14th International Conference on Information Fusion
Chicago, Illinois, USA, July 5-8, 2011

Introducing the Laplace approximation in particle filtering

Christian Musso
Onera – The French Aerospace Lab
F-91761 Palaiseau, France
christian.musso@onera.fr

Paul Bui Quang
Onera – The French Aerospace Lab
F-91761 Palaiseau, France
paul.bui_quang@onera.fr

François Le Gland
INRIA
Rennes, France
francois.le_gland@inria.fr

***Abstract*—The situations where particle filtering fails (so-called weight degeneracy) can be detected with the asymptotic variance of the particle approximation. However, this asymptotic variance is in general intractable, and in the case of weight degeneracy, computing it by Monte Carlo sampling is inefficient. We propose to compute the asymptotic variance of the particle approximation via the Laplace method for multidimensional integrals. We present this method, and illustrate how it can be used to improve particle filtering robustness.**

Keywords: Tracking, particle filtering, asymptotic variance, Laplace approximation, regularization.

I. Introduction

Given a dynamic system, Bayesian filtering consists in online estimation of the distribution of the hidden state conditionally to observations, with the help of a prior distribution on the hidden state. From the obtained posterior distribution, one can extract many point estimates, such as posterior moments.

When the state dynamic or the observation process are nonlinear, sequential Monte Carlo methods, or particle filtering methods, are an efficient way to approximate the posterior distribution. The strength of the particle approximation is that it converges to the true posterior distribution as the number of Monte Carlo samples increases.

The principle of particle filtering is as follows. A population of weighted particles is propagated over time. Each particle undergoes at each time step a mutation step and a selection step, so that the empirical distribution of the particles population approximates online the posterior distribution.

The main weakness of particle filtering is known as weight degeneracy. When the number of time iterations increases, a few particles tends to get all the population weight, and most of the particles have a very low weight. Most of the particle do not contribute to the density approximation, thus impoverishing filtering.

The typical situation in which Monte Carlo sampling is difficult and weight degeneracy occurs is when the discrepancy between the sampling distribution (usually the prior distribution) and the weighting function (the likelihood function) is too important. Indeed, when the state space area with a high a priori probability differs too much from the area with a high likelihood, a small number of the particles sampled from the prior distribution has a significantly positive weight. This can happen in particular when the state space dimension is high [1] and lead to the divergence of the particle filter.

The purpose of this article is to propose an accurate evaluation of the variance of the particle approximation as the number of particles grows to infinity. Typically, in particle filtering, we estimate the hidden state by the posterior expectation and the asymptotic variance gives the accuracy of this estimate. Unfortunately, in a nonlinear context, its theoretical expression is in general intractable as it involves a ratio of multidimensional integrals. To compute this ratio one can use Monte Carlo sampling. However, when the prior distribution does not overlap the likelihood function (i.e., they are discrepant), this method fails. We propose to evaluate the asymptotic variance by using the Laplace method for integrals approximation. Indeed, the Laplace method is very accurate in terms of relative error, even in case of non-overlap between the supports of the prior and the likelihood. For example, it allows to approximate the variance of the particle weights and to control weights degeneracy in a more robust manner than usual empirical criteria.

In section 2, we present the basis of Bayesian filtering and its resolution by particle approximation. The Laplace approximation is introduced in details in section 3. The usefulness of this method in a Bayesian framework is demonstrated. To illustrate the accuracy of the Laplace method, we consider in section 4 a simple study case where we aim at localizing a target by triangulation. The Laplace method is shown to perform well, while importance sampling fails. In section 5, we consider a realistic bearings-only target tracking problem with weakly noised observations. The asymptotic variance of the weights is approximated online by the Laplace method. This criterion is used to decide at each time step if the particles cloud must be resampled/regularized. With a low additional computational burden, the resulting algorithm performs better in terms of robustness than the one using a classical criterion for resampling/regularization.

