

# PARTICLE FILTERS FOR SYSTEM IDENTIFICATION WITH APPLICATION TO CHAOS PREDICTION

Fredrik Gustafsson and Paul Hriljac

*Department of Electrical Engineering, Linköpings universitet,  
SE-581 83 Linköping, Sweden  
Tel: +46 13 282226; fax: +46 13 282622  
e-mail: fredrik@isy.liu.se  
College of Engineering, Embry-Riddle Aeronautical University,  
3200 Willow Creek Road USA  
e-mail: hriljap@erau.edu*

Abstract: The theory of the particle filter, or sequential Monte Carlo methods, has made substantial progress the last decade. The number of applications has increased substantially the last three years, in particular in navigation and telecommunication areas. In this contribution, we will first point out how the particle filter can be used for system identification, using a quite general problem formulation, and it is pointed out in which kind of application the particle filter can be an attractive alternative to classical system identification methods. This is then demonstrated on prediction of time series arising from chaotic dynamical systems. The specific dynamical system considered is the so called logistical map with an unknown parameter, which belongs to the chaotic regime.

Keywords: system identification, particle filter, dynamical systems, non-linear systems, estimation, chaos, prediction, Cramer-Rao

## 1. INTRODUCTION

Since the seminal paper Gordon et al. (1993a), a large number of papers developing the theory of particle filtering have appeared. State of the art is summarized in the recent monograph Doucet et al. (2001a). During the five years following the seminal paper above, the theory was developed mainly by statisticians. When the theory was mature enough, the signal processing community quickly adopted the results, and developed practical algorithms for applications. Since 2000, it can be noted that the number of publications on applications has increased substantially, in particular in navigation, Gustafsson et al. (2002), and telecommunications areas. Special sessions on particle filter are now frequent on major signal processing conferences. However, so far the impact on the automatic control and system identification societies have been quite limited.

Anyhow, the particle filter offers a general tool for estimating unknown parameters in non-linear models of moderate complexity. A general parametric non-linear state space model, see Chapter 5.3 in Ljung (1999), with additive noise processes is given by:

$$z_{t+1} = f(z_t; \theta_t) + v_t^z, \quad (1a)$$

$$y_t = h(z_t; \theta_t) + e_t, \quad (1b)$$

The functions  $f(z_t; \theta)$  and  $h(z_t; \theta)$  are given by the model and may contain unknown parameters as given by the vector  $\theta$ . Joint state and parameter estimation aims at estimating both the state  $z_t$  and the parameters  $\theta$  simultaneously.

The literature on system identification mainly describes black-box approaches to system identification for non-linear systems. Particular algorithms for special structures as Wiener and Hammerstein models are known. In the general case, with a completely known model structure, prediction error methods can

be applied. These involve differentiation of the non-linearities, and often work quite well when the noise is well-behaved (uni-modal and symmetric) and the non-linearities are smooth. Maximum-likelihood approaches are possible for some particular structures, where the 'certainty equivalence' principle can be applied (iterate between estimating states and parameters) where for instance the EM-algorithm can work well. For general models, including arbitrary non-linearities (discontinuous ones for instance), hard constraints on the parameters and states, or multi-modal noise distributions (aircraft maneuver as state noise, radar lobes as noise distribution *etc.*), no general theory applies or standard methods give poor performance.

To put this in a framework where the particle filter can be applied, the state vector is augmented with the parameter vector. The new state vector is denoted

$$x_t = \begin{pmatrix} z_t \\ \theta_t \end{pmatrix} \quad (2)$$

which is governed by the relations

$$\begin{bmatrix} z_{t+1} \\ \theta_{t+1} \end{bmatrix} = \begin{bmatrix} f(z_t; \theta_t) \\ \theta_t \end{bmatrix} + \begin{bmatrix} v_t^z + w_t^z \\ v_t^\theta + w_t^\theta \end{bmatrix} \quad (3a)$$

$$y_t = h(z_t; \theta_t) + e_t, \quad (3b)$$

Here we have distinguished the physical noise  $v_t$  from the instrumental roughening noise  $w_t$ . As will be described later, this is something needed for the particle filter to explore the whole state space, which fills in the gaps between the finite number of particles. By introducing an additional noise to the samples the depletion problem can be reduced. This technique is called *jittering* Fearnhead (1998) or *roughening* Gordon et al. (1993a). To summarize the stochastic assumptions, the different noise processes in (3) are:

- Physical state noise  $v_t^z$ .
- Roughening state noise  $w_t^z$ , which has turned out to be beneficial for the particle filter performance. Loosely speaking, it helps the particles explore the whole state space.
- Random walk noise  $v_t^\theta$  on the parameters, for making the algorithm adaptive to slow changes in the parameters. In system identification,  $v_t^\theta = 0$ , which we will mostly assume here.
- Roughening parameter noise  $w_t^\theta$ , which makes the particles explore a small neighborhood. This noise should decay with time in system identification, to get a converging estimate.

The particle filter provides an appealing framework for this nonlinear and non-Gaussian estimation problems. The aim of the particle filter is to recursively estimate the posterior density function  $p(X_t|Y_t)$ , where  $X_t = \{x_0, \dots, x_t\}$ . According to the Bayesian philosophy,  $p(X_t|Y_t)$  contains all there is to know about the process at time  $t$ . From this density, we can then obtain any point estimate we like. Loosely speaking the particle filter can be interpreted as a large number of simulations, where each simulation consists of a sample from the distribution we want to estimate.

There is a weight associated with each sample, which contains information on how *likely* the corresponding sample is. These samples together with the corresponding weights will constitute a discrete approximation of the posterior density. Hence, for a sufficiently large number of samples, the particle filter provides a tool to approximate  $p(z_t, \theta_t|Y_t)$  arbitrarily well. However, the large state dimension might be prohibitive for the practical use of the particle filter, and this is its main drawback, besides the obvious demand on sufficient computational resources. As a coarse rule of thumb, the particle filter should not be applied to problems with more than five states.

In an accompanying paper, Schön and Gustafsson (2003), marginalization techniques are applied to possible linear sub-structures in the model, which makes it possible to apply the filter to larger problems and furthermore decreases the requirement on computational power.

## 2. THE PARTICLE FILTER

### 2.1 The idea

The particle filter provides an approximative solution for the problem of recursively estimating the *posterior* density function  $p(X_t|Y_t)$ , for a nonlinear discrete time system on the form (1). In this article we are interested in one of the marginals of the posterior density, namely the *filtering* density,  $p(x_t|Y_t)$ . Using this density function we can then compute an estimate of any inference function  $g(x_t)$  we like, for instance the state estimator  $g(x_t) = z_t$  or the output prediction error variance  $g(x_t) = \text{Var}(y_t - h(z_t; \theta_t))$ . We will use  $I(g(x_t))$  to denote this estimate, according to

$$I(g(x_t)) = E_{p(x_t|Y_t)}[g(x_t)] = \int g(x_t)p(x_t|Y_t)dx_t \quad (4)$$

More specifically, the particle filter provides an approximative solution to the optimal recursive Bayesian filter given by the prediction density,  $p(x_{t+1}|Y_t)$ , and the filtering density,  $p(x_t|Y_t)$ , Jazwinski (1970)

$$p(x_{t+1}|Y_t) = \int p(x_{t+1}|x_t)p(x_t|Y_t) dx_t, \quad (5a)$$

$$p(x_t|Y_t) = \frac{p(y_t|x_t)p(x_t|Y_{t-1})}{p(y_t|Y_{t-1})}. \quad (5b)$$

These equations are in general very hard to solve analytically, except in a few special cases, *i.e.* when the model is linear and the noise is Gaussian. In that case the solution is given by the Kalman filter, see Anderson and Moore (1979). The particle filter provides us with an approximative solution to these integrals by using a large set of samples (also called particles, hence the name particle filter),  $\{x_t^{(i)}\}_{i=1}^N$ , which constitutes a discrete approximation of  $p(x_t|Y_t)$ , according to

$$\hat{p}_N(x_t|Y_t) = \sum_{i=1}^N \bar{q}_t^{(i)} \delta(x_t^{(i)} - x_t), \quad (6)$$

