An Insight into the Issue of Dimensionality in Particle Filtering

Paul Bui Quang
ONERA
Châtillon, France.
paul.bui_quang@onera.fr

Christian Musso
ONERA
Châtillon, France.
christian.musso@onera.fr

François Le Gland
INRIA
Rennes, France.
legland@irisa.fr

Abstract – *Particle filtering is a widely used Monte Carlo method to approximate the posterior density in non-linear filtering. Unlike the Kalman filter, the particle filter deals with non-linearity, multi-modality or non Gaussianity. However, recently, it has been observed that particle filtering can be inefficient when the dimension of the system is high. We discuss the effect of dimensionality on the Monte Carlo error and we analyze it in the case of a linear tracking model. In this case, we show that this error increases exponentially with the dimension.*

Keywords: Particle filtering, Bayesian estimation, tracking, curse of dimensionality.

1 Introduction

Particle filters are non-linear/non-Gaussian Bayesian estimation techniques with a wide range of applications, like target tracking [3], [10], navigation [11], geophysics [17], etc. It typically applies on sequential data generated by so-called hidden Markov models.

It is well known that the Kalman filter gives optimal estimation when the model is linear and Gaussian. Versions of the Kalman filter that can handle non-linear and/or non-Gaussian models have been developped, such as the extended Kalman filter (EKF) and the unscented Kalman filter (UKF) [12]. In these Kalman-like methods, the posterior density is approximated by its first two moments.

Unlike the EKF and the UKF, the particle filter estimates sequentially the posterior density by a discrete density, and not only by its approximated mean and covariance. The discrete density is computed thanks to Monte Carlo (MC) sampling methods. The strength of particle methods is that they can handle highly non-linear non-Gaussian models.

Recently, it has been observed that particle filtering is not efficient when applied to a high dimensional estimation problem [2], [6], [17]. The purpose of this paper is to study theoretically the behaviour of particle filter in case of a high-dimensional linear system.

The outline of the paper is the following. In section 2, we shortly present the particle filter and recall a convergence result of the particle estimator which gives an approximation of the root mean squared error (RMSE). In section 3, the degradation of this convergence in a high-dimensional state space is discussed from the view point of the RMSE. The main contribution of the paper is the section 4. In this section we analyze precisely the influence of the dimension on a linear target tracking model. We show that the particle filter error increases exponentially with the state dimension. Some numerical simulations illustrate these theoretical results.

2 Particle filtering

The typical Bayesian models on which particle filtering methods can be applied are non-linear/non-Gaussian hidden Markov models. Such models involve a hidden (i.e. unobserved) Markov chain of states $(X_n, n \geq 0)$ (denoted by $X_{0:n}$) and of a sequence of measurements $(Y_n, n \geq 1)$ (denoted by $Y_{1:n}$). The states transition law

$$X_n | X_{n-1} = x_{n-1} \sim f(x_n | x_{n-1})$$

is known (throughout the paper, probability laws are characterized by their densities). The measurements are independent conditionally to the state process, with known distribution

$$Y_n | X_n = x_n \sim h(y_n | x_n).$$

The objective of filtering is to estimate at each time step n the law $p(x_n | y_{1:n})$ of the hidden state conditionally to the past observations, which is the Bayesian posterior density.

At each time step n, a particle filter iteration consists in a prediction step and a correction step. In the prediction step, a set of particles $\xi_n^1, \ldots, \xi_n^N$ is sampled from a given importance distribution (also called proposal distribution) that depends on the past estimates. In the correction step, a weight w_i is assigned to each particle ξ_n^i. This weight depends on the likelihood $h(y_n | \xi_n^i)$.

The empirical density

$$\hat{p}_n^N(x_n) = \sum_{i=1}^{N} w_i \delta_{\xi_n^i}(x_n) \quad (1)$$

is a discrete estimate of the posterior density $p_n(x_n) = p(x_n|y_{1:n})$. Particle filtering has been introduced in [10]. It can be seen as a sequential application of importance sampling, that is why it belongs to the so-called sequential Monte Carlo methods [8].

A known weakness of particle filtering is weight degeneracy [8], or weight collapse. The variance of the weights can only increase over time. In practice, after a few time iterations, all but one particle have a very low weight and the particle density does not fit well anymore the true posterior. A means to avoid weight degeneracy is to add a resampling step after the correction step. All the particles are resampled with a probability equal to their weight, which discards low weight particles and regenerates high weight particles. The sequential importance resampling (SIR) particle filter, that includes such a resampling step, is presented in Table 1. Generally, resampling is done only when weight degeneracy is severe, which can observed with the effective sample size criterion [1].