II. Bayesian filtering and particle approximation

A. Bayesian filtering in hidden Markov models

We consider the hidden Markov model $\{X_n, Y_n\}_{n\geq 0}$, composed of the state process $\{X_n\}_{n\geq 0}$ and of the observation process $\{Y_n\}_{n\geq 0}$. The unobserved state process is a Markov

chain, defined by

$$\begin{cases} X_0 \sim q_0(x_0)dx_0 \\ X_n|X_{n-1} = x_{n-1} \sim q_n(x_n|x_{n-1})dx_n, \ n \geq 1, \end{cases} \quad (1)$$

and taking values in an open subset of $\mathbb{R}^d$. At each time step n, the hidden state X_n delivers an observation Y_n,

$$Y_n|X_n = x_n \sim g_n(y_n|x_n)dy_n, \ n \geq 0, \quad (2)$$

where $g_n(y_n|x_n)$ is the likelihood function. We assume that for each n, the conditional distribution of Y_n given the past state path $X_{0:n}$ equals the conditional distribution of Y_n given the current state X_n, and that the observations $Y_{0:n}$ are mutually independent given $X_{0:n}$.

The objective of Bayesian filtering is to get at each time step the posterior density $p_n(x_n|y_{0:n})$. The sequence of posterior densities $\{p_n(x_n|y_{0:n})\}_{n\geq 0}$ is called the Bayesian filter. The Bayesian filter evolves in time according to a recursive dynamic. At each time step, the posterior density is propagated through a prediction step and an update step. The predictive density is given by the action of the Markovian kernel q_n on the previous posterior density p_{n-1},

$$p_{n|n-1}(x_n|y_{0:n-1}) = \int q_n(x_n|x_{n-1})p_{n-1}(x_{n-1}|y_{0:n-1})dx_{n-1}. \quad (3)$$

Then, the predictive density, which plays the role of prior distribution, is updated with the likelihood according to Bayes formula,

$$p_n(x_n|y_{0:n}) = \frac{g_n(y_n|x_n)p_{n|n-1}(x_n|y_{0:n-1})}{\int g_n(y_n|x_n)p_{n|n-1}(x_n|y_{0:n-1})dx_n}. \quad (4)$$

These step are summarized as follows:

$$p_{n-1}(x_{n-1}|y_{0:n-1}) \overset{\text{prediction}}{\longrightarrow} p_{n|n-1}(x_n|y_{0:n-1}) \quad (5)$$

$$\overset{\text{correction}}{\longrightarrow} p_n(x_n|y_{0:n}). \quad (6)$$

The purpose of particle filtering is to approximate at each time step the posterior density by the empirical distribution based on the weighted particles,

$$p_n(x_n|y_{0:n}) \approx \sum_{i=1}^N w_n^i \delta_{\xi_n^i}(x_n) \quad (7)$$

where $(\xi_n^1, \ldots, \xi_n^N)$ denote the particles, $(w_n^1, \ldots, w_n^N)$ their weights, and $\delta_{\xi_n^i}$ denotes the Dirac measure centered at ξ_n^i.

B. Particle filtering

We describe here recursively the principle of particle filtering. Let $(\xi_{n-1}^1, \ldots, \xi_{n-1}^N)$ be the particles population at time $n-1$, with weights $(w_{n-1}^1, \ldots, w_{n-1}^N)$. The particle approximation of the posterior density at time $n-1$ is

$$p_{n-1}^N(x_{n-1}|y_{0:n-1}) = \sum_{i=1}^N w_{n-1}^i \delta_{\xi_{n-1}^i}(x_{n-1}). \quad (8)$$

The predictive density is approximated by propagating the particles according to q_n,

$$p_{n|n-1}^N(x_n|y_{0:n-1}) = \sum_{i=1}^N w_{n-1}^i q_n(x_n|\xi_{n-1}^i). \quad (9)$$

This is the mutation step. Then, a new population of particles $(\xi_n^1, \ldots, \xi_n^N)$ is sampled from $p_{n|n-1}^N$ and weighted according to their likelihood. This is the selection step. The empirical distribution of this new population is

$$p_n^N(x_n|y_{0:n-1}) = \frac{\sum_{i=1}^N w_{n-1}^i g_n(y_n|\xi_n^i)\delta_{\xi_n^i}(x_n)}{\sum_{i=1}^N w_{n-1}^i g_n(y_n|\xi_n^i)} \quad (10)$$

$$= \sum_{i=1}^N w_n^i \delta_{\xi_n^i}(x_n), \quad (11)$$

where

$$w_n^i = \frac{w_{n-1}^i g_n(y_n|\xi_n^i)}{\sum_{j=1}^N w_{n-1}^j g_n(y_n|\xi_n^j)}. \quad (12)$$

Thus, p_n^N approximates the posterior density p_n. The corresponding algorithm, called sequential importance sampling (SIS), is presented in Appendix A.

