

Particle filtering with dependent noise

Fredrik Gustafsson and Saikat Saha

Department of Electrical Engineering

Linköping University

Linköping, Sweden

[fredrik, saha}@isy.liu.se](mailto:{fredrik, saha}@isy.liu.se)

Abstract – *The theory and applications of the particle filter (PF) have developed tremendously during the past two decades. However, there appear to be no version of the PF readily applicable to the case of dependent process and measurement noise. This is in contrast to the Kalman filter, where the case of correlated noise is a standard modification. Further, the fact that sampling continuous time models give dependent noise processes is an often neglected fact in literature. We derive the optimal proposal distribution in the PF for general and Gaussian noise processes, respectively. The main result is a modified prediction step. It is demonstrated that the original Bootstrap particle filter gets a particular simple and explicit form for dependent Gaussian noise. Finally, the practical importance of dependent noise is motivated in terms of sampling of continuous time models.*

Keywords: nonlinear filtering, particle filter, dependent noise, optimal proposal density

1 Problem Formulation

The standard model in particle filtering literature (see e.g. [3]– [7]) is based on the prediction model and likelihood function, respectively,

$$p(x_{k+1}|x_k), \quad (1a)$$

$$p(y_k|x_k). \quad (1b)$$

The case of dependent noise can be phrased as $p(y_k, x_{k+1}|x_k)$, where

$$p(y_k, x_{k+1}|x_k) \neq p(x_{k+1}|x_k)p(y_k|x_k), \quad (2)$$

that is, y_k and x_{k+1} are not independent. In engineering literature, the following state space model, and variants of it, is commonly used,

$$x_{k+1} = f_k(x_k) + Gv_k, \quad (3a)$$

$$y_k = h_k(x_k) + e_k. \quad (3b)$$

The inequality (2) applies in case the process noise v_k and the measurement noise e_k are dependent. The proposal distribution for this case is derived in Section 2. To get instructive and explicit expressions, the Gaussian case,

$$x_0 \sim N(\hat{x}_{1|0}, P_{1|0}), \quad (3c)$$

$$\begin{pmatrix} v_k \\ e_k \end{pmatrix} \in N\left(0, \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix}\right), \quad (3d)$$

is studied in detail in 3. The standard PF applies for the case $S = 0$ only. This Gaussian model can also be written as

$$p\left(\begin{pmatrix} x_{k+1} \\ y_k \end{pmatrix} | x_k\right) = N\left(\begin{pmatrix} f(x_k) \\ h(x_k) \end{pmatrix}, \begin{pmatrix} GQG^T & GS \\ S^TG^T & R \end{pmatrix}\right). \quad (4)$$

Note that this noise covariance is the standard representation in the classic text book [1] on Kalman filters (KF).

The case of dependent noise might be more common in practice than is acknowledged in literature. More specifically, it occurs whenever a (linear or nonlinear) filter is based on a discrete time model that is derived from a continuous time model. The only exception is when a piecewise constant noise process is assumed. Section 4 motivates the importance of dependent noise processes in these terms.

2 Optimal Proposal for Dependent Noise

The proposal distribution has the functional form $q(x_k|x_{1:k}, y_{1:k})$. In the standard PF, the Markovian property and the independence of $y_{1:k-1}$ and x_k are used to get

$$(1) \Rightarrow q(x_k|x_{1:k-1}, y_{1:k}) = q(x_k|x_{k-1}, y_k). \quad (5a)$$

For the case (2), there is a correlation between y_{k-1} and x_k , so we get

$$(2) \Rightarrow q(x_k|x_{1:k-1}, y_{1:k}) = q(x_k|x_{k-1}, y_k, y_{k-1}). \quad (5b)$$

Based on this, we derive the following theorem for the optimal proposal function.

Theorem 1 [Optimal proposal for dependent noise] *For the model specified by (2), the optimal proposal function is given by*

$$q(x_k|x_{k-1}, y_k, y_{k-1}) = \frac{p(x_k, y_{k-1}|x_{k-1})p(y_k|x_k)}{p(y_k, y_{k-1}|x_{k-1})}. \quad (6)$$

Proof: The optimal proposal is given by the posterior distribution of the same form, which can be rewritten using Bayes' law twice and two conditional independence properties in the last equality,

$$q(x_k|x_{k-1}, y_k, y_{k-1}) \quad (7a)$$

$$= p(x_k|x_{k-1}, y_k, y_{k-1}) \quad (7b)$$

$$= \frac{p(x_k, y_{k-1}|x_{k-1}, y_k)}{p(y_{k-1}|x_{k-1}, y_k)} \quad (7c)$$

$$= \frac{p(x_k, y_{k-1}|x_{k-1})p(y_k|x_k, x_{k-1}, y_{k-1})}{p(y_{k-1}|x_{k-1}, y_k)p(y_k|x_{k-1})} \quad (7d)$$

$$= \frac{p(x_k, y_{k-1}|x_{k-1})p(y_k|x_k)}{p(y_{k-1}|x_{k-1})p(y_k|x_{k-1})} \quad (7e)$$

$$= \frac{p(x_k, y_{k-1}|x_{k-1})p(y_k|x_k)}{p(y_k, y_{k-1}|x_{k-1})}. \quad (7f)$$

