadia Frigo).

- (M2) Conditionally on  $U_0, U_1, \dots, U_T$ , the observations  $Y_0, Y_1, \dots, Y_T$  are independent random vectors, for all  $t$ ,  $Y_t$  depends only on  $U_t$  and the conditional distribution of  $Y_t$  given  $U_t = u_t$  admits a density  $b_t(y|u_t)$ , with respect to Lebesgue measure on  $\mathbb{R}^{d_Y}$ .

The filtering problem consists, in its most complete form, in computing, for each  $t \in \mathbb{N}$ , the conditional distribution of  $U_t$  given  $(Y_0, \dots, Y_t)$ , called *the filtering distribution*. Using the following standard notation: For  $s < t$  we define  $y_{s:t} = (y_s, y_{s+1}, \dots, y_t)$  and the same for  $Y_{s:t}$ ,  $u_{s:t}$  and  $U_{s:t}$ , the conditional density of  $U_s$  given  $Y_{0:t} = y_{0:t}$  is expressed by  $p(u_s|y_{0:t})$ . We shall call  $p(u_t|y_{0:t})$  the filtering density,  $p(u_t|y_{0:t-1})$  the prediction density and  $b_t(y_t|u_t)$  the observation density. It is easy to see that the following equations hold:

$$\begin{aligned} \text{prediction} \quad p(u_t|y_{0:t-1}) &= \int p(u|y_{0:t-1})a_t(u, u_t) \, du \\ \text{updating} \quad p(u_t|y_{0:t}) &= \frac{b_t(y_t|u_t)p(u_t|y_{0:t-1})}{\int b_t(y_t|u)p(u|y_{0:t-1}) \, du}. \end{aligned}$$

We call an exact filter system the above algorithm that determines recursively the filtering densities. Often an exact filter system is not very useful in practice, since the number of quantities involved in the computation of the required distributions increases with  $t$ , making the analytic derivation almost impossible. However, if we are able to prove that the filtering density belongs to a given family of probability densities parameterized by a finite-dimensional parameter set, the problem becomes much more tractable and we say that the system admits a *finite dimensional filter*. Starting from the well-known Kalman-Bucy filter, very few finite dimensional filters are known. Some examples can be found in Smith and Miller (1986), Shephard (1994), Ferrante and Vidoni (1998), Ferrante and Vidoni (1999), Vidoni (1999), Runggaldier and Spizzichino (2001) and in the recent paper by Genon-Catalot (2003). Although these models do not present the extreme flexibility and wide range of applicability of the Kalman filter, they are still of extreme interest. Besides many applications to physical and engineering problems, where the process  $U_t$  represents an unknown quantity and the observable process  $Y_t$  its measuring, stochastic filtering techniques have been also applied to economics and finance (see e.g. Barndorff-Nielsen, 1998), where the observable process usually describes a financial time series.

The aim of this paper is twofold: first we shall present a partially observable stochastic process, solution to a system of nonlinear stochastic difference equations, which admits a finite dimensional filter. This system slightly generalizes what considered in Ferrante and Vidoni (1999) and can be interpreted as a model where the unobservable and the observable part, respectively, will provide a stochastic modeling for both the mean and the variance of the observation distribution. Then, relaxing some of the assumptions, we shall develop a Particle Filtering (PF in the sequel) approximation scheme, using a Rao

Blackwellisation procedure, for this and a more general model, where a set of constants is assumed to be stochastic itself. We will be able to provide sufficient conditions under which the approximated distribution converges to the exact one, as the number of particles goes to infinity.

The standard notation  $X \sim N(\mu, \sigma^2)$  will mean:  $X$  is a Gaussian random variable with mean  $\mu$  and variance  $\sigma^2$ . Similar notation will be adopted for other classes of distributions. The less common notation  $(X, Z) \sim N(\mu + \beta Z, Z) \times GIG(\lambda, \delta, \gamma)$  will mean: the joint density of  $(X, Z)$  is the product of the conditional density of  $X$  given  $Z = z$ , which is Gaussian with mean  $\mu + \beta z$  and variance  $z$ , and the density of  $Z \sim GIG(\lambda, \delta, \gamma)$ , where  $GIG$  denotes the generalized inverse Gaussian distribution,

$$\frac{\left(\frac{\gamma}{\delta}\right)^\lambda}{2K_\lambda(\delta\gamma)} x^{\lambda-1} \exp\left\{-\frac{1}{2}(\delta^2 x^{-1} + \gamma^2 x)\right\}, \quad x > 0,$$

with  $K_\lambda(\cdot)$  the modified Bessel function of the second kind and  $\lambda$ ,  $\delta$  and  $\gamma$  belonging to the set  $D = \{(\lambda, \delta, \gamma) \in \mathbb{R} \times [0, +\infty)^2 : \lambda^+ + \delta > 0, \lambda^- + \gamma > 0\}$ . This distribution is called the bivariate Gaussian-generalized inverse Gaussian distribution (see Barndorff-Nielsen and Blæsild, 1981). We shall finally denote by  $\text{supp}X$  the support of the law of the random variable  $X$ .