A common problem in particle filtering is weight degeneracy: if nothing is done, the variance of the weights increases over time. As a result, only very few particles have a significantly positive weight after a few time steps and the empirical density p_n^N becomes very poor. A solution to this problem is to resample the particles population once in a time, thus discarding particles with a low weight and replicating those with a large weight. This additional resampling procedure has first been proposed in [2] where it is done at each time step. See Appendix B for the algorithm.

There exists many other particle filter versions, in which the particles are sample and weighted in various ways. See [3], [4] or [5] for reviews.

C. Some asymptotic results

Let us consider the SIS algorithm. For any test function ϕ, the following central limit theorem (CLT) holds [6]:

$$\sqrt{N}\left(\int \phi(x)p_n^N(x)dx - \int \phi(x)p_n(x)dx\right) \longrightarrow \mathcal{N}(0, v_n(\phi)) \quad (13)$$

in distribution when $N \to \infty$, where

$$v_n(\phi) = \frac{\int (\phi(x_n) - \bar{\phi}_n)^2 g_{0:n}(y_{0:n}|x_{0:n})^2 q_{0:n}(x_{0:n})dx_{0:n}}{\left(\int g_{0:n}(y_{0:n}|x_{0:n})q_{0:n}(x_{0:n})dx_{0:n}\right)^2}, \quad (14)$$

with

$$g_{0:n}(y_{0:n}|x_{0:n}) = \prod_{k=0}^n g_k(y_k|x_k), \quad (15)$$

$$q_{0:n}(x_{0:n}) = q_0(x_0)\prod_{k=1}^n q_k(x_k|x_{k-1}), \quad (16)$$

and

$$\bar{\phi}_n = \int \phi(x) p_n(x) dx = E[\phi(X)|Y_{1:n}]. \tag{17}$$

Another CLT concerning the variance of the weights is the following. Let $w_n^{*i} = w_{n-1}^{*i} g_n(\xi_n^i)$, for $i = 1, \ldots, N$ denote the unormalized weights. It verifies

$$\sqrt{N}\left(\frac{\frac{1}{N}\sum_{i=1}^N w_n^{*i}}{\int g_{0:n}(y_{0:n}|x_{0:n}) q_{0:n}(x_{0:n}) dx_{0:n}} - 1\right) \longrightarrow \mathcal{N}(0, V_n) \tag{18}$$

in distribution when $N \to \infty$, where

$$V_n = \frac{\int g_{0:n}(y_{0:n}|x_{0:n})^2 q_{0:n}(x_{0:n}) dx_{0:n}}{\left(\int g_{0:n}(y_{0:n}|x_{0:n}) q_{0:n}(x_{0:n}) dx_{0:n}\right)^2} - 1, \tag{19}$$

Thoughout the article, we will consider the asymptotic variances $v_n(\phi)$ and V_n as measures of the difficulty of Bayesian estimation with Monte Carlo methods. The Laplace method, presented in the next section, will allow us to get a good approximation of these quantities. In section 4 and 5, it will be used to detect weight degeneracy.

III. The Laplace approximation in Bayesian estimation

The Laplace approximation is a method to calculate integrals of the form

$$I_n = \int b(x) e^{nL(x)} dx \tag{20}$$

where b and L are smooth functions defined on an open subset of $\mathbb{R}^d$. Suppose L has a maximum and attains it at $\hat{x}$. Then I_n can be approximated by

$$\hat{I}_n = (2\pi)^{d/2} b(\hat{x}) |\det[-n\nabla^2 L(\hat{x})]|^{-1/2} e^{nL(\hat{x})} \tag{21}$$

where $\nabla^2 L$ is the $d \times d$ Hessian matrix of L. Under some regularity conditions, the Laplace approximation satisfies [7]

$$I_n = \hat{I}_n\{1 + O(n^{-1})\}. \tag{22}$$

In the field of Bayesian estimation, this technique allows to calculate posterior moments of the form

$$E[\phi(X)|Y_{1:n}] = \frac{\int \phi(x) g_{1:n}(y_{1:n}|x) q(x) dx}{\int g_{1:n}(y_{1:n}|x) q(x) dx}, \tag{23}$$

where $q(x)$ is the prior density on the state of interest X, and $g_{1:n}(y_{1:n}|x)$ is the likelihood associated with the data set $Y_1, \ldots, Y_n$.