$n = 0$	sample $\xi_0^i \sim f(x_0)$ for $i = 1, \ldots, N$
$n \geq 1$	**for** $i = 1, \ldots, N$
	sample $\xi_n^i \sim f(x_n|\xi_{n-1}^i)$ $(i = 1, \ldots, N)$
	calculate $\tilde{w}_i = h(y_n|\xi_n^i)$
	end
	calculate $s = \sum_{j=1}^N \tilde{w}_j$
	for $i = 1, \ldots, N$
	normalize $w_i = \tilde{w}_i / s$
	end
	resample (ξ_n^i, w_i)

Table 1: SIR particle filter

Improved particle filters have been developpeded in the past few years, such as the Rao-Blackwellized particle filter, the regularized particle filter, the auxiliary particle filter, etc. (see [1] for a review of these algorithms).

Theoretical results insure the convergence of the particle density (1) to the real posterior density when the number of particles N goes to infinity [5], [7]. The following classical result formulates the convergence of the RMSE:

$$E\left[(\langle \hat{p}_n^N, \phi \rangle - \langle p_n, \phi \rangle)^2\right]^{1/2} \leq c_n \frac{\|\phi\|}{\sqrt{N}} \quad (2)$$

where $\langle \mu, \phi \rangle = \int \phi(x)\mu(dx)$ and $\|\phi\| = \sup_x \phi(x)$, with ϕ a bounded measurable test function. Expectation is taken with respect to the particles distribution. The convergence rate of the RMSE w.r.t. the number of particles is thus $N^{-1/2}$. The upper bound (2) does not show explicit dependency on the dimension of the hidden state. However, authors have noticed that the particle filter behaves poorly when the dimension increases [2], [5], [6], [17]. The aim of the following work is to give some more insight into the behaviour of particle approximation in a high-dimensional state space.

3 The effect of dimensionality

3.1 The particle approximation error

Let d denote the dimension of the system state X_n. Although d does not explicitly appear in the upper bound in (2), authors in [5], [6] have underlined that the dependency on d of the constant c_n can be strong.

To study the impact of the dimension, let us get out of the sequential data framework and consider that a whole trajectorial data batch is available simultaneously. This will allow us to simplify calculation without considerable loss of generality. One observes simultaneously the data batch $Y_{1:n}$ and wants to estimate the hidden state X_n. In this non-sequential context, particle filtering simply becomes importance sampling applied to Bayesian inference. Let $h(y_{1:n}|x)$ be the likelihood and $f(x)$ the prior density. Suppose $f(x)$ is chosen as the importance density. We want to estimate the posterior density $p(x|y_{1:n})$ and we approximate it with the empirical density

$$\hat{p}^N(x) = \sum_{i=1}^{N} w_i \delta_{\xi^i}(x)$$

where the particles ξ^i are independent and identically distributed (i.i.d.) according to the importance distribution

$$\xi^i \sim f(x),$$

and where the normalized weights w_i are calculated with the likelihood:

$$w_i = \frac{h(y_{1:n}|\xi^i)}{\sum_{j=1}^N h(y_{1:n}|\xi^j)}. \quad (3)$$

The RMSE convergence can here be bounded [14] as

$$E\left[(\langle \hat{p}^N, \phi \rangle - \langle p, \phi \rangle)^2\right]^{1/2} \leq \frac{c_0}{\sqrt{N}} I(f,h)\|\phi\| \quad (4)$$

where

$$I(f,h) = \frac{\sup_x h(y_{1:n}|x)}{\int_{\mathbb{R}^d} h(y_{1:n}|x) f(x) dx} \quad (5)$$

and where c_0 is a constant independent of d.

Throughout the paper, we consider that the term $I(f,h)$ characterizes the MC error. Specifically, when the integral $\int_{\mathbb{R}^d} h(y_{1:n}|x) f(x) dx$ tends towards 0, the MC error increases. Hereafter, we analyze this term as a function of the dimension. In a way, this integral represents the discrepancy between the prior density $f(x)$ and the likelihood $h(y_{1:n}|x)$ in the state space $\mathbb{R}^d$. Discrepancy between the prior and the likelihood is a classical problem in importance sampling: when it occurs, the particles ξ^i sampled from $f(x)$ fall on the tails of $h(y_{1:n}|x)$ (taken as a function of x), and therefore most of the weights $w_i \propto h(y_{1:n}|\xi_i)$ are close to 0. Numerically, this implies that $h(y_{1:n}|x)$ is small when $f(x)$ is high, thus making the integral $\int h(y_{1:n}|x)f(x)dx$ small and enlarging the bound in (4). In section 4, we show in a linear case that this upper bound grows exponentially with the state dimension by studying the term $I(f,h)$.