## 2 A Gaussian-generalized inverse Gaussian model

In this section we shall consider a partially observable stochastic process, solution to a system of nonlinear stochastic difference equations. Its interest lies in the fact that under specific assumptions it admits a finite dimensional filter. This model, except for the family of constants  $\alpha_t$ , has been considered in Ferrante and Vidoni (1999); although the new model is a slight generalization of that, it seems more suitable for a practical use and further future generalizations, since it could be used to model a wider class of real data.

### 2.1 Finite dimensional filter

Let us consider the system of nonlinear stochastic difference equations:

$$X_{t+1} = \alpha_t X_t + \beta (Z_{t+1} - \alpha_t Z_t / 2) + \sqrt{Z_{t+1} - \alpha_t^2 Z_t / 2} V_{t+1}, \quad (1)$$

$$Z_{t+1} = Z_t + W_{t+1}, \quad (2)$$

$$Y_t = X_t + \sqrt{Z_t} V'_t, \quad t \in \mathbb{N} \quad (3)$$

where  $(X_0, Z_0)$  is a given random vector,  $\alpha_t, \beta \in \mathbb{R}$ , with  $|\alpha_t| < \sqrt{2}$ , and  $\{V_{t+1} : t \in \mathbb{N}\}$ ,  $\{V'_t : t \in \mathbb{N}\}$  and  $\{W_{t+1} : t \in \mathbb{N}\}$  are three independent sequences of independent random variables, with  $\text{supp}W_{t+1} \subset [0, +\infty)$ . The process  $\{(X_t, Z_t) : t \in \mathbb{N}\}$  represents the unobservable part, while  $\{Y_t : t \in \mathbb{N}\}$  the observable process.

**Remark 1** *If we assume that  $\{V'_t : t \in \mathbb{N}\}$  is a White noise, equation (3) assures that the conditional distribution of  $Y_t$  given  $(X_t, Z_t) = (x_t, z_t)$ , is that of a  $N(x_t, z_t)$  distributed random variable. The process  $\{(X_t, Z_t, Y_t) : t \in \mathbb{N}\}$  provides hence a model where both the mean  $X_t$  and the variance  $Z_t$  of the Gaussian observation distribution are stochastic.*

**Remark 2** *Note that it is not possible to express the unobservable process in model (1)-(2) in the typical state space formulation*

$$U_{t+1} = U_t + b(U_t) + \sigma(U_t)\xi_t$$

for a given noise sequence  $\{\xi_t\}$ . Moreover, the observable process has the “diffusion” coefficient  $\sqrt{z}$  not bounded away from zero, while a typical assumption is to take it equal to a constant. These two facts make this model interesting to be studied, even if one faces more challenging technical difficulties.

In order to prove that the system (1)-(3) admits a finite dimensional filter, we have to assume the following set of conditions on the noise sequences and the parameter involved:

**(A.1)**  $\{V_{t+1} : t \in \mathbb{N}\}$ ,  $\{W_{t+1} : t \in \mathbb{N}\}$  and  $\{V'_t : t \in \mathbb{N}\}$  are independent sequences of independent random variables. For any  $t \geq 0$ ,  $V_{t+1}$  and  $V'_t$  are standard Gaussian, while  $W_{t+1}$  is Gamma, with shape parameter  $|\lambda_t|$  and scale parameter  $\frac{\gamma^2}{2}$ , where

$$\lambda_0 = \lambda - \frac{1}{2}, \quad \lambda_t = |\lambda_{t-1}| - \frac{1}{2}, \quad t \geq 1, \quad (4)$$

$$\gamma_0^2 = \gamma^2 + \frac{\beta^2}{2}, \quad \gamma_t^2 = \gamma_{t-1}^2 + \frac{\beta^2}{2} = \gamma^2 + (t+1)\frac{\beta^2}{2}, \quad t \geq 1. \quad (5)$$

**(A.2)**  $(X_0, Z_0) \sim N(\mu + \beta Z_0, Z_0) \times \text{GIG}(\lambda, \delta, \gamma)$ , with  $\lambda \in [0, \frac{1}{2})$ , and is independent of the noise processes.

**Remark 3** *The model presented in Ferrante and Vidoni (1999) is (1)-(3) with  $\alpha_t \equiv 1$ . The addition of these coefficients makes this model more flexible and their presence in front of  $X_t$  is quite natural. Indeed, conditioned to  $Z_{0:t}$ , the process  $(X_t, Y_t)$  is an additive Gaussian model, where these coefficients are usually present. Since we look for a finite dimensional filter for the joint process  $(X_t, Z_t)$ , we need, however, to modify carefully the other terms in (1). A similar update for (2), i.e. add a family of constants in front of the term*

$Z_t$ , does not work so well. Indeed, to exploit the peculiar closure property of the GIG distribution with respect to the convolution (Ferrante and Vidoni, 1999, p.168), we would have to multiply also the noise term  $W_{t+1}$  by the same constant. Analogously, adding variable coefficients in (3) is not very useful, since additional terms, depending on these coefficients, would appear in (1) and (2), making the unobservable process dependent on the observable one.

Under the previous assumptions, the partially observable stochastic process  $\{(X_t, Z_t), Y_t, t \in \mathbb{N}\}$ , solution to (1)-(3), admits a finite dimensional filter system, as stated in the following proposition.