When ϕ is a positive function, the posterior moment (23) is well approximated by the ratio of the Laplace approximations of the numerator and the denominator. For the numerator, let be $b^{\text{num}}(x) \equiv 1$ and $L^{\text{num}}(x) = L_n^{\text{num}}(x) = \frac{1}{n}[\log \phi(x) + \log g_{1:n}(y_{1:n}|x) + \log q(x)]$. For the denominator, let be $b^{\text{den}}(x) \equiv 1$ and $L^{\text{den}}(x) = L_n^{\text{den}}(x) = \frac{1}{n}[\log g_{1:n}(y_{1:n}|x) + \log q(x)]$. Therefore,

$$E[\phi(X)|Y_{1:n}] = \frac{\int b^{\text{num}}(x) e^{nL_n^{\text{num}}(x)} dx}{\int b^{\text{den}}(x) e^{nL_n^{\text{den}}(x)} dx}, \tag{24}$$

which can be approximated by

$$E[\phi(X)|Y_{1:n}] \approx \left(\frac{\det[-n\nabla^2 L_n^{\text{den}}(\hat{x})]}{\det[-n\nabla^2 L_n^{\text{num}}(\tilde{x})]}\right)^{1/2} \times \exp\{n(L_n^{\text{num}}(\tilde{x}) - L_n^{\text{den}}(\hat{x}))\}, \tag{25}$$

where $\tilde{x} = \arg\max L_n^{\text{num}}(x)$ and $\hat{x} = \arg\max L_n^{\text{den}}(x)$, i.e. $\hat{x}$ is the maximum a posteriori (MAP) estimate. Approximation (25) corresponds to the ratio of the so-called fully exponential Laplace approximations of the numerator and the denominator. It is closely related to the saddlepoint approximation, which is also useful in non-Bayesian statistics [8], [9]. The main regularity condition under which (25) holds is that the posterior density is unimodal, see [7], [10], and [11] for mathematical details. When ϕ is not a positive function (like the identity function, if we seek the posterior expectation), a simple way to calculate the fully exponential Laplace approximation of the numerator is the following: add a large constant c to ϕ, calculate the Laplace approximation of $E[\phi(X)+c|Y_{1:n}]$, then substract c [12].

The Laplace approximation can be used to calculate other quantities, like posterior covariances, or posterior marginal densities [7], [11]. Note that, when it is used to calculate posterior moments, the Laplace approximation depends essentially on the determinant of the observed Fisher information matrix $\nabla^2 \log g_{1:n}(y_{1:n}|\hat{x}) + \nabla^2 \log q(\hat{x})$.

IV. Study case: evaluation of the Laplace approximation

In this section, we consider a nonlinear Bayesian estimation study case with two goals. We illustrate the accuracy of the Laplace approximation in terms of relative error. We also demonstrate that it can be used to quantify the discrepancy between the prior information and the information brought by the observations, which measures the difficulty of Monte Carlo sampling for Bayesian estimation.

A. The model

Let us consider a fixed target in the real plane, that we are interested in locating. The hidden state $X = (P^x, P^y)$ is the cartesian position of the target. It a priori follows a Gaussian distribution with mean m and covariance matrix $Q = \sigma_Q^2 I_2$ (I_d denotes the $d \times d$ identity matrix), whose density is denoted by $q(x)$. Two sensors, with respective positions (s_1^x, s_1^y) and (s_2^x, s_2^y), measure independently n azimuth angles. The observation model is

$$Y_k = h(X) + \varepsilon_k, \tag{26}$$

for $k = 1, \ldots, n$, where

$$h(X) = \left[\arctan\frac{P^y - s_1^y}{P^x - s_1^x}, \arctan\frac{P^y - s_2^y}{P^x - s_2^x}\right]^T \tag{27}$$

is the vector made by the two azimuth measurements. The observation noises ε_k are independent with distribution

$\mathcal{N}(0_2, R)$, where $R = \sigma_R^2 I_2$. The likelihood function associated with an observation Y_k writes

$$g_k(y_k|x) = \frac{1}{(2\pi)\det(R)} \times \exp\left[-\frac{1}{2}(y_k - h(x))^T R^{-1}(y_k - h(x))\right]. \quad (28)$$

Since the observations are independent, the joint likelihood function is $g_{1:n}(y_{1:n}|x) = \prod_{k=1}^n g_k(y_k|x)$. The true target position (i.e., the state from which the observations are generated) is randomly generated from the prior distribution.