We assume that $I(f,h)$ is a relevant indicator of the difficulty of the estimation problem. Another criterion for evaluating the performance of particle filtering is the variance of the normalized weights (3) (w.r.t. to the prior density $f(x)$). Using the Delta method [13], we have for a large N that

$$Var(w_i) \approx \frac{1}{N^2}\left(\frac{\int_{\mathbb{R}^d} h(y_{1:n}|x)^2 f(x)dx}{\left(\int_{\mathbb{R}^d} h(y_{1:n}|x)f(x)dx\right)^2} - 1\right)$$

for all $i = 1,\dots,N$. The term $\frac{\int_{\mathbb{R}^d} h(y_{1:n}|x)^2 f(x)dx}{\left(\int_{\mathbb{R}^d} h(y_{1:n}|x)^2 f(x)dx\right)^2}$ can be shown to increase with d at the same rate $I(f,h)$ does in our case study in section 4. Therefore, in this study, dimensionality actually aggravates weight degeneracy.

Moreover, the ratio $I(f,h)$ can also be seen as inverse of the probability of acceptation in the rejection algorithm [16]. The rejection algorithm is used to sample from the posterior distribution $p(x|y_{1:n})$ by sampling from the prior distribution $f(x)$ and accepting each realization with probability $p_a = \frac{\int h(y_{1:n}|x)f(x)dx}{\sup_x h(y_{1:n}|x)} = \frac{1}{I(f,h)}$. Consequently, when importance sampling behaves poorly ($I(f,h)$ is large), the rejection algorithm is slow (p_a is small). $I(f,h)$ thus characterizes the accuracy of the MC approximation.

3.2 Weight collapse

The problem raised by high-dimensional systems has been studied in the very last years from the viewpoint of weight collapse. Authors in [2], [17] make the data size n and the state dimension d grow simultaneously to infinity, which allow them to do a Gaussian approximation of the measurement model. Thanks to this approach, they show that the number of particles N needs to increase exponentially in d to avoid weight collapse.

In the next section, we rather focus on the upper bound in (4) to study the degradation of particle methods in high-dimensional state space. We consider a simple model to exhibit the behavior of the bound when d increases from 1 to some fixed data size n.

4 Analysis of the MC error in a linear framework

To analyze the influence of the state dimension on particle filtering, let us consider a simple target tracking model. This model will allow us to calculate explicitly the upper bound (4). Let x_k be the position of the target at time k. The target follows a deterministic dynamical model defined by a $(d-1)$th order Taylor expansion of the motion equation

$$x_k = x_{k-1} + \Delta\dot{x}_{k-1} + \cdots + \frac{\Delta^{d-1}}{(d-1)!}x_{k-1}^{(d-1)} \qquad (6)$$

for $k \geq 1$, with the initial state $(x_0, \dot{x}_0, \dots, x_0^{(d-1)})^T$. $x_k^{(l)}$ denotes the lth derivative of the target position. The d-dimensional state of the system at each time k is defined as $X_k = (x_k, \dot{x}_k, \dots, x_k^{(d-1)})^T$. Let F denote the $d \times d$ state transition matrix. The target dynamics is

$$X_k = FX_{k-1}$$

where

$$F = \begin{pmatrix} 1 & \Delta & \frac{\Delta^2}{2} & \cdots & \frac{\Delta^{d-1}}{(d-1)!} \\ 0 & 1 & \Delta & \cdots & \frac{\Delta^{d-2}}{(d-2)!} \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}.$$

The data consist in n observations. At each time step $k = 1,\dots,n$, a sensor measures the current position. The measurement equation is thus

$$y_k = x_k + \epsilon_k \qquad (7)$$

for $k = 1,\dots,n$ where $\epsilon_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\sigma_R^2)$ is a Gaussian measurement noise. The y_k are 1-dimensional. Let us define $Y_{1:n} = (y_1,\dots,y_n)^T$ the trajectorial data. The number of measurements n is taken greater than the state dimension d, in order to insure that the state X_k is observable.

The hidden state to estimate is the state of the target at initial time 0, that is

$$X_0 = (x_0, \dot{x}_0, \dots, x_0^{(d-1)})^T.$$

The continuous version of the dynamics gives for $0 \leq t \leq n$:

$$\dot{X}_t = \begin{pmatrix} \dot{x}_t \\ \ddot{x}_t \\ \vdots \\ x_t^{(d-1)} \\ x_t^{(d)} \end{pmatrix} = AX_t \qquad (8)$$

where A is a $d \times d$ matrix such that

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

The solution of (8) is then $X_t = \exp(At)X_0$. By discretizing this solution, one obtains

$$X_n = \exp(n\Delta A)X_k = F^n X_0$$

for all $k = 0, \ldots, n$. Using the fact that $A^d = 0$ and denoting $L_1(\cdot)$ the linear operator that associates to a matrix its first line, we can re-write x_k as

$$\begin{aligned} x_k &= L_1(X_k) \\ &= L_1(\exp(n\Delta A)X_0) \\ &= L_1\left(\sum_{j=0}^{\infty} \frac{(k\Delta)^j A^j}{j!}\right) X_0 \\ &= \sum_{j=0}^{d-1} L_1\left(\frac{(k\Delta)^j A^j}{j!}\right) X_0 \\ &= \sum_{j=0}^{d-1} \frac{(k\Delta)^j}{j!} x_0^{(j)} \\ &= \left(1, k\Delta, \ldots, \frac{(k\Delta)^{d-1}}{(d-1)!}\right) \begin{pmatrix} x_0 \\ \vdots \\ x_0^{(d-1)} \end{pmatrix} \end{aligned}$$