where  $\bar{q}_t^{(i)} = q_t^{(i)} / \sum_{j=1}^N q_t^{(j)}$  are the normalized *importance* weights. These weights are introduced due to the fact that we cannot sample from the true density functions. For a more thorough discussion of these weights see *e.g.* Doucet et al. (2001b); Doucet (1998). These weights are updated using the likelihood function according to

$$q_{t+1}^{(i)} = p(y_{t+1}|x_{t+1}^{(i)}) \bar{q}_t^{(i)}, \quad (7)$$

which means that the most *likely* samples, *i.e.* the samples that correspond to a large likelihood, will be assigned a large weight. There is still one problem that remains to be solved, and that is that the approach described above will lead to that the variance of the importance weights increases over time and thus the estimate will finally diverge, see *e.g.* Doucet (1998) for a formal proof of this fact. What happens is that the samples spread out and the weights will be almost zero for most of the samples. This can be avoided, using *resampling*. This key-step, which made the particle filter work in practice was introduced in Gordon et al. (1993b), based on the weighted bootstrap presented in Smith and Gelfand (1992). The resampling step consists of drawing  $N$  samples with replacement, where the probability of drawing  $X_t^{(i)}$  is given by the corresponding importance weight,  $\bar{q}_t^{(i)}$ . This makes sense since the importance weight will be large if the corresponding sample is close to the true state.

Apart from the resampling step the basic ideas for the particle filter have been around since the 1940:s. The first article, known to the authors, introducing the overall ideas is Metropolis and Ulam (1949). In the automatic control community the ideas were introduced in the late 1960:s by Handschin and Mayne (1969); Handschin (1970), but then they were forgotten again until the late 1980:s, when more computer power became available, and since then there has been a lot of research activity in this area.

## 2.2 Obtaining the estimates

As described above the particle filter provides us with an estimate of the filtering density,  $p(x_t|Y_t)$ , from which we can deduce various point estimates according to (4). An estimate of the mean of the state, can now be obtained by combining (4), with  $g(x_t) = x_t$ , and (6),

$$\begin{aligned} \hat{x}_{t|t} &= E_{p(x_t|Y_t)}[x_t] = \int x_t p(x_t|Y_t) dx_t \\ &\approx \int x_t \hat{p}_N(x_t|Y_t) dx_t \\ &= \int x_t \sum_{i=1}^N \bar{q}_t^{(i)} \delta(x_t^{(i)} - x_t) dx_t = \sum_{i=1}^N \bar{q}_t^{(i)} x_t^{(i)}. \end{aligned}$$

Similarly an estimate of the variance can be obtained using  $g(x_t) = (x_t - \hat{x}_{t|t})(x_t - \hat{x}_{t|t})^T$ .

A numerical approximation to (5) is given in the following algorithm.

### Algorithm 1. The Particle Filter

**Given:** A parametric state space model (3) with known parametric functions  $f(z; \theta)$  and  $h(z; \theta)$ . Prior densities  $p_{z_0}, p_{\theta_0}$  and noise densities  $p_{v_t, z}, p_{e_t}$  and possibly also  $p_{v_t, \theta}$  for adaptive filtering.

**Design parameters:** Number of particles  $N$ . Roughening densities  $p_{w_t, z}$  and  $p_{w_t, \theta}$ .

- (1) *Initialization:* Generate  $x_0^{(i)} \sim p_{x_0}, i = 1, \dots, N$ . Each sample of the state vector is referred to as a *particle*.
- (2) *Measurement update:* Update the weights by the likelihood (more generally, any importance function, see Doucet et al. (2001a)):

$$\begin{aligned} q_t^{(i)} &= q_{t-1}^{(i)} p(y_t|x_t^{(i)}) \\ &= q_{t-1}^{(i)} p_{e_t}(y_t - h(x_t^{(i)}, \theta_t^{(i)})) \end{aligned}$$

for  $i = 1, 2, \dots, N$ , and normalize to  $\bar{q}_t^{(i)} = q_t^{(i)} / \sum_i q_t^{(i)}$ . As an approximation of  $E(x_t|Y_t)$ , take