**Proposition 4** *Under the assumptions (A.1)–(A.2) the conditional density of  $(X_t, Z_t)$  given  $Y_{0:t} = y_{0:t}$  is for each  $t \in \mathbb{N}$  that of a  $N\left(\mu_t + \beta \frac{Z_t}{2}, \frac{Z_t}{2}\right) \times \text{GIG}(\lambda_t, \delta_t, \gamma_t)$  distributed random vector, where  $\mu_t(\cdot), \delta_t(\cdot)$  are recursively defined by*

$$\mu_0 = \frac{1}{2}(y_0 + \mu) \quad , \quad \mu_t = \frac{1}{2}(y_t + \alpha_{t-1}\mu_{t-1}) \quad , \quad t \geq 1, \quad (6)$$

$$\delta_0^2 = \delta^2 + \frac{1}{2}(y_0 - \mu)^2 \quad , \quad \delta_t^2 = \delta_{t-1}^2 + \frac{1}{2}(y_t - \alpha_{t-1}\mu_{t-1})^2 \quad , \quad t \geq 1. \quad (7)$$

**Proof.** To prove this result, one can follow the same lines of the proof of Proposition 3.1 in Ferrante and Vidoni (1999), with this slight adjustment in the prediction step. Assume that the conditional distribution of  $(X_{t-1}, Z_{t-1})$  given  $Y_{0:t-1} = y_{0:t-1}$  is that of a  $N\left(\mu_{t-1} + \beta \frac{Z_{t-1}}{2}, \frac{Z_{t-1}}{2}\right) \times \text{GIG}(\lambda_{t-1}, \delta_{t-1}, \gamma_{t-1})$  distributed random vector. Denoting by  $p(x_{t-1}, z_{t-1} | y_{0:t-1})$  the filtering density at step  $t-1$ , we get that the conditional density of  $(X_t, Z_t)$  given  $Y_{0:t-1} = y_{0:t-1}$  is given by:

$$p(x_t, z_t | y_{0:t-1}) = \int_{[0, +\infty)} \left\{ \int_{\mathbb{R}} p(x_t | x_{t-1}, z_t, z_{t-1}) p(x_{t-1} | z_{t-1}, y_{0:t-1}) dx_{t-1} \right\} \\ \times p(z_t | z_{t-1}) p(z_{t-1} | y_{0:t-1}) dz_{t-1}.$$

By (1), the closure with respect to the convolution product of the Gaussian densities and applying the change of variable  $s = \alpha_{t-1}x_{t-1}$ , is easy to see that

$$\int_{\mathbb{R}} p(x_t | x_{t-1}, z_t, z_{t-1}) p(x_{t-1} | z_{t-1}, y_{0:t-1}) dx_{t-1}$$

does not depend actually on  $z_{t-1}$  and coincides with the density of a Gaussian r.v. with mean  $\alpha_{t-1}\mu_{t-1} + \beta z_t$  and variance  $z_t$ . Now, replacing the coefficient  $\mu_{t-1}$  with  $\alpha_{t-1}\mu_{t-1}$  for the rest of the proof of Proposition 3.1 in Ferrante and Vidoni (1999), we can complete this proof.  $\square$

In order to derive a PF approximation procedure, we will need at the end of the next sections to be able to compute the conditional density of  $Y_{t+1}$  given  $Y_{0:t} = y_{0:t}$ , which is called in filtering theory the *y-prediction density*. Under the assumptions (A.1)-(A.2), it is possible to prove (see the similar result in Proposition 3.2 of Ferrante and Vidoni (1999)) that the conditional density of  $Y_{t+1}$  given  $Y_{0:t} = y_{0:t}$ , is that of a generalized hyperbolic distribution of parameters  $(|\lambda_t|, \sqrt{\frac{\gamma_t^2}{2} + \frac{\beta^2}{4}}, \frac{\beta}{2}, \alpha_t \mu_t, \sqrt{2}\delta_t)$ , whose definition can be find e.g. in Barndorff-Nielsen and Blæsild (1981).

## 2.2 Model behavior

Before proceeding with the approximation procedure, let us see how the behavior of these processes change according to the choice of the parameters  $\beta$  and  $\alpha_t$  involved in the equations (1)-(3).

The parameter  $\beta$  allows us to calibrate the process  $Z_t$ , which represents the variance of the observation distribution at time  $t$ . For example, it is easy to see that

$$E[Z_t] = E[Z_{t-1}] + E[W_t] = E[Z_{t-1}] + \frac{4|\lambda_{t-1}|}{2\gamma^2 + t\beta^2}.$$

So, if the parameter  $\beta$  is relatively big (in absolute value), the mean of the variables  $Z_t$  increases at a smaller rate. It is interesting to point out that for this model we are able to evaluate the conditional expectation of the process  $\{Z_t : t \in \mathbb{N}\}$ , given the observations  $Y_{0:t}$ . A simple computation (see e.g. Jørgensen, 1982) gives

$$E[Z_t | Y_{0:t} = y_{0:t}] = R_{\lambda_t}(\delta_t \gamma_t) \frac{\delta_t}{\gamma_t}, \quad \text{where} \quad R_\lambda(x) = \frac{K_{\lambda+1}(x)}{K_\lambda(x)}.$$

Since  $\delta_t > 0$ ,  $\gamma_t > 0$  and  $\lambda_t \geq -1/2$ ,  $R_{\lambda_t}(\delta_t \gamma_t)$  goes to 1 as  $t$  goes to  $+\infty$  (see again Jørgensen, 1982), and, for large  $t$ ,  $E[Z_t | Y_{0:t} = y_{0:t}]$  may be approximated by  $\delta_t / \gamma_t$ . From the recursions (5) and (7), we have that in order to control the increase of this conditional expectation, we need again to choose a parameter  $\beta$  not too small, in absolute value. In the Figure 1 the different behavior of the process  $Z_t$  is shown by taking  $\beta = 0.3$  and  $\beta = 3$ .