The above calculation leads to a compact formulation of the measurement model (7):

$$Y_{1:n} = HX_0 + \varepsilon \tag{9}$$

where

$$H = \begin{pmatrix} 1 & \Delta & \cdots & \frac{\Delta^{d-1}}{(d-1)!} \\ 1 & 2\Delta & \cdots & \frac{(2\Delta)^{d-1}}{(d-1)!} \\ \vdots & & & \vdots \\ 1 & n\Delta & \cdots & \frac{(n\Delta)^{d-1}}{(d-1)!} \end{pmatrix} \tag{10}$$

is the $n \times d$ measurement matrix and where $\varepsilon \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, R)$, with $R = \sigma_R^2 I_n$.

To get a Bayesian model, one needs a prior distribution on the hidden state X_0. The prior density is chosen as

$$X_0 \sim \mathcal{N}(m, Q) \tag{11}$$

with $Q = \sigma_Q^2 I_d$, and where m and σ_Q are known hyperparameters. For the sake of calculation simplicity, we chose to estimate the initial state, which is equivalent to estimate X_k at any time $k = 0, \ldots, n$ since the dynamics (6) is deterministic.

We have at our disposal a non-sequential Bayesian model, that enables us to make a case study on how the particle approximation upper bound (4) is degraded when the state dimension d increases from 1 to n.

4.1 Relationship between the MC error and the dimension

In our target tracking model (9)–(11), the initial state model is $X_0 \sim f(x)$ where the prior density is

$$f(x) = \frac{\exp\left[-\frac{1}{2}(x-m)^T Q^{-1}(x-m)\right]}{(2\pi)^{d/2}\sqrt{det(Q)}}$$

and the measurement model is $Y_{1:n}|X_0 = x \sim h(y_{1:n}|x)$, with the likelihood

$$h(y_{1:n}|x) = \frac{\exp\left[-\frac{1}{2}(y_{1:n} - Hx)^T R^{-1}(y_{1:n} - Hx)\right]}{(2\pi)^{n/2}\sqrt{detR}}$$

Let us calculate the upper bound of the particle approximation error (4)

$$\frac{c_0}{\sqrt{N}} I(f,h)\|\phi\|$$

where

$$I(f,h) = \frac{\sup_x h(y_{1:n}|x)}{\int_{\mathbb{R}^d} h(y_{1:n}|x)f(x)dx}.$$

As we are in presence of a Gaussian linear model, the Kalman equations arise naturally. Let

$$\begin{aligned} S &= HQH^T + R \\ K &= QH^T S^{-1} \\ \hat{m} &= m + K(Y_{1:n} - Hm) \\ P &= (I_d - KH)Q \end{aligned}$$

where S is the innovation covariance matrix, K is the gain matrix, $\hat{m}$ is the posterior mean estimator and P is the covariance of m. Note that, since the Kalman filter provides minimum mean squared error estimator, P is the posterior Cramér-Rao bound associated with the parameter m. Straightforward calculations give us the desired ratio (5) (see Appendix A):

$$I(f,h) = e^{\frac{1}{2}B}\sqrt{\frac{det(Q)}{det(P)}} \tag{12}$$

where

$$B = (\hat{m}-m)^T Q^{-1}(\hat{m}-m) + (y_{1:n} - H\hat{m})^T R^{-1}(y_{1:n} - H\hat{m}).$$

The following lemma is useful in a linear Gaussian context.

Lemma 1. *Let Q and R be two definite positive matrix, with dimension d and n respectively, and let H be a $d \times n$ matrix. Then,*

$$(H^T R^{-1} H + Q^{-1})^{-1} = Q - QH^T(HQH^T + R)^{-1}HQ.$$

The next proposition gives a relation between the MC error and the dimension in our model, through a lower bound of the term $I(f,h)$ where d explicitly appears.

Proposition 1. *For all d such that $2 \le d \le n$,*

$$I(f,h) \ge e^{\frac{1}{2}B}\sqrt{1 + \frac{\sigma_Q^2}{\sigma_R^2}\frac{e^{\alpha(d-1)}}{2\pi(2d-1)e^{\frac{1}{6}}}}$$

where

$$\begin{aligned} \alpha &= 2\min(\Delta,1)\left(1 + \log\max\left(1, \frac{n\Delta}{d-1}\right)\right) \\ &\ge 2\min(\Delta,1). \end{aligned}$$

Proof. Using lemma 1, we can write the posterior covariance matrix P as

$$P = (H^T R^{-1} H + Q^{-1})^{-1} = \left(\frac{1}{\sigma_R^2} H^T H + \frac{1}{\sigma_Q^2} I_d \right)^{-1}.$$