$$\hat{x}_t \approx \sum_{i=1}^N \bar{q}_t^{(i)} x_t^{(i)}.$$

- (3) *Re-sampling:*
  - (a) *Bayesian bootstrap.* Take  $N$  samples with replacement from the set  $\{x_t^{(i)}, \theta_t^{(i)}\}_{i=1}^N$  where the probability to take sample  $i$  is  $\bar{q}_t^{(i)}$ . Let  $q_t^{(i)} = 1/N$ . This step is also called *Sampling Importance Re-sampling (SIR)*.
  - (b) *Importance sampling.* Only resample as above when the effective number of samples is less than a threshold  $N_{\text{th}}$ ,

$$N_{\text{eff}} = \frac{1}{\sum_i (\bar{q}_t^{(i)})^2} < N_{\text{th}},$$

see Bergman (1999); Doucet et al. (2000); Kong et al. (1994); Liu (1996). Here  $1 \leq N_{\text{eff}} \leq N$ , where the upper bound is attained when all particles have the same weight, and the lower bound when all probability mass is at one particle. The threshold can be chosen as  $N_{\text{th}} = 2N/3$ .

- (4) *Prediction:* Take  $v_t^{z,(i)} \sim p_{v,z}, v_t^{\theta,(i)} \sim p_{v,\theta}, w_t^{z,(i)} \sim p_{w,z}$ , and  $w_t^{\theta,(i)} \sim p_{w,\theta}$ , and simulate

$$\begin{aligned} x_{t+1}^{(i)} &= f(x_t^{(i)}; \theta_t^{(i)}) + v_t^{z,(i)} + w_t^{z,(i)} \\ \theta_{t+1}^{(i)} &= \theta_t^{(i)} + v_t^{\theta,(i)} + w_t^{\theta,(i)} \end{aligned}$$

for  $i = 1, 2, \dots, N$ .

- (5) Let  $t := t + 1$  and iterate to item 2.

The key point with re-sampling is to prevent high concentration of probability mass at a few particles. Without this step, some  $\bar{q}_t^{(i)}$  will converge to 1 and the filter would brake down to a pure simulation. The

re-sampling can be efficiently implemented using a classical algorithm for sampling  $N$  ordered independent identically distributed variables Bergman (1999); Ripley (1988).

The roughening state noise processes are just instrumental, which have no physical counterpart. The rationale is to fill in the gap in between the finite number particles to make sure that the whole state space will be covered. Furthermore, when the particle cloud becomes denser around its true value, the roughening can be decreased. A simple choice is a Gaussian distribution  $p_{w,z} = N(0, (\sigma_{w,z}^2/t)I)$  and  $p_{w,\theta} = N(0, (\sigma_{w,\theta}^2/t)I)$ , whose variance decays to zero. One could here compare to recursive implementations of the least squares method, where the step size in the parameter update decreases as  $1/t$ .

As a generalization, in adaptive algorithms the parameter vector is assumed slowly (compared to the dynamics) time-varying. The step size in the recursive least squares algorithm with forgetting factor or the Kalman filter for parameter estimation decays as  $1/t$  initially in the transient phase, and then converges/fluctuates around a constant value. Algorithm 1 can be made adaptive by having a non-zero parameter noise  $w_t^\theta$ .

### 3. APPLICATION EXAMPLE: CHAOS

The prediction and control of chaotic dynamical systems is a problem relevant to a number of fields in science and engineering Schuster (1995); Kapitaniak (1998); Lorenz (1993); Devaney (1989). Chaotic behavior has been postulated and/or observed in classical mechanics Henon and Heiles (1964), quantum mechanics Gutzwiller (1990), chemical reactions Vidal and Pacault (1981), civil engineering Naschie (1990), electrical circuits Chua (1992), and climatology Vallis (1988). Prediction and understanding of chaotic phenomena are frequently based on approximations of the underlying attractor and its dynamics Farmer and Sidorowich (1987), Casdagli (1989), Sugihara and May (1990), Sauer (1993). This approximation is typically performed by using values of the time series from embedding this attractor in some phase-space. Then approximations of the dynamics on the attractor are obtained in various neighborhoods using various methods, some linear, some not. Two important issues encountered are dealing with noise in the observations and dealing with relatively small amounts of data. These are important issues because the underlying geometry of chaotic attractors is frequently very fine, with varying structure at all levels of magnification. Another method sometimes used is neural networks Lapedes and Farber (1987) and Casdagli and Eubank (1992). The point of this paper is to examine the possible use of a particle filter in this context. In this paper, it is assumed that one has a time series of observations of some potentially chaotic system, and also a model of the underlying system, but with unknown param-

eters. It is further assumed that both the system and the observations have a small amount of error.