In order to control the process  $X_t$ , the parameters  $\alpha_t$  play a key role. It is not easy to see if this process could be stationary under suitable assumptions (for example, when  $\alpha_t \equiv 0$ , its expected value increases to infinity). However, we have empirical evidences that, if  $|\alpha_t| < 1$  for every  $t \geq 0$ , then the process  $\{X_t : t \in \mathbb{N}\}$  increases, in absolute value, very slowly, while for  $|\alpha_t| \geq 1$  it diverges to infinity very fast. Figure (1) shows a simulated path of the process  $X_t$  under the assumption  $\alpha_t \equiv 0.5$  and  $\alpha_t \equiv 1$ .

To conclude this subsection, let us consider the parameters  $\lambda_t$ . Since  $\lambda \in [0, \frac{1}{2})$ , it is immediate that  $\lambda_t \leq 0$ , with  $\lambda_{2t} = \lambda - \frac{1}{2}$  and  $\lambda_{2t+1} = -\lambda$ . Two special cases may be emphasized: First, if  $\lambda = 0$ , then  $\lambda_t$  takes the values  $-\frac{1}{2}$  and 0. Then,  $(X_t, Z_t)$  given  $Y_{0:t}$ ,  $t \in \mathbb{N}$ , follows, alternatively, a bivariate Gaussian-inverse Gaussian distribution and a bivariate Gaussian-hyperbola distribution. Secondly, whenever  $\lambda = \frac{1}{4}$ ,  $\lambda_n$  equals  $-\frac{1}{4}$  for each  $t \in \mathbb{N}$  and we can define the process  $Z_t$  by the difference equation

$$Z_{t+1} = Z_t + \left( \frac{\gamma^2}{2} + \frac{t+1}{4} \beta^2 \right)^{-1} W'_{t+1}.$$

where  $\{W'_{t+1}, t \in \mathbb{N}\}$  represents a sequence of i.i.d. gamma random variables, with shape parameter  $\frac{1}{4}$  and scale parameter 1.

### 3 Particle filtering approximation

It appears extremely unlikely that the model (1)-(3) still admits a finite dimensional filter under a different set of assumptions. Then, in order to study this model under less restrictive conditions, a possible alternative is to use numerical approximation methods. The Sequential Monte Carlo methods (aka PF) is a recursive algorithm that produces, at each time  $t$ , a set of simulated values, called *particles*, whose empirical distribution approximates the posterior distribution  $p(dx_{0:t}, dz_{0:t} | y_{0:t})$  (see Doucet et al. , 2001, for a comprehensive account on this topic). The particles are propagated over time using a combination of sequential importance sampling (SIS) and resampling steps. In this chapter we will first describe a PF algorithm, then prove the  $L^2$ -convergence of the approximated distribution to the exact one and at the end we shall extend the numerical results to a Jump Markov model.

#### 3.1 Rao-Blackwellised Particle Filtering

In this subsection we will first describe an algorithm that we can apply in this case in order to derive the approximated filtering density. We shall need the weak assumptions that the initial variable  $X_0$  will be Gaussian and the sequences  $V_t$  and  $V'_t$  White noises. Given these assumptions and the structure of our model, it appears quite natural to marginalize out the variable  $X_{0:t}$  from the posterior. In fact, the density  $p(x_{0:t}, z_{0:t} | y_{0:t})$  can be factorized as

$$p(x_{0:t}, z_{0:t} | y_{0:t}) = p(x_{0:t} | y_{0:t}, z_{0:t}) p(z_{0:t} | y_{0:t})$$

and from the results in Proposition 4, the conditional posterior density  $p(x_{0:t} | z_{0:t}, y_{0:t})$  is Gaussian and then it is analytically tractable using the Kalman filter. We

only need to focus on estimating  $p(dz_{0:t}|y_{0:t})$ , which lies in a space of reduced dimension. Marginalize out some of the variables is an example of the technique called Rao-Blackwellisation, because it is related to the Rao-Blackwell formula (Casella and Robert, 1996). This is a standard technique to increase the efficiency of sampling schemes by reducing the size of the state-space. The Rao-Blackwellised Particle filtering (RBPF) (see e.g. Doucet et al., 2001) combines this technique with the PF procedure. RBPF is very similar to the standard PF, except that each particle now maintains not just a sample from  $p(dz_{0:t}|y_{0:t})$ , which we will denote  $z_{0:t}^{(i)}$ , but also a parametric representation of  $p(dx_t|z_{0:t}^{(i)}, y_{0:t})$ . The  $Z_t$  samples are updated as in standard PF, and then the  $X_t$  distributions are updated using the Kalman filter, conditioned to  $Z_{0:t}$ .