Since H is full rank, $H^T H$ is definite positive. Let $\lambda_1, \dots, \lambda_d$ be its positive eigen values. We have

$$\begin{aligned}
\frac{det(Q)}{det(P)} &= det\left(\frac{1}{\sigma_R^2} H^T H + \frac{1}{\sigma_Q^2} I_d \right) det\left(\sigma_Q^2 I_d\right) \\
&= \frac{1}{(\sigma_Q^2)^d} det\left(\frac{\sigma_Q^2}{\sigma_R^2} H^T H + I_d \right) (\sigma_Q^2)^d \\
&= \left(1 + \frac{\sigma_Q^2}{\sigma_R^2}\lambda_1\right) \times \cdots \times \left(1 + \frac{\sigma_Q^2}{\sigma_R^2}\lambda_d\right) \\
&= 1 + \frac{\sigma_Q^2}{\sigma_R^2} Tr(H^T H) + \cdots \\
&\quad + \left(\frac{\sigma_Q^2}{\sigma_R^2}\right)^d det(H^T H) \\
&\geq 1 + \frac{\sigma_Q^2}{\sigma_R^2} Tr(H^T H).
\end{aligned}$$

Consequently, $I(f,h)$ is bounded from below in the following way:

$$I(f,h) \geq e^{\frac{1}{2}B} \sqrt{1 + \frac{\sigma_Q^2}{\sigma_R^2} Tr(H^T H)}. \tag{13}$$

In Appendix B, we show that

$$Tr(H^T H) \geq \frac{e^{\alpha(d-1)}}{2\pi(2d-1)e^{\frac{1}{6}}}$$

for any d, so that

$$\begin{aligned}
I(f,h) &\geq e^{\frac{1}{2}B} \sqrt{1 + \frac{\sigma_Q^2}{\sigma_R^2} \frac{e^{\alpha(d-1)}}{2\pi(2d-1)e^{\frac{1}{6}}}} \qquad (14) \\
&\geq e^{\frac{1}{2}B} \frac{\sigma_Q e^{\frac{\alpha}{2}(d-1)}}{\sigma_R \sqrt{2\pi(2d-1)} e^{\frac{1}{12}}}.
\end{aligned}$$

□

Since $B \geq 0$, $e^{\frac{1}{2}B} \geq 1$ and the upper bound (4) grows with d faster than

$$c_0 \frac{\sigma_Q e^{\frac{\alpha}{4}(d-1)}}{\sigma_R \sqrt{2\pi} e^{\frac{1}{12}}} \|\phi\| \tag{15}$$

where the coefficient α is larger than $2\min(\Delta, 1) > 0$ uniformly in d.

We observe that when the state noise is close to 0 ($\sigma_Q \approx 0$), the bound (15) does not depend anymore on d. The next proposition gives the behavior of $I(f,h)$ in this case.

Proposition 2.

$$I(f,h) \underset{\sigma_Q \to 0}{\longrightarrow} e^{\frac{1}{\sigma_R^2}(y_{1:n} - Hm)^T (y_{1:n} - Hm)}$$

Proof. Recall

$$I(f,h) = e^{\frac{1}{2}B} \sqrt{\frac{det(Q)}{det(P)}}.$$

From the proof of Proposition 1, we have that

$$\frac{det(Q)}{det(P)} \underset{\sigma_Q \to 0}{\longrightarrow} 1.$$

Moreover, B is defined as

$$B = (\hat{m} - m)^T Q^{-1} (\hat{m} - m) + (y_{1:n} - H\hat{m})^T R^{-1} (y_{1:n} - H\hat{m}).$$

Recall $Q = \sigma_Q^2 I_d$ and $R = \sigma_R^2 I_n$. We have

$$\begin{aligned}
\hat{m} - m &= K(y_{1:n} - Hm) \\
&= \sigma_Q^2 H^T (\sigma_Q^2 H H^T + \sigma_R^2 I_n)^{-1} (y_{1:n} - Hm) \\
&\underset{\sigma_Q \to 0}{\sim} \frac{\sigma_Q^2}{\sigma_R^2} H^T (I_n - \frac{\sigma_Q^2}{\sigma_R^2} H H^T)(y_{1:n} - Hm),
\end{aligned}$$

so that

$$(\hat{m} - m)^T Q^{-1} (\hat{m} - m) \underset{\sigma_Q \to 0}{\longrightarrow} 0.$$

Besides, $\hat{m} \underset{\sigma_Q \to 0}{\longrightarrow} m$. Therefore,

$$B \underset{\sigma_Q \to 0}{\longrightarrow} \frac{1}{\sigma_R^2} (y_{1:n} - Hm)^T (y_{1:n} - Hm).$$

□

4.2 Discussion

In our target tracking example, we can assert by Proposition 1 that increasing the dimension degrades drastically the particle approximation. More precisely, the MC error grows exponentially with the dimension faster than the rate $O\left(\frac{e^{\frac{\alpha}{4}d}}{\sqrt{N}}\right)$. Consequently, to maintain a given MC accuracy, the number of particles must increase exponentially.