B. Posterior expectation estimation

Let p_n be the density of the posterior distribution. In this framework, the particle approximation of p_n is computed by importance sampling (IS):

$$p_n^N(x) = \sum_{i=1}^N w_n^i \delta_{\xi^i}(x) \quad (29)$$

where the ξ^i's are sampled according to $q(x)dx$ and where the importance weights are

$$w_n^i = \frac{g_{1:n}(y_{1:n}|\xi^i)}{\sum_{j=1}^N g_{1:n}(y_{1:n}|\xi^j)}. \quad (30)$$

Let $\mu_n = E(X|Y_{1:n})$ be the sought state estimate, which is the expectation a posteriori (EAP).

To illustrate the accuracy of the Laplace approximation (25), let us compare the results of the EAP calculation by the Laplace method and by the above importance sampling method. The parameters are set to $m = (20, 30)^T$, $\sigma_Q = 3$, $(s_1^x, s_1^y) = (0, 0)$, $(s_2^x, s_2^y) = (1, 0)$, and $\sigma_R = 10\pi/180$. The number of observations per sensor is $n = 2$. As μ_n^N is strongly consistent approximation of μ_n, we consider a very large number of particles (say, $N = 10^6$) and use the resulting approximation, denoted by μ_n^*, as the true EAP. We compute the IS approximation $\mu_n^N = \sum_{i=1}^N w_n^i \xi^i$ with $N = 50$ particles, and the Laplace approximation (25). The average relative error (over 100 independent trials) is less than 0.1% on each component for the two approximations. Thus, the Laplace method is here as accurate as the particles mean, for a data size as small as $n = 2$, although it is an asymptotic approximation with respect to n.

C. Asymptotic covariance estimation

We are now interested in calculating the covariance matrix of μ_n^N. The CLT holds [6] that

$$\sqrt{N}(\mu_n^N - \mu_n) \longrightarrow \mathcal{N}(0, \Sigma_n) \quad (31)$$

in distribution when $N \to \infty$. The components of the asymptotic covariance matrix Σ_n are

$$\Sigma_{n,ij} = \frac{\int (x_i - \mu_{n,i})(x_j - \mu_{n,j}) g_{1:n}(y_{1:n}|x)^2 q(x)dx}{\left(\int g_{1:n}(y_{1:n}|x)q(x)dx\right)^2} \quad (32)$$

N	50	100	200	400
$\|\Sigma_n^N - \Sigma_n^*\|_F$	0.1413	0.0606	0.0466	0.0193
$\|\Sigma_n^N ./ \Sigma_n^* - J_2\|_F$	0.0116	0.0048	0.0037	0.0027
$\|\Sigma_n^{\text{Lap}} - \Sigma_n^*\|_F$	0.0893	0.0893	0.0893	0.0893
$\|\Sigma_n^{\text{Lap}} ./ \Sigma_n^* - J_2\|_F$	0.0068	0.0068	0.0068	0.0068

Table I
$n = 4$, $\sigma_Q = 3$, $\sigma_R = 10\pi/180$. $\|\Sigma_n^*\|_F = 19.4326$.

for $i, j \in \{1, 2\}$, where $\mu_n = (\mu_{n,1}, \mu_{n,2})^T$. A first way to approximate Σ_n is by importance sampling, that is

$$\Sigma_{n,ij}^N = \frac{\frac{1}{N}\sum_{i=1}^N (\xi_1^i - \mu_{n,1}^N)(\xi_2^i - \mu_{n,2}^N) g_{1:n}(y_{1:n}|\xi^i)^2}{\left(\frac{1}{N}\sum_{i=1}^N g_{1:n}(y_{1:n}|\xi^i)\right)^2} \quad (33)$$

for $i, j \in \{1, 2\}$. An alternative way to approximate Σ_n is to use the Laplace method on the numerator and the denominator of (32). The corresponding approximation is denoted Σ_n^{Lap}. Σ_n^N is a consistent approximation of Σ_n, so that the approximation (33) with $N = 10^6$ particles is used as a reference, which is denoted Σ_n^*.