The paper in particular studies one of the simplest known models which can produce chaotic behavior, the logistics map,

$$z_{t+1} = \theta z_t(1 - z_t). \quad (8)$$

For  $\theta$  in the the range  $[3.56994568, 4]$  chaotic behaviour is frequently observed Schuster (1995). To reflect the fact that one may have a model of a potentially chaotic system but not the underlying parameters, our paper assumes that  $\theta$  is unknown and includes it as a state variable.

The logistic map can in our notation be formulized as

$$z_{t+1} = \theta z_t(1 - z_t) + v_t \quad (9a)$$

$$y_t = z_t + e_t. \quad (9b)$$

In such an application, the state noise may simply be quantization noise from a finite precision simulation, and the measurement noise may be quantization noise from the sensor. However, we will investigate the case of Gaussian noise in Section 6, to be able to compute the Cramer-Rao lower bound.

The state space form (3) with state  $x_t = (z_t, \theta)^T$  now becomes

$$x_{t+1} = \begin{pmatrix} \theta z_t(1 - z_t) \\ \theta \end{pmatrix} + \begin{pmatrix} v_t^z + w_t^z \\ w_t^\theta \end{pmatrix} \quad (10a)$$

$$y_t = z_t + e_t \quad (10b)$$

The parameter random walk noise  $v_t^\theta = 0$ , since a time constant  $\theta$  is assumed.

### 4. PREDICTION

The simplest idea to predict future states in (1) is to estimate the state  $x_t$  and parameter  $\theta$  using the particle filter, then use these values for a pure simulation of the state equation (1a). However, we get no information of the uncertainty in the prediction in this way. This is a particularly pronounced drawback for a chaotic model as (10). What we want to do is to evolve  $p(x_t|Y_t)$  to  $p(x_{t+k}|Y_t)$  using (1a). One idea to approximate this distribution is to use the particle approximation  $x_t^{(i)} \sim p(x_t|Y_t)$ , and to simulate each particle according to (1a). If this is done without both state and roughening noise  $v_t^z = v_t^\theta = w_t^z = w_t^\theta = 0$ , then a the state prediction with the smallest variance is obtained. On the other hand, keeping the state noise  $v_t$  with no roughening  $w_t^z = w_t^\theta = 0$ , the *a posteriori* distribution  $x_{t+k}^{(i)} \sim p(x_{t+k}|Y_t)$  is approximated, which may be needed in decision theory, statistical approaches to control, and risk calculations.

### 5. CRAMER-RAO LOWER BOUND

The Cramer-Rao lower bound for non-linear filtering takes a quite simple form for Gaussian noise. Intuitively, the best one can hope for is to get the same performance as the Kalman filter on a linearized model,

which happens when the estimation error is small enough so the non-linearities  $f, h$  are approximately linear. One can show formally, see N. Bergman and Gordon (2001), that

$$\text{Cov}(\hat{x}_{t|t}) \geq J^{-1} = P_{t|t},$$

where  $J$  is the Fisher information matrix, for any estimator  $\hat{x}_{t|t}$ . The Cramer-Rao lower bound  $P_{t|t}$  for this case is given by the Ricatti equation

$$\begin{aligned} P_{t|t-1} &= F_t P_{t-1|t-1} F_t^T + Q_t \\ P_{t|t} &= P_{t|t-1} - \\ & P_{t|t-1} H_t^T (H_t P_{t|t-1} H_t^T + R)^{-1} H_t P_{t|t-1}. \end{aligned}$$

For the chaos model (9), we get

$$\begin{aligned} F_t &= \left. \frac{df}{dx} \right|_{x=x_t} = \begin{pmatrix} \theta(1-2z_t) & z_t(1-z_t) \\ 0 & 1 \end{pmatrix} \\ H_t &= \left. \frac{dh}{dx} \right|_{x=x_t} = (1, 0), \end{aligned}$$

which is the associated linear system with (10), evaluated at the trajectory  $x_t$ .