This concludes the proof. \square
The optimal proposal as described in (6) has been used as a special case for estimating the stochastic volatility in [8]. This optimal proposal should be compared to the standard one given by

$$q(x_k|x_{k-1}, y_k) = \frac{p(x_k|x_{k-1})p(y_k|x_k)}{p(y_k|x_{k-1})}. \quad (8)$$

One can just as for the standard PF define two extreme cases of sub-optimal proposal distributions

$$\text{Prior :} \quad q(x_k|x_{k-1}, y_{k-1}) \propto p(x_k, y_{k-1}|x_{k-1}) \quad (9a)$$

$$\text{Likelihood :} \quad q(x_k|y_k) \propto p(y_k|x_k). \quad (9b)$$

The first proposal corresponds to the model (2), while the second proposal requires a marginalization of the model using

$$p(y_k|x_k) = \int p(x_{k+1}, y_k|x_k)dx_{k+1}. \quad (10)$$

3 Gaussian Model

We next study the Gaussian model in (4). The main result is given in Theorem 1.

Theorem 2 [Optimal proposal for Gaussian dependent noise] *For the model specified by (4), the op-*

timal proposal function is given by

$$\begin{aligned} & q(x_k|x_{k-1}, y_k, y_{k-1}) \propto \\ & \text{N}\left(x_k - f(x_{k-1}) + G_{k-1}S_{k-1}R_{k-1}^{-1}(y_{k-1} - h(x_{k-1})), \right. \\ & \quad \left. G_{k-1}(Q_{k-1} - S_{k-1}R_{k-1}^{-1}S_{k-1}^T)G_{k-1}^T\right) \\ & \times \text{N}(y_k - h(x_k), R_k). \end{aligned} \quad (11)$$

Proof: The result follows from (6) by studying the two factors in the numerator. The second factor follows by straightforward marginalization (10). The first factor $(x_k|x_{k-1}, y_{k-1})$ follows by using Lemma 1 below. \square

Lemma 1 [Conditional Gaussian Distributions] *Suppose the vectors X and Y are simultaneously Gaussian distributed*

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \text{N}\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} P_{xx} & P_{xy} \\ P_{xy} & P_{yy} \end{pmatrix}\right) = \text{N}\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, P\right). \quad (12)$$

Then, the conditional distribution for X , given the observed $Y = y$, is Gaussian distributed:

$$(X|Y = y) \sim \text{N}(\mu_x + P_{xy}P_{yy}^{-1}(y - \mu_y), P_{xx} - P_{xy}P_{yy}^{-1}P_{yx}). \quad (13)$$

Since the optimal proposal in (6) can be evaluated point-wise for each sample x_k up to an unknown scaling factor $p(y_k, y_{k-1}|x_{k-1})$, the principle of sampling importance resampling is applicable to generate random numbers from the proposal $q(x_k|x_{k-1}, y_k, y_{k-1})$. Note that another proposal function is needed here.

The prior proposal in (9a) becomes more explicit as summarized in the Corollary 1.

Corollary 1 [Prior proposal for Gaussian dependent noise] *For the model specified by (4), the optimal proposal function is given by*

$$\begin{aligned} & q(x_k|x_{k-1}, y_{k-1}) = \\ & \text{N}\left(x_k; f(x_{k-1}) - G_{k-1}S_{k-1}R_{k-1}^{-1}(y_{k-1} - h(x_{k-1})), \right. \\ & \quad \left. G_{k-1}(Q_{k-1} - S_{k-1}R_{k-1}^{-1}S_{k-1}^T)G_{k-1}^T\right). \end{aligned} \quad (14)$$

Proof: In this case, the proposal is $p(x_k|x_{k-1}, y_{k-1})$ which is directly given by Lemma 1. \square

It is thus straightforward to generate samples from this proposal using the Gaussian random number generator. It is the correction terms that are non-trivial. The standard SIR PF is obtained by letting $S = 0$ in (11). The generalized version of the SIR PF is summarized in Algorithm 1.

Algorithm 1 [SIR PF for dependent noise] *Iteration: For $k = 1, 2, \dots$*

1. Measurement update: For $i = 1, 2, \dots, N$,

$$w_{k|k}^i = \frac{1}{c_k} w_{k|k-1}^i \mathcal{N}(y_k; h(x_k^i), R_k) \quad (15a)$$

where the normalization weight is chosen such that $\sum_{i=1}^N w_{k|k}^i = 1$.