The algorithm we use can be specified as follows:

- (1) Initialization,  $t = 0$ 
  - For  $i = 1, \dots, M$  sample  $z_0^{(i)}$  from a chosen distribution and  $x_0^{(i)}$  from a normal distribution;
  - set  $t = 1$ .
- (2) Importance sampling step and exact step
  - For  $i = 1, \dots, M$  sample  $\tilde{z}_t^{(i)} \sim q(z_t|z_{t-1}^{(i)})$  (we use the prior distribution as proposal distribution) and set  $\tilde{z}_{0:t}^{(i)} = (z_{0:t-1}^{(i)}, \tilde{z}_t^{(i)})$ ;
  - for  $i = 1, \dots, M$  update  $m_{t|t-1}^{(i)}$  and  $P_{t|t-1}^{(i)}$  using the prediction step of the Kalman filter, where  $m_{t|t-1}^{(i)}$  and  $P_{t|t-1}^{(i)}$  are the mean and the variance of  $p(dx_t|\tilde{z}_{0:t}^{(i)}, y_{0:t-1})$ ;
  - for  $i = 1, \dots, M$  evaluate the importance weights  $\tilde{\omega}_t^{(i)} \propto p(y_t|y_{0:t-1}, \tilde{z}_{0:t}^{(i)})$ , that is  $N(m_{t|t-1}^{(i)}, P_{t|t-1}^{(i)} + \tilde{z}_t^{(i)})$ ;
  - normalize the importance weights.
- (3) Selection step and exact step
  - Re-sample with replacement  $M$  particles  $(z_{0:t}^{(1)}, \dots, z_{0:t}^{(M)})$  from the set  $\{\tilde{z}_{0:t}^{(1)}, \dots, \tilde{z}_{0:t}^{(M)}\}$  according to the importance weights;
  - for  $i = 1, \dots, M$  update  $m_{t|t}^{(i)}$  and  $P_{t|t}^{(i)}$  using the update step of the Kalman filter, where  $m_{t|t}^{(i)}$  and  $P_{t|t}^{(i)}$  are the mean and the variance of  $p(dx_t|z_{0:t}^{(i)}, y_{0:t})$ ;
  - set  $t:=t+1$  and go to step 2.

**Remark 5** *It is clear that the previous algorithm can be used to approximate the filtering distribution for a wide class of partially observable stochastic processes. Indeed, we can take  $Z_t$  a generic Markov process, with a given transition Kernel, and  $(X_t, Y_t)$  defined by the set of difference equations:*

$$\begin{aligned} X_t &= A_t(Z_{0:t})X_t + B_t(Z_{0:t}) + C_t(Z_{0:t})V_{t+1}, \\ Y_t &= X_t + D_t(Z_{0:t})V_t', \quad t \in \mathbb{N}, \end{aligned}$$

under the assumptions that the noises  $V_{t+1}$  and  $V_t'$  and the initial data  $X_0$  are independent and Gaussian.

If in the step (1) of the previous algorithm, we sample  $z_0^{(i)}$  from a  $\text{GIG}(\lambda, \delta, \gamma)$ ,  $x_0^{(i)}$  from a  $\text{N}(\mu + \beta z_0^{(i)}, z_0^{(i)})$  and the hypothesis (A.1)-(A.2) are satisfied, we get an approximation of the exact distribution of the finite dimensional filter. In this case we can use the exact filtering distribution to test "empirically" this algorithm and the extremely nice results are presented in Figure 2.

### 3.2 Convergence of the empirical measure

A basic result in particle filtering is to prove that the approximated distribution converges to the optimal one, as the number of particles goes to infinity. Here, under some additional assumptions, that sadly do not match with hypothesis (A.2), we are able to prove the convergence of the PF algorithm, while we shall give just an empirical evidence that the approximated distribution still converges to the exact one under the assumptions (A1)-(A2).

We will apply here a result proved in Crisan and Doucet (2000) for a general class of PF, which include our RBPF. Fix a given observation record  $Y_{0:t} = y_{0:t}$  we denote by  $\pi_t^M$  the empirical measure

$$\frac{1}{M} \sum_{i=1}^M \delta_{z_{0:t}^{(i)}}(dz_{0:t})$$

and by  $\pi_t$  the distribution of  $Z_{0:t}$  given  $Y_{0:t} = y_{0:t}$ . Under the assumption that the conditional distribution of the observation process with respect to  $Z_{0:t} = z_{0:t}$ , is bounded as a function of the variables  $z_{0:t}$ , they are able to prove in Theorem 1, p.16, that for any  $t \geq 0$ , there exist a constant  $c_t$ , independent of  $M$ , such that, for any measurable function  $f \in \mathcal{B}(\mathbb{R}^t)$ ,

$$E\left[\left((\pi_t^M, f) - (\pi_t, f)\right)^2\right] \leq c_t \frac{\|f\|^2}{M}.$$

To apply this result, it will be sufficient to put some restriction on the support of the Law of the initial distribution  $Z_0$  and we are able to prove the following proposition:

**Proposition 6** *Let us consider the model (1)-(3) and assume that  $V_{t+1}$  and  $V_t'$  are independent White noises,  $X_0$  is a Gaussian random variable, independent of the noise sequences, and  $\text{supp} Z_0 \subset (0, +\infty)$ . Fixed a given observation record  $Y_{0:t} = y_{0:t}$ , the empirical measure  $\pi_t^M$  converges in  $L^2$  to the conditional distribution  $\pi_t$ , as the number of particles  $M$  goes to infinity.*