By Proposition 2, we have that $I(f,h)$ has no more dependency on d when $\sigma_Q \to 0$. That is, when the prior density $f(x)$ converges weakly to a Dirac measure on $x = m$, the MC error does not depend on the dimension, but only on the distance between the true measure $y_{1:n}$ and the predicted measure Hm. The MC error increases exponentially with this distance.

Another remark can be done about the measurement noise. From (15), we see that the MC error grows when $\sigma_R \to 0$. Indeed, when the likelihood is narrow, almost no particles have a sufficiently large weight, thus degrading the particle approximation.

Note that the particle approximation is also getting weaker when the number of observations n increases.

The Cramér-Rao bound P, which decreases with n, appears on the denominator in equation (12). More precisely, calculations in Appendix B show that, for a fixed dimension d, the MC error grows with n at rate $O(n^d)$. The explication is that, for a large n, the likelihood is narrow (the measurement noise variance is roughly $\frac{\sigma_R^2}{n}$) so that the weight degeneracy is severe, just as it has been explained above in the case where $\sigma_R \to 0$.

4.3 Numerical results

For several values of the prior covariance σ_Q^2, Figure 1 shows the bevahiour of the "true" MC error upper bound $I(f,h)$ (12) and of the bounds (13) and (14) we found out in Proposition 1. For these simulations, we chose $\sigma_R = 1$ as the standard deviation of the measurement noise, $\Delta = 1$ as the time period, and $m = (1,\ldots,1)^T \in \mathbb{R}^d$ as the mean value for the prior distribution of X_0. We set σ_Q to 0.1, 0.01 and 0.001 successively. We increase the dimension from $d = 2$ to $d = 10$. The number of observations is $n = 10$.

We observe the expected increase of the MC error with d at exponential rate, which confirms our derivations. Moreover, when σ_Q is small, the dependency on the dimension vanishes as explained in section 4.2.

In Figure 2, we increase further the dimension. It shows $I(f,h)$ and its lower bounds (13) and (14) for $d = 2,\ldots,20$ with $n = 20$. σ_Q is set to 1 and Δ to 0.5. The other parameters remain the same as in Figure 1.

The vertical scale on Figures 1 and 2 is logarithmic. Note that the curves on both figures seem to increase less than linearly. This is due to the fact that the coefficient α in front of d in Proposition 1 actually depends on the dimension. However, α remains greater than a strictly positive value for all d ($\alpha \geq 2\min(\Delta, 1) > 0$), which insures that the growth of the bounds is at least exponential.

5 Conclusion

Using a target tracking example, we established that the performance of the particle filter is drastically weakened when the state dimension is high. More precisely, in a linear context, the MC error increases exponentially with the dimension. The linear assumption allows explicit derivations of the MC error. We expect that this result can be generalized to a non-linear framework. Indeed, by using a local linearization of the joint distribution $h(y_{1:n}|x)f(x)$ around the maximum a posteriori estimate, one could proceed as in section 4.

Several solutions can be considered in order to attenuate this curse of dimensionality. A known solution for improving importance sampling is to use information about the observations to sample the particles. This means that one does not simply use the prior $f(x)$ as a proposal distribution, but preferably some distribution $\pi(x|y_{1:n})$, where the data is taken into account. However, the distribution $\pi(x|y_{1:n})$ can be hard to obtain in general [8]. In the case where the likelihood is very narrow (σ_R in (7) is small), an alternative solution is to use the progressive correction algorithm [15]. Both of these techniques allow the reduction of $I(f,h)$ but the issue of dimensionality remains.

Rao-blackwellization on the other hand uses particle approximation on a reduced state vector [1]. Another promising tool is the notion of effective dimension [4]: retaining dimensions of interest only and discarding non-informative ones could attenuate the problem of state dimensionality.