Let us compare the IS approximation and the Laplace approximation. We consider $n = 4$ obervations and keep the same model parameters as in the section IV-B. Table I gives the Frobenius norm[1] of the absolute and relative errors[2] between the approximation Σ_n^N and Σ_n^* averaged over 10^4 independent runs, for a growing number of particles. As expected, the error decreases with N. Besides, we can check that the Laplace approximation Σ_n^{Lap} is accurate; it is here as good as the importance sampling approximation for a reasonable number of particles.

Moreover, we argue that Σ_n is a good indicator of the limits of the IS method. To illustrate this assertion, let us change the model parameters in order to decrease the overlap between the sampling density q and the weighting function g_n. A way to do it is to increase the standard deviation of the prior distribution and to decrease the standard deviation of the observation noise. In this case it is easy to chek that the asymptotic covariance matrix (32) grows up. Roughly speaking, fewer particles fall in the state space area with a high likelihood, and more particles have a very low weight. The particle approximation p_n^N of the posterior density will then be poorer, and we expect that this phenomenon can be captured by the Laplace approximation Σ_n^{Lap}. Let us set $\sigma_Q = 7$, $\sigma_R = \pi/180$, and $n = 4$. Results are summarized in Table II. First, we observe that $\|\Sigma_n^*\|_F$ has increased here compared to the first case (Table I), which confirms that Σ_n quantifies the difficulty of importance sampling. Moreover, Σ_n^{Lap} is a good approximation of Σ_n unlike Σ_n^N. Indeed, Σ_n^{Lap} does not suffer from the introduced discrepancy as it is a good approximation in terms of relative error, see (22). Note that the Monte Carlo approximation of Σ_n (32) can be very poor when

[1]The Frobenius norm of a matrix A is defined by $\sqrt{\text{trace}(A^T A)}$.

[2]We define the relative error between two matrix A and B with the same size $d \times d$ as $\|A./B - J_d\|_F$, where $./$ is the term-by-term ratio and J_d is the matrix with 1 on each component.

N	50	100	200	400
$\|\Sigma_n^N - \Sigma_n^*\|_F$	208.0893	181.7208	98.0890	41.1918
$\|\Sigma_n^N ./ \Sigma_n^* - J_2\|_F$	0.4844	0.4194	0.2273	0.0950
$\|\Sigma_n^{\text{Lap}} - \Sigma_n^*\|_F$	19.8261	19.8261	19.8261	19.8261
$\|\Sigma_n^{\text{Lap}} ./ \Sigma_n^* - J_2\|_F$	0.0583	0.0583	0.0583	0.0583

Table II
$n = 4$, $\sigma_Q = 7$, $\sigma_R = \pi/180$. $\|\Sigma_n^*\|_F = 865.3311$.

the denominator is close to zero. Indeed, Monte Carlo gives an approximation in terms of absolute error. Consequently, Σ_n^{Lap} is an approximation of Σ_n that can be trusted in situations where importance sampling fails.

V. Bearings-only tracking

In this section, we consider a more realistic bearings-only target tracking problem, for which we use the regularized particle filtering (RPF) algorithm. At each time step, to decide whether to regularize or not, we consider two methods, one of them being based on the Laplace approximation.

A. The model

The target follows a linear uniform motion in the real plane. Its state at time n is defined as $X_n = (P_n^x, \dot{P}_n^x, P_n^y, \dot{P}_n^y)^T$, where P_n^x and P_n^y are the x- and y-coordinates, and $\dot{P}_n^x$ and $\dot{P}_n^y$ are the corresponding velocities. The initial state is random but the state transition is deterministic. More precisely, the state process is

$$\begin{cases} X_0 \sim q_0(x_0)dx_0, \\ X_n = FX_{n-1}, \ n \geq 1 \end{cases} \tag{34}$$

where

$$F = \begin{pmatrix} 1 & \Delta & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{35}$$

with Δ the sampling period. We assume that the distribution of the initial state is Gaussian with mean m_0 and covariance matrix Q_0.