## 6. SIMULATIONS

The chaotic system

$$z_{t+1} = \theta z_t(1 - z_t) + v_t \quad (11a)$$

$$y_t = z_t + e_t \quad (11b)$$

is simulated with 100 samples,  $\theta = 3.95$ ,  $\text{Var}(v_t^z) = 7 \cdot 10^{-14}$ ,  $\text{Var}(e_t) = 8 \cdot 10^{-5}$  and initial state  $x(1) = 0.513$ ,

The particle filter with  $N_p = 1000$ ,  $\text{Var}(w_t^z) = \text{Var}(w_t^\theta) = 10^{-4}$  is applied initialized with uniformly distributed random numbers in  $\theta \in [3.8, 4]$  and  $x(1) \in [0, 1]$ . Figure 1 shows the particle filter output and its resulting simulation. Figure 2 illustrates the estimation error. We see that the proposed filter provides a converging parameter estimate, good state estimates and also gives results comparable to the Cramer-Rao lower bound.

Finally, Figure 3 illustrates how the predictive ability of chaos quickly deteriorates with the horizon  $k$ . Still, better performance than just guessing is achieved up to five time steps ahead, after the initial transient (the 25 first samples are removed here) has vanished. Note that a simulation initialized at any  $x(1)$ , would give a state variance of 0.3. The standard deviation of such a difference is thus  $\sqrt{2}$  larger than the variance of  $x(t)$  itself. That is, since the model is chaotic, long term predictions, even using the true  $\theta$ , tend to be approximately 0.45 in the limit  $k \rightarrow \infty$ , which is the value the curve in Figure 3 converges to.

## 7. CONCLUSIONS

The particle filter shows promise in the problem of simultaneous state and parameter estimation in general, which was exemplified in an application of prediction of chaotic time series. The filter is applicable when:

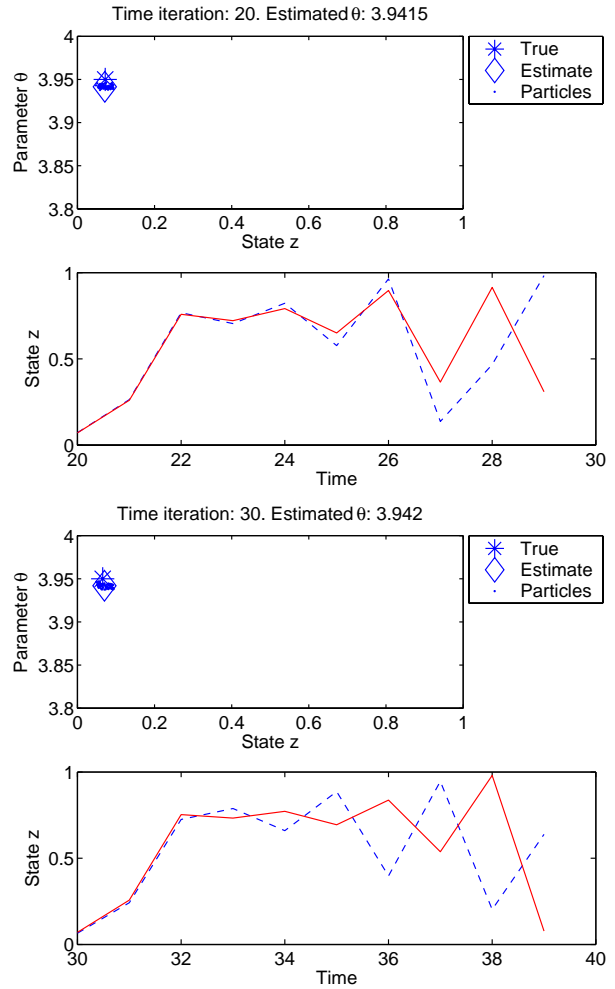


Fig. 1. Particle cloud and particle filter estimate of  $x, \theta$  at time  $i = 20$  and  $i = 30$ , respectively, compared to the true values. Lower plot shows the predictions based on the estimated parameter and state compared to the true state.