2. Resampling: Optionally at each time, take N samples with replacement from the set $\{x_{1:k}^i\}_{i=1}^N$ where the probability to take sample i is $w_{k|k}^i$ and let $w_{k|k}^i = 1/N$.

3. Time update: Generate predictions according to the proposal distribution

$$x_k^i \sim \mathcal{N}\left(f(x_{k-1}^i) - G_{k-1}S_{k-1}R_{k-1}^{-1}(y_{k-1} - h(x_{k-1}^i)), G_{k-1}(Q_{k-1} - S_{k-1}R_{k-1}^{-1}S_{k-1}^T)G_{k-1}^T\right). \quad (15b)$$

4 Dependent Noise in Sampled Systems

To motivate the importance of dependent noise, consider a continuous time linear model with a noisy input,

$$\dot{x}(t) = Ax(t) + B(u(t) + v(t)), \quad (16a)$$

$$y(t) = Cx(t) + D(u(t) + v(t)) + e(t). \quad (16b)$$

For the case $D \neq 0$, the correlation between process noise $v(t)$ and measurement noise $Dv(t) + e(t)$ is obvious. However, even in the case $D = 0$, correlation will occur in the discretized system due to sampling. Table 1 summarizes different sampling strategies (see Chapter 13.1 in [2]), and the corresponding discrete time model is given by,

$$x_{k+1} = Fx_k + G(u_k + v_k) = Fx_k + Gu_k + \bar{v}_k, \quad (17a)$$

$$y_k = Hx_k + J(u_k + v_k) + e_k = Hx_k + Ju_k + \bar{e}_k, \quad (17b)$$

$$\begin{pmatrix} \bar{v}_k \\ \bar{e}_k \end{pmatrix} \in \mathcal{N}\left(0, \begin{pmatrix} GQG^T & GQJ^T \\ JQG^T & R \end{pmatrix}\right). \quad (17c)$$

A typical example of such models is in inertial navigation, where $u(t)$ is the acceleration measured by an accelerometer, and the model for position is given by $\ddot{p}(t) = u(t)$. The accelerometer signal is usually low-pass filtered before the sampling, and the bilinear sampling formula is appropriate.

The only exception where correlation does not occur, is when the process noise is assumed piecewise constant between the sampling instants (ZOH), which is clearly unphysical but still the dominating assumption in the filtering literature.

Virtually all models used in navigation (see e.g. [9]) suffer from dependent noise. Other examples than the

acceleration model $\ddot{p}(t) = u(t)$ include angular rate integration $\dot{q}(t) = u(t)$, where $q(t)$ is an orientation vector, and odometry $\dot{p}(t) = u(t)$, where velocity is measured. All motion models used in target tracking have basically the same problem, the difference is that there is no measured input $u(t)$.

5 Practical Examples

Here we mention some popular applications where the dependency between the process noise and the observation noise in the state space model appears quite naturally.

5.1 Sensor Fusion Application

Consider the problem of fusing GPS position measurements with inertial measurements from an IMU (inertial measurement unit). This is one of the most classical sensor fusion applications. However, one less investigated issue is time alignment of these two sensors [10]. Let the timing error in the GPS receiver compared to the IMU be Δ , which is not an integer multiple of the sampling interval, but rather a fractional delay. As a simplified example of the application in [10], consider a one-dimensional model, where the IMU measures linear acceleration a_k with noise w_k ,

$$p_{k+1} = p_k + T_s v_k + T_s^2/2(a_k + w_k), \quad (18)$$

$$v_{k+1} = v_k + T_s a_k + T_s w_k, \quad (19)$$

$$\begin{aligned} y_k &= p(kT_s - \Delta) + e_k \\ &\approx p_k - v_k \Delta + (a_k + w_k) \Delta^2/2 + e_k, \end{aligned} \quad (20)$$

where a second order Taylor expansion of the position $p_k = p(kT_s)$ has been included. Since the acceleration input enters both the dynamics and measurement equation with its measurement noise, we have an example of dependent noise.

The example above can be generalized to any sensor model $h(p_k)$ that depends on the position, including all the applications in [9], using the Taylor expansion

$$\begin{aligned} y_k &= h(p(kT - \Delta)) + e_k \\ &\approx h(p_k - v_k \Delta + (a_k + w_k) \Delta^2/2) + e_k. \end{aligned} \quad (21)$$

5.2 Econometric Application

Consider the so called GARCH-M model for estimating the volatility from the observed mean return data. This model is very popular in financial literature and can be described as follows :

$$h_t = \gamma_1 + \gamma_2 h_{t-1} + \gamma_3 \epsilon_{t-1}^2 \quad (22)$$

$$r_t = \mu + \delta \sqrt{h_t} + \epsilon_t \quad (23)$$

$$(24)$$

Table 1: Correlation due to sampling of the state $x(t) = (p(t), v(t))^T$ in a double integrator using zero-order hold (ZOH, assuming piecewise constant input), first-order hold (FOH, assuming piecewise linear input), and bilinear transformation (BIL), which is an often used approximation for band-limited signals.