At each time step, a sensor delivers a noised measurement of the azimuth. The observation process is

$$Y_n = h_n(X_n) + \varepsilon_n, \ n \geq 0 \tag{36}$$

where

$$h_n(X_n) = \arctan \frac{P_n^y - s_n^y}{P_n^x - s_n^x}, \tag{37}$$

with (s_n^x, s_n^y) the sensor position at time n. Note that the observer must maneuver to insure the observability of the target. The observation noises ε_n are i.i.d. with distribution $\mathcal{N}(0, \sigma_R^2)$. The likelihood function (the density of Y_n given $X_n = x_n$) is therefore

$$g_n(x_n|y_n) = \frac{1}{\sqrt{2\pi}\sigma_R} \exp\left[-\frac{1}{2\sigma_R^2}(y_n - h_n(x_n))^2\right]. \tag{38}$$

As the target motion is deterministic here, the Markov transition density kernel writes

$$q_n(x_n|x_{n-1}) = \delta(x_n = Fx_{n-1}), \tag{39}$$

so that the predictive density is

$$p_{n|n-1}(x_n) = p_{n-1}(F^{-1}x_n|y_{0:n-1}), \tag{40}$$

because $\det(F) = 1$, and the posterior density is

$$p_n(x_n|y_{0:n}) = g_{0:n}(y_{0:n}|x_n)q_0(F^{-n}x_n), \tag{41}$$

where

$$\begin{aligned} g_{0:n}(y_{0:n}|x_n) &= \prod_{k=0}^{n} g_k(y_k|F^{k-n}x_n) \\ &= \frac{1}{(2\pi)^{n/2}\sigma_R^n} \exp\left[-\frac{1}{2\sigma_R^2}\sum_{k=0}^{n}(y_k - h_k(F^{k-n}x_n))^2\right]. \end{aligned} \tag{42}$$

B. Algorithm

We use the post-regularized particle filtering algorithm to estimate online the target state in the above model. Details on this technique can be found in [13]. The asymptotic variance of the unormalized weights (cf. equation (18)) writes

$$V_n = \frac{\int g_{0:n}(y_{0:n}|x_n)^2 q_0(F^{-n}x_n)dx_n}{\left(\int g_{0:n}(y_{0:n}|x_n)q_0(F^{-n}x_n)dx_n\right)^2} - 1 \tag{43}$$

if the time step n is before the first regularization time; else, if regularization has occured before time n, it is expressed as

$$V_n = \frac{\int g_{k_0:n}(y_{k_0:n}|x_n)^2 q_{k_0}(F^{k_0-n}x_n)dx_n}{\left(\int g_{k_0:n}(y_{k_0:n}|x_n)q_{k_0}(F^{k_0-n}x_n)dx_n\right)^2} - 1 \tag{44}$$

where $g_{k_0:n}(y_{k_0:n}|x_n) = \prod_{k=k_0}^{n} g_k(y_k|F^{k-n}x_n)$, k_0 is the last regularization time before n, and q_{k_0} is the density of the particles after resampling at time k_0. The Laplace approximation of (44) is

$$\begin{aligned} &V_n^{\text{Lap}} \\ &= (2\pi)^{-d/2} \frac{g_{k_0:n}(y_{k_0:n}|\tilde{x}_n)^2 q_{k_0}(F^{k_0-n}\tilde{x}_n)}{g_{k_0:n}(y_{k_0:n}|\hat{x}_n)^2 q_{k_0}(F^{k_0-n}\hat{x}_n)^2} \\ &\quad \times \frac{\det[-\nabla^2 \log g_{k_0:n}(y_{k_0:n}|\hat{x}_n) - \nabla^2 \log q_{k_0}(F^{k_0-n}\hat{x}_n)]}{\det[-2\nabla^2 \log g_{k_0:n}(y_{k_0:n}|\tilde{x}_n) - \nabla^2 \log q_{k_0}(F^{k_0-n}\tilde{x}_n)]^{1/2}} \\ &\quad - 1, \end{aligned} \tag{45}$$

where

$$\hat{x}_n = \arg\max\{g_{k_0:n}(y_{k_0:n}|x)q_{k_0}(F^{k_0-n}x)\} \tag{46}$$

is the MAP and where

$$\tilde{x}_n = \arg\max\{g_{k_0:n}(y_{k_0:n}|x)^2 q_{k_0}(F^{k_0-n}x)\}. \tag{47}$$