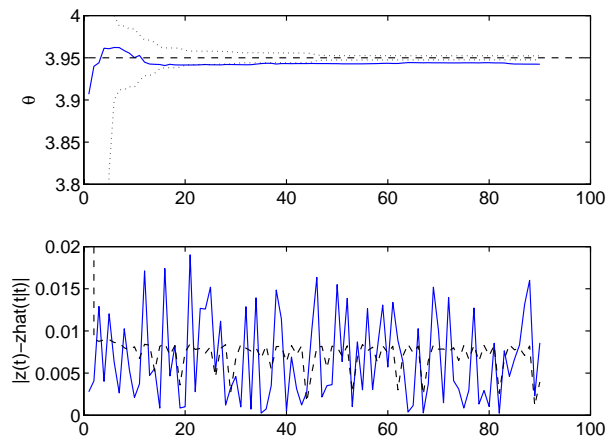


Fig. 2. Estimated  $\hat{\theta}$  with Cramer-Rao confidence interval around true parameter, filter error  $|x_t - \hat{x}_{t|t}|$  compared to Cramer-Rao lower bound.

- the problem is of moderate complexity (not more than, roughly, five states and parameters),
- the computational resources are good enough, compared to the sample interval,

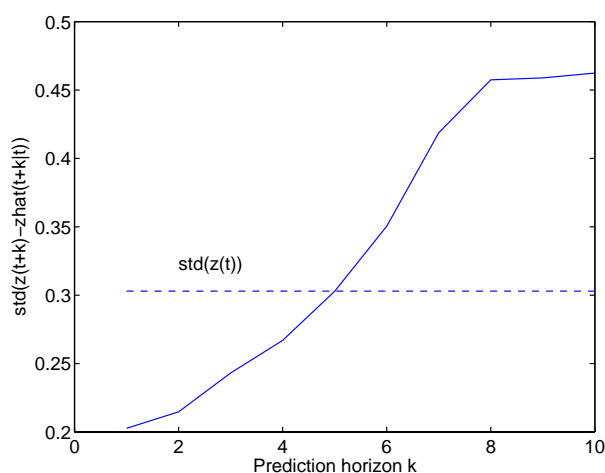


Fig. 3. Prediction error  $\sqrt{E(z_{t+k} - \hat{z}_{t+k|t})^2}$  ( $k=1,2,\dots,10$ ). As a comparison, the state variance itself is 0.3.

- prior knowledge on a region containing the true initial state and parameter vector is available.

Having said these limitation, the particle filter has been demonstrated to be quite a flexible tool, handling all kind of non-linear systems, even including by hard constraints or geographical information systems. Any known noise distribution can be used to model *e.g.* sensor characteristic, state disturbances and prior knowledge on stability *etc.*. For system identification, where a part of the state vector corresponds to the time invariant parameter vector, the known special trick of adding artificial roughening noise is needed. With a carefully decaying roughening noise variance, a converging estimate is obtained.

The principle was demonstrated on a model for chaos, which is non-linear in the states and where the Cramer-Rao lower bound can be calculated.

#### References

B.D.O. Anderson and J.B. Moore. *Optimal Filtering*. Information and system science series. Prentice Hall, Englewood Cliffs, New Jersey, 1979.

N. Bergman. *Recursive Bayesian Estimation: Navigation and Tracking Applications*. Dissertation nr. 579, Linköping University, Sweden, 1999.

M. Casdagli. Nonlinear prediction of chaotic time series. *Physica D*, 35:335, 1989.

M. Casdagli and S. Eubank. *Nonlinear Modeling and Forecasting*. Addison-Wesley, 1992.

L.O. Chua. The genesis of chua's circuit. *Archiv fur Elektronik und Ubertragungstechnik*, 46:250–257, 1992.

Robert Devaney. *An Introduction to Chaotic Dynamical Systems, Second Edition*. Addison-Wesley, 1989.

A. Doucet. On sequential simulation-based methods for Bayesian filtering. Technical Report CUED/F-INFENG/TR.310, Signal Processing Group, Department of Engineering, University of Cambridge, 1998.