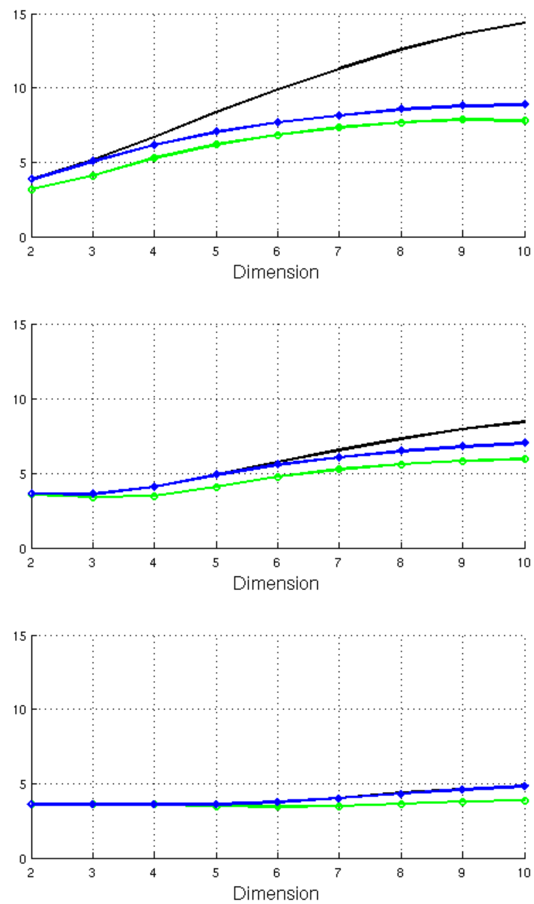

Figure 1: $I(f,h)$ (black solid line) and its lower bounds (13) (diamond blue line) and (14) (dotted green line) for $\sigma_Q = 0.1$ (top), $\sigma_Q = 0.01$ (middle), and $\sigma_Q^2 = 0.001$ (bottom). The vertical scale is logarithmic.

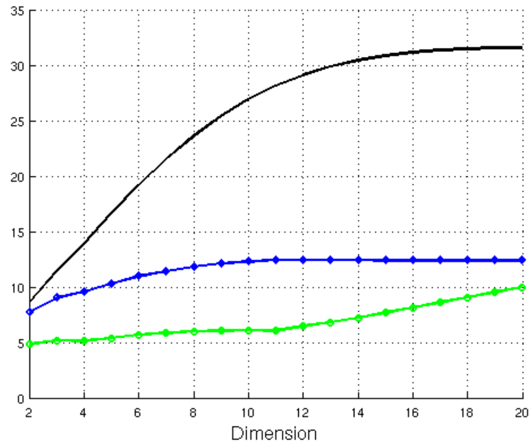

Figure 2: The MC error behaviour from $d = 2$ to $d = 20$ with $\sigma_Q = 1$. The vertical scale is logarithmic.

Appendix A

The integral of interest is

$$\begin{aligned}
&\int_{\mathbb{R}^d} h(y_{1:n}|x) f(x) dx \\
&= \frac{1}{(2\pi)^{\frac{d+n}{2}} \sqrt{det(R)det(Q)}} \\
&\times \int_{\mathbb{R}^d} \exp\Big[-\frac{1}{2}((y_{1:n} - Hx)^T R^{-1}(y_{1:n} - Hx) + \\
&(x-m)^T Q^{-1}(x-m)\Big] dx \\
&= \frac{e^{-\frac{1}{2}B}}{(2\pi)^{\frac{d+n}{2}} \sqrt{det(R)det(Q)}} \\
&\times \int_{\mathbb{R}^d} \exp\Big[-\frac{1}{2}(x-\hat{m})^T P^{-1}(x-\hat{m})\Big] dx
\end{aligned}$$

where $\hat{m}$ and P are defined in section 3.1 and where B is such that

$$\begin{aligned}
&(x-\hat{m})^T P^{-1}(x-\hat{m}) + B \\
&= (y_{1:n} - Hx)^T R^{-1}(y_{1:n} - Hx) \\
&+ (x-m)^T Q^{-1}(x-m)
\end{aligned}$$

for all $x \in \mathbb{R}^d$. By setting $x = \hat{m}$ we get

$$B = (\hat{m}-m)^T Q^{-1}(\hat{m}-m) + (y_{1:n} - H\hat{m})^T R^{-1}(y_{1:n} - H\hat{m}).$$

Finally, since

$$\int_{\mathbb{R}^d} \exp\Big[-\frac{1}{2}(x-\hat{m})^T P^{-1}(x-\hat{m})\Big] dx = (2\pi)^{\frac{d}{2}} \sqrt{det(P)}$$

and

$$\sup_x h(y_{1:n}|x) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{det(R)}},$$

we have

$$\frac{\sup_x h(y_{1:n}|x)}{\int_{\mathbb{R}^d} h(y_{1:n}|x) f(x) dx} = \sqrt{\frac{det(Q)}{det(P)}} e^{\frac{1}{2}B}.$$

Appendix B

Let us bound from below $Tr(H^T H)$ with a quantity that grows up with d.