**Proof.** In order to prove that the empirical measure converges, it will be

sufficient to prove that the conditional density of  $Y_t$  given  $Y_{0:t-1} = y_{0:t-1}$ ,  $Z_{0:t} = z_{0:t}$  is bounded as a function of  $z_{0:t}$ . A simple computation gives that

$$\begin{aligned} p(y_t|y_{0:t-1}, z_{0:t}) &= \int_{\mathbb{R}} p(y_t|y_{0:t-1}, z_{0:t}, x_t)p(x_t|y_{0:t-1}, z_{0:t})dx_t = \\ &= \int_{\mathbb{R}} p(y_t|z_t, x_t)p(x_t|y_{0:t-1}, z_{0:t})dx_t . \end{aligned}$$

This is a Gaussian density with mean  $m_{t,t-1}$  and variance  $P_{t,t-1} + z_t$ , where  $m_{t,t-1}$  and  $P_{t,t-1}$  are, respectively, the mean and the variance of the prediction density in the (conditioned) Kalman filter. It is easy to see that this density is unbounded as a function of  $z_{0:t}$ , unless there exists  $\varepsilon > 0$  such that  $z_s \geq \varepsilon$  for any  $s \leq t$ . Since the assumption that the support of  $Z_0$  is a subset of  $(0, +\infty)$  implies that there exists  $\varepsilon > 0$  such that  $Z_s \geq \varepsilon > 0$  almost surely for any  $s \leq t$ , we can use the result in Crisan and Doucet (2000) to prove the convergence.

□

**Remark 7** *The same result holds true if we consider a standard PF algorithm instead of the present RBPF. In that case, the density that have to be bounded is  $p(y_t|x_t, z_t)$  and this is true if and only if again  $Z_0$  is almost surely bounded away from 0.*

The convergence result of Proposition 6 does not apply to our model under the assumptions (A1)-(A2), since  $\text{supp } GIG(\lambda, \delta, \gamma) = [0, +\infty)$ . Nevertheless, we can use an empirical approach to see whether or not the approximated filtering distributions converge to the exact one as the number of particles increase. Let us consider the case in which a series of  $T = 50$  values are simulated exactly according to the recursive formulas for the parameters involved. We fix the parameters of the model equal to:  $\lambda = 0.4$ ,  $\delta = 1$ ,  $\gamma = 4$ ,  $\mu = 0$ ,  $\beta = 3$  and  $\alpha_t = 0.5$  for every  $t \geq 0$ . We can first apply the algorithm described in Subsection 3.1 in order to estimate the filtering distribution of the model. In this case, we know the exact finite dimensional filter, so we can test the approximation using the result given in Proposition 4. For a fixed value of  $M$ , we repeat the algorithm  $R = 100$  times and we compare the true mean of  $p(dx_t|y_{0:t})$  and of  $p(dz_t|y_{0:t})$  with the empirical ones obtained by using the algorithm. The procedure is run by setting the number of particles equal to  $M = 50$  and  $M = 200$ . In order to obtain information about the distribution of the empirical mean, we collect the mean and the standard deviation of the samples obtained through the algorithm. For time  $T = 50$ , we report these values in the Table 1, where instead of the mean values we record the absolute value of the difference between the true and the sample mean. As we expect, increasing the number of particles, these distributions concentrate around the true values. We implemented also a standard PF procedure for

	RBPF		PF	
	$M = 50$	$M = 200$	$M = 50$	$M = 200$
$E(X_t y_{0:t})$				
$error_T$	0.0136	0.0004	0.0099	0.0019
$std_T$	0.0223	0.0135	0.1377	0.0852
$\max_{i=1,\dots,T}(error_i)$	0.0154	0.0055	0.0530	0.0152
$E(Z_t y_{0:t})$				
$error_T$	0.0990	0.0184	0.1536	0.0353
$std_T$	0.1452	0.0853	0.1825	0.1049
$\max_{i=1,\dots,T}(error_i)$	0.0990	0.0184	0.1536	0.0366

Table 1

Error and standard deviation of the empirical distribution of the mean of the processes  $X_t|y_{0:t}$  and  $Z_t|y_{0:t}$ .  $T = 50$ . Maximum of the error over the time is also reported.

both the variables. As we can see in Table 1, the convergence property is still obtained, but the magnitude of the errors is bigger, as theoretically proved, than using the RBPF algorithm.

### 3.3 A Jump Markov Gaussian-generalized inverse Gaussian

A natural generalization of the previous model is to assume that the sequence of constants  $\{\alpha_t, t \in \mathbb{N}\}$  will be a finite state Markov chain with known initial distribution and transition probabilities. Such a process is usually called a Jump Markov model (see e.g. Andrieu, Davy and Doucet, 2003) and in this way we obtain a more flexible model, with a wider range of potential application. To fix the notation, let us assume that  $\alpha_t$  will be a discrete-time, time homogeneous Markov chain with transition probabilities  $p_{m,n} = P\{\alpha_{t+1} = n | \alpha_t = m\}$ , with  $m, n \in S$  and initial probability distribution  $p_m = P\{\alpha_0 = m\}$ . Under the assumptions (A1)-(A2) and with this extension, we are no more able to compute the exact filter, but we can apply a RBPF procedure to the model  $(\alpha_{0:t}, X_{0:t}, Z_{0:t}, Y_{0:t})$ . Indeed, conditionally on  $\alpha_{0:t}$ ,  $p(X_t, Z_t | \alpha_{0:t}, y_{0:t})$  is analytically tractable and so we can marginalized out  $(X_t, Z_t)$  from the posterior and focus on  $p(\alpha_{0:t} | y_{0:t})$  using a RBPF technique. The algorithm acts in the same way as before, but instead of using the Kalman filter in the exact step, we update the parameters  $\mu_t$  and  $\delta_t$  of the exact filter given in Proposition 4. Note that in the importance sampling step we use the result on the y-prediction density proved at the end of the Subsection 2. The new algorithm that we obtain is as follows:

- (1) Initialization,  $t = 0$ 
  - For  $i = 1, \dots, M$  sample  $\alpha_0^{(i)}$  from the initial probability distribution and set  $\mu_0^{(i)} = \mu_0$  and  $\delta_0^{(i)} = \delta_0$ ;
  - set  $t = 1$ .
- (2) Importance sampling step
  - For  $i = 1, \dots, M$  sample  $\tilde{\alpha}_t^{(i)}$  from the transition probability  $P\{\alpha_t|\alpha_{t-1}^{(i)}\}$  and set  $\tilde{\alpha}_{0:t}^{(i)} = (\alpha_{0:t-1}^{(i)}, \tilde{\alpha}_t^{(i)})$ ;
  - for  $i = 1, \dots, M$  evaluate the importance weights  $\tilde{\omega}_t^{(i)} \propto p(y_t|y_{0:t-1}, \tilde{\alpha}_{0:t}^{(i)})$  that is GH  $(|\lambda_{t-1}|, \sqrt{\frac{1}{2}\gamma_{t-1}^2 + \frac{1}{4}\beta^2}, \frac{\beta}{2}, \tilde{\alpha}_{t-1}^{(i)}\mu_{t-1}^{(i)}, \sqrt{2}\delta_{t-1}^{(i)})$ ;
  - normalize the importance weights.
- (3) Selection step and exact step
  - Re-sample with replacement  $M$  particles  $(\alpha_{0:t}^{(1)}, \dots, \alpha_{0:t}^{(M)})$  from the set  $\{\tilde{\alpha}_{0:t}^{(1)}, \dots, \tilde{\alpha}_{0:t}^{(M)}\}$  according to the importance weights;
  - for  $i = 1, \dots, M$  update the parameters of the distribution  $p(dx_t, dz_t|\alpha_{0:t}^{(i)}, y_{0:t})$  that is the exact filtering distribution  $N(\mu_t^{(i)} + \beta Z_t/2, Z_t/2) \times \text{GIG}(\lambda_t, \delta_t^{(i)}, \gamma_t)$ ;
  - set  $t:=t+1$  and go to step 2.

As an application, let us take  $S = \{0.5, 1\}$  and the stationary distribution as initial probability. We consider the case  $p_{m,n} = 0.5$  for every  $m, n \in S$  and  $p_{m,n} = 0.9$  for  $m \in S, n = 1$ . It is interesting to note that, even if the probability that the Markov chain visit the state 1 is very high, the filtering distribution is much closer to that of the deterministic case  $\alpha_t \equiv 0.5$ , since even occasional visits to the value 0.5 let decrease the process very much (see Figure 3).

#### 4 Conclusions

In this paper we have first proved that a class of discrete time partially observable stochastic processes admits a finite dimensional filter, under strong assumptions and then, relaxing these assumptions, we have considered a PF approximation of the filtering distribution. Then we have proved the convergence of the approximated filtering distribution and we have extended the numerical approximation to a Jump Markov model. We plan to continue the present research, in order to prove the stability and the uniform convergence of the PF approximation for this and similar models, and to study further statistical properties.

## References

- Andrieu, C., Davy, M. and Doucet, A., 2003. Efficient particle filtering for jump Markov systems. Application to time-varying autoregressions. *IEEE Trans. Signal Process.* 51, 1762–1770.
- Barndorff-Nielsen, O. E., Blæsild, P., 1980. Hyperbolic distributions and ramifications: contributions to theory and application, in *Statistical distributions in scientific work*, Vol. 4, pp. 19-44.
- Barndorff-Nielsen, O. E., 1998. Processes of normal inverse Gaussian type. *Finance Stoch.* 2, 41-68.
- Casella, G., Robert, C.P., 1996. Rao-Blackwellisation of sampling schemes. *Biometrika* 83, 81-94.
- Crisan, D., Doucet, A., 2000. Convergence of sequential Monte Carlo methods. Technical Report.
- Crisan, D., Doucet, A., 2002. A survey of convergence results on particle filtering methods for practitioners. *IEEE Trans. Signal Process.* 3, 736-46.
- Doucet, A., de Freitas, N. and Gordon, N., 2001. *Sequential Monte Carlo methods in practice*. Springer-Verlag, New York.
- Doucet, A., de Freitas, N., Murphy, K. and Russel, S., 2000. Rao-Blackwellised particle filtering for dynamic Bayesian networks. In *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, 176-83. Stanford, 2000.
- Ferrante, M. and Vidoni, P., 1998. Finite-dimensional filters for nonlinear stochastic difference equations with multiplicative noises. *Stochastic Process. Appl.* 77, 69-81.
- Ferrante, M. and Vidoni, P., 1999. A Gaussian-generalized inverse Gaussian finite-dimensional filter. *Stochastic Process. Appl.* 1, 165-76.
- Genon-Catalot, V., 2003. A non-linear explicit filter. *Statist. Probab. Lett.* 61, 145-54.
- Jørgensen, B., 1982. *Statistical properties of the generalized inverse Gaussian distribution*. Lecture Notes in Statistics 9. Springer-Verlag, New York.
- Meyn, S.P. and Tweedie, R.L., 1993. *Markov Chains and Stochastic Stability*. Springer, London.
- Runggaldier, W. J. and Spizzichino, F., 2001. Sufficient conditions for finite dimensionality of filters in discrete time: a Laplace transform-based approach. *Bernoulli* 7, 211-21.
- Shephard, N., 1994. Local scale models: state space alternative to integrated GARCH processes. *J. Econometrics* 60, 181-202.
- Smith, R. L. and Miller, J. E., 1986. A non-Gaussian state space model and application to prediction of records. *J. Roy. Statist. Soc. Ser. B* 48, 79-88.
- Vidoni, P., 1999. Exponential family state space models based on a conjugate latent process. *J. Roy. Statist. Soc. Ser. B* 61, 213-21.