A. Doucet, N. de Freitas, and N. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Springer Verlag, 2001a.

A. Doucet, S.J. Godsill, and C. Andrieu. On sequential simulation-based methods for Bayesian filtering. *Statistics and Computing*, 10(3):197–208, 2000.

A. Doucet, N. Gordon, and V. Krishnamurthy. Particle Filters for State Estimation of Jump Markov Linear Systems. *IEEE Transactions on Signal Processing*, 49(3):613–624, 2001b.

J.D. Farmer and J. Sidorowich. Predicting chaotic time series. *Phys. Rev. Lett.*, 59:845, 1987.

P. Fearnhead. *Sequential Monte Carlo methods in filter theory*. PhD thesis, University of Oxford, 1998.

N.J. Gordon, D.J. Salmond, and A.F.M. Smith. A novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEE Proceedings on Radar and Signal Processing*, volume 140, pages 107–113, 1993a.

N.J. Gordon, D.J. Salmond, and A.F.M. Smith. A novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEE Proceedings on Radar and Signal Processing*, volume 140, pages 107–113, 1993b.

F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P.-J. Nordlund. Particle filters for positioning, navigation and tracking. *IEEE Transactions on Signal Processing*, 50(2), February 2002.

M.C. Gutzwiller. *Chaos in Classical and Quantum Mechanics*. Springer, 1990.

J.E. Handschin. Monte carlo techniques for prediction and filtering of non-linear stochastic processes. *a*, 6:555–563, 1970.

J.E. Handschin and D.Q. Mayne. Monte carlo techniques to estimate the conditional expectation in multi-stage non-linear filtering. *International journal of control*, 9:547–559, 1969.

M. Henon and C. Heiles. The applicability of the third integral of the motion: Some numerical results. *Astron. J.*, 69:73, 1964.

A.H. Jazwinski. *Stochastic processes and filtering theory*. Mathematics in science and engineering. Academic Press, New York, 1970.

Tomaz Kapitaniak. *Chaos for Engineers, Theory, Applications, and Control*. Springer-Verlag, 1998.

A. Kong, J. S. Liu, and W. H. Wong. Sequential imputations and Bayesian missing data problems. *J. Amer. Stat. Assoc.*, 89(425): 278–288, 1994.

A. Lapedes and R. Farber. Nonlinear signal processing using neural networks: Prediction and signal modeling. *preprint, Los Alamos*, 1987.

J.S. Liu. Metropolisized independent sampling with comparison to rejection sampling and importance sampling. *Statistics and Computing*, 6:113–119, 1996.

L. Ljung. *System identification, Theory for the user*. Prentice Hall, Englewood Cliffs, NJ, second edition, 1999.

Edward Lorenz. *The Essence of Chaos*. University of Washington Press, 1993.

N. Metropolis and S. Ulam. The Monte Carlo Method. *Journal of the American Statistical Association*, 44(247):335–341, 1949.

A. Doucet N. Bergman and N.J. Gordon. Optimal estimation and cramer-rao bounds for partial non-gaussian state-space model. *Ann. Inst. Stat. Math*, 52(1):97–112, 2001.

M.S. El Naschie. *Stress Stability and Chaos*. McGraw-Hill, 1990.

B.D. Ripley. *Stochastic Simulation*. John Wiley, 1988.

T. Sauer. Time series prediction using delay coordinate embedding. *Time Series Prediction: Forecasting and Understanding the Past*, 1993.

T. Schön and F. Gustafsson. Particle filters for system identification of state-space models linear in either parameters or states. In *SYSID*, Rotterdam, NL, 2003.

Heinz Georg Schuster. *Deterministic Chaos: An Introduction, Third Edition*. VCH, 1995.

A.F.M. Smith and A.E. Gelfand. Bayesian statistics without tears: A sampling-resampling perspective. *The American Statistician*, 46(2):84–88, May 1992.

G. Sugihara and R.M. May. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, 334:734, 1990.

G.K. Vallis. Conceptual models of el nino and the southern oscillation. *J. Geophys. Res.*, 93:13979–14019, 1988.

C. Vidal and A. Pacault. *Nonlinear Phenomena in Chemical Dynamics*. Springer, 1981.