Knowing $H_{ij} = \frac{(i\Delta)^{j-1}}{(j-1)!}$ (see (10)), we have

$$Tr(H^T H) = \sum_{i=1}^{d} \frac{\Delta^{2i-2}}{(i-1)!^2} \sum_{k=1}^{n} k^{2i-2}.$$

Noticing

$$\sum_{k=1}^{n} k^{2i-2} \geq \int_0^n x^{2i-2} dx = \frac{n^{2i-1}}{2i-1},$$

we have

$$Tr(H^T H) \geq \frac{n}{2d-1} \sum_{i=1}^{d} \frac{(n\Delta)^{2i-2}}{(i-1)!^2}$$

The sequence $\left(\frac{(n\Delta)^{2i}}{i!^2}\right)_{i \geq 0}$ reaches its maximum for $i = \lfloor n\Delta \rfloor$ ($\lfloor \cdot \rfloor$ is the floor function). Therefore,

$$\sum_{i=0}^{d-1} \frac{(n\Delta)^{2i}}{i!^2} \geq \frac{(n\Delta)^{2m}}{m!^2}$$

where $m = \min(\lfloor n\Delta \rfloor, d-1)$.

Suppose $d \geq 2$. Knowing the following inequality for all integer m [9]:

$$\frac{m!}{\sqrt{2\pi} m^{m+\frac{1}{2}} e^{-m}} \leq e^{\frac{1}{12m}} \leq e^{\frac{1}{12}},$$

we have

$$\begin{aligned}
\sum_{i=1}^{d} \frac{(n\Delta)^{2i-2}}{(i-1)!^2} &\geq \frac{e^{2m\log(n\Delta)} e^{-\frac{1}{6}}}{2\pi m^{2m+1} e^{-2m}} \\
&= \frac{e^{2m(\log(n\Delta) - \log m + 1) - \frac{1}{6}}}{2\pi m}.
\end{aligned}$$

We observe that

$$\begin{aligned}
\log(n\Delta) - \log m + 1 &= 1 + \log \frac{n\Delta}{\min(\lfloor n\Delta \rfloor, d-1)} \\
&\geq 1 + \log \max\left(1, \frac{n\Delta}{d-1}\right),
\end{aligned}$$

then

$$\sum_{i=1}^{d} \frac{(n\Delta)^{2i-2}}{(i-1)!^2} \geq \frac{e^{2m\left(1 + \log\max\left(1, \frac{n\Delta}{d-1}\right)\right) - \frac{1}{6}}}{2\pi m}.$$

Since $m \geq (d-1)\min(\Delta, 1)$,

$$\begin{aligned}
Tr(H^T H) &\geq \frac{(n-1)}{2d-1} \frac{e^{2m\left(1 + \log\max\left(1, \frac{n\Delta}{d-1}\right)\right) - \frac{1}{6}}}{2\pi m} \\
&\geq \frac{e^{2(d-1)\min(\Delta,1)\left(1 + \log\max\left(1, \frac{n\Delta}{d-1}\right)\right) - \frac{1}{6}}}{2\pi(2d-1)}
\end{aligned}$$

where the last inequality comes from $n \geq d$.

References

[1] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, 2002.

[2] T. Bengtsson, P. Bickel, and B. Li. Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. *Probability and Statistics: Essays in Honor of David A. Freedman*, 2:316–334, 2008.

[3] Y. Boers, H. Driessen, and K. Grimmerink. A particle-filter-based detection scheme. *IEEE Signal Processing Letters*, 10(10):300–302, 2003.

[4] R.E. Caflisch, W. Morokoff, and A. Owen. Valuation of mortgage backed securities using Brownian bridges to reduce effective dimension. *Journal of Computational Finance*, 1(1):27–46, 1997.

[5] D. Crisan and A. Doucet. A survey of convergence results on particle filtering methods for practitioners. *IEEE Transactions on Signal Processing*, 50(3):736–746, 2002.

[6] F. Daum, R. Co, and J. Huang. Curse of Dimensionality and Particle Filters. In *Proceedings of the IEEE Aerospace Conference*, 2003.

[7] P. Del Moral. Measure-valued processes and interacting particle systems. Application to nonlinear filtering problems. *Annals of Applied Probability*, 8(2):438–495, 1998.

[8] A. Doucet, S. Godsill, and C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10(3):197–208, 2000.

[9] W. Feller. *An Introduction to Probability Theory and its Applications, vol 2*. Wiley India Pvt. Ltd., 2008.

[10] N.J. Gordon, D.J. Salmond, and A.F.M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEE Proceedings-F*, volume 140, pages 107–113, 1993.

[11] 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):425–437, 2002.

[12] S.J. Julier and J.K. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–422, 2004.

[13] A. Kong. A Note on Importance Sampling using Standardized Weights. Technical Report 348, University of Chicago, Dept. of Statistics, 1992.

[14] F. Le Gland and N. Oudjane. Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters. *Annals of Applied Probability*, 14(1):144–187, 2004.

[15] N. Oudjane and C. Musso. Progressive correction for regularized particle filters. In *Proceedings of the 3rd International Conference on Information Fusion*, volume 2, 2000.

[16] C. P. Robert and G. Casella. *Monte Carlo statistical methods*. Springer-Verlag, 2004.

[17] C. Snyder, T. Bengtsson, P. Bickel, and J. Anderson. Obstacles to high-dimensional particle filtering. *Monthly Weather Review*, 136(12):4629–4640, 2008.