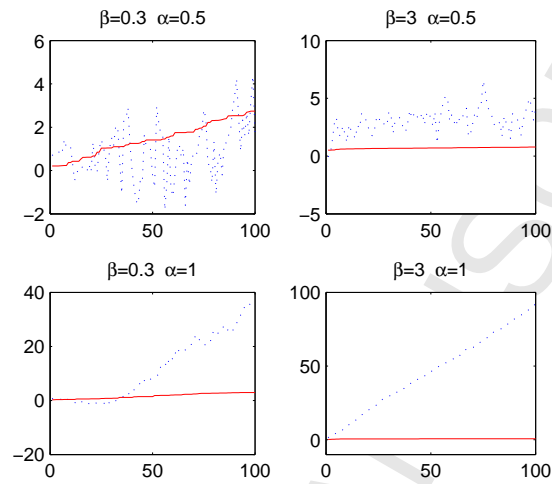


Fig. 1. Effect of changing in the value of  $\beta$  and  $\alpha_t \equiv \alpha$  with  $\mu = 0$ ,  $\lambda = 0.4$ ,  $\delta = 1$ ,  $\gamma = 4$  fixed.  $\{X_t\}$  dotted line and  $\{Z_t\}$  solid line.

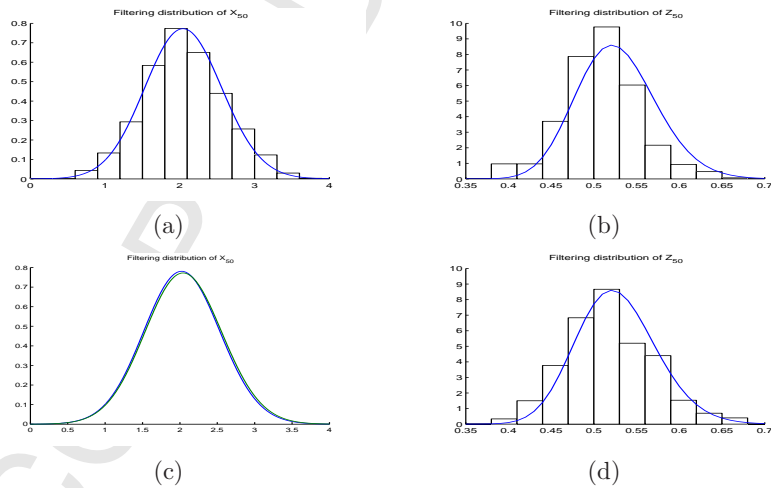


Fig. 2. Approximations of  $p(X_{50}|y_{0:50})$  (a) and  $p(Z_{50}|y_{0:50})$  (b). Exact distributions (solid lines) coincide with the estimated one when (A1)-(A.2) hold (PF algorithm (a,b), RBPF algorithm (c,d)).  $M = 1000$ .

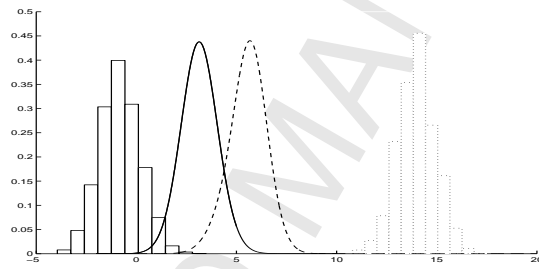


Fig. 3. Comparison between the approximated filtering distribution  $p(X_{50}|y_{0:50})$  obtained by using a standard PF algorithm in the case in which  $\alpha_t \equiv 0.5$  (solid histogram) and  $\alpha_t \equiv 1$  (dotted histogram). The lines represent the approximated filtering distribution  $p(X_{50}|y_{0:50}, \alpha_{0:50})$  when  $\alpha_t$  is a 2-state Markov chain with  $p_{m,n} = 0.5$  for every  $m, n \in S$  (solid line) and  $p_{m,n} = 0.9$  for  $m \in S, n = 1$  (dotted line). Always  $M = 3000$ .