Continuous time	$A = \begin{pmatrix} 0_n & I_n \\ 0_n & 0_n \end{pmatrix}$	$B = \begin{pmatrix} 0_n \\ I_n \end{pmatrix}$	$C = (I_n, 0_n)$	$D = 0_n$
ZOH	$F = \begin{pmatrix} I_n & TI_n \\ 0_n & I_n \end{pmatrix}$	$G = \begin{pmatrix} \frac{T^2}{2}I_n \\ TI_n \end{pmatrix}$	$H = (I_n, 0_n)$	$J = 0_n$
FOH	$F = \begin{pmatrix} I_n & TI_n \\ 0_n & I_n \end{pmatrix}$	$G = \begin{pmatrix} T^2I_n \\ TI_n \end{pmatrix}$	$H = (I_n, 0_n)$	$J = \frac{T^2}{6}I_n$
BIL	$F = \begin{pmatrix} I_n & TI_n \\ 0_n & I_n \end{pmatrix}$	$G = \begin{pmatrix} \frac{T^2}{4}I_n \\ \frac{T}{2}I_n \end{pmatrix}$	$H = (I_n, \frac{T}{2}I_n)$	$J = \frac{T^2}{2}I_n$

where h_t is the unobserved volatility, r_t is the observed mean return, $\epsilon_t = \sqrt{h_t}z_t$ with $z_t \sim N(0, 1)$. The rest of the parameters are the same as described in [11] (see equation (2)–(3) of section 3.1 there). This is another example where the noises are found to be dependent.

6 Conclusions

The fact that the process noise is correlated with the measurement noise in sampled models is often neglected in literature. There might be several reasons for this. The process noise is often instrumental for the tuning, so its physical is of less importance. Another reason is that the correlation can be quite small for fast sampling compared to the time constant of the system. A final reason might be that the PF theory is not yet adapted to dependent noise, in contrast to the KF literature where this is more of a standard assumption with a rather simple remedy.

Our contribution is to point out that the basic model for the PF should be $p(y_k, x_{k+1}|x_k)$. The corresponding optimal proposal distribution was derived, and the two most common approximations (prior and likelihood, respectively) were stated as special cases. The special case of additive Gaussian noise processes was studied in detail, and the common Bootstrap/SIR PF was modified by a new prediction step.

7 Acknowledgement

S. Saha is supported by grants from the Linnaeus Center for Control, Autonomy, and Decision-making in Complex Systems (CADICS).

References

- [1] T. Kailath, A. H. Sayed and B. Hassibi, *Linear Estimation*, Prentice-Hall, Upper Saddle River, NJ, 2000.
- [2] F. Gustafsson, *Statistical Sensor Fusion*, Studentlitteratur, 2010.
- [3] S. Arulampalam, S. Maskell, N. Gordon and T. Clapp, "A Tutorial on Particle Filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transaction on Signal Processing*, vol. 50(2), pp. 174–188, Feb. 2002.
- [4] J. V. Candy, "Bootstrap particle filtering," *IEEE Signal Processing Magazine*, vol. 24(4), pp. 73–85, 2007.
- [5] O. Cappe, S. J. Godsill and E. Moulines, "An overview of existing methods and recent advances in sequential Monte Carlo," *IEEE Proceedings*, vol. 95(5), pp. 899–924, 2007.
- [6] P. M. Djuric, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo and J. Miguez, "Particle filtering," *IEEE Signal Processing Magazine*, vol. 20(5), pp. 19–38, 2003.
- [7] A. Doucet and A. M. Johansen, "A Tutorial on Particle Filtering and Smoothing: Fifteen years Later," *Handbook of Nonlinear Filtering (eds. D. Crisan et B. Rozovsky)*, Oxford University Press, to appear.
- [8] S. Saha, *Topics in Particle Filtering and Smoothing*, PhD thesis, Univ. of Twente, 2009, ISBN 978-90-365-2864-1.
- [9] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson and P. Nordlund, "Particle Filters for Positioning, Navigation and Tracking," *IEEE Transaction on Signal Processing*, vol. 50(2), Feb. 2002.
- [10] I. Skog and P. Händel, "Effects of time synchronization errors in GNSS-aided INS," *Proceedings of PLANS 2008*, Monterey, CA, pp. 82–88.
- [11] A. Goyal, "Predictability of Stock Return Volatility from GARCH Models," *Working paper*, 2000, available online at www.bus.emory.edu/agoyal/docs/Garch.pdf.