Note that, with this approximation, we can see that the error of the particle approximation is high when the observed Fisher information matrix, which is defined as $-\nabla^2 \log g_{k_0:n}(y_{k_0:n}|\hat{x}_n) - \nabla^2 \log q_{k_0}(F^{k_0-n}\hat{x}_n)$, is large (for example, in the case of low measurement noise). That is, the sharper the posterior density mode, the more particles are needed in order to maintain a given accuracy.

The Laplace approximation (45) is computed in the proposed algorithm, which we call the Laplace-RPF. It is taken as a criterion to decide whether resample/regularize: it is done when this approximated variance exceeds a threshold T. However, equation (45) involves unknown quantities, like $\log q_{k_0}$ and its Hessian, and the maximum of the integrands on the numerator and the denominator. We compute it in the following way:

$$\begin{aligned} & V_n^{\text{Lap}} \\ & \approx (2\pi)^{-d/2} \frac{g_{k_0:n}(y_{k_0:n}|\tilde{\xi}_n^N)^2 \bar{q}_{k_0}(F^{k_0-n}\tilde{\xi}_n^N)}{g_{k_0:n}(y_{k_0:n}|\hat{\xi}_n^N)^2 \bar{q}_{k_0}(F^{k_0-n}\hat{\xi}_n^N)^2} \\ & \times \frac{\det[-\nabla^2 \log g_{k_0:n}(y_{k_0:n}|\hat{\xi}_n^N) - \hat{Q}_{k_0}]}{\det[-2\nabla^2 \log g_{k_0:n}(y_{k_0:n}|\tilde{\xi}_n^N) - \tilde{Q}_{k_0}]^{1/2}} - 1, \end{aligned} \quad (48)$$

where

$$\hat{\xi}_n^N = \frac{\sum_{i=1}^N w_n^i \xi_n^i}{\sum_{i=1}^N w_n^i}, \quad (49)$$

$$\tilde{\xi}_n^N = \frac{\sum_{i=1}^N (w_n^i)^2 \xi_n^i}{\sum_{i=1}^N (w_n^i)^2}, \quad (50)$$

and

$$\hat{Q}_{k_0} = \tilde{Q}_{k_0} = \frac{1}{N} \sum_{i=1}^N (\xi_n^i - \hat{\xi}_n^N)(\xi_n^i - \hat{\xi}_n^N)^T \quad (51)$$

is the empirical covariance matrix of the particles cloud after regularization at time k_0 ($\hat{Q}_{k_0} = Q$ if $k_0 = 0$). Besides, the density of the sampling distribution q_{k_0} is approximated by the Gaussian density $\bar{q}_{k_0}$ with mean $\frac{1}{N}\sum_{i=1} \xi_{k_0}^i$ and covariance $\hat{Q}_{k_0}$ ($\bar{q}_{k_0} = q_0$ if $k_0 = 0$). Computation (48) is justified by the fact that the Laplace approximation is robust to the substitution of the modes (46) and (47) by the empirical means (49) and (50). The algorithm is presented below.

Laplace-RPF

- $k = 0$
  Set k_0 to 0.
  [Mutation] For $i = 1, \ldots, N$, sample independently $\xi_0^i \sim q_0(x_0)dx_0$.
  [Weighting] For $i = 1, \ldots, N$, compute the weights $w_0^i = \frac{g_0(\xi_0^i)}{\sum_{j=1}^N g_0(\xi_0^j)}$.
- $k = 1, \ldots, n$
  Compute V_n^{Lap} according to (48) with the observations $Y_{k_0:k}$.
  - If $V_n^{\text{Lap}} > T$, set k_0 to k.
    [Mutation] For $i = 1, \ldots, N$, $\xi_{k_0}^i = F\xi_{k_0-1}^i$.
    [Weighting] For $i = 1, \ldots, N$, compute the weights $w_{k_0}^i = \frac{w_{k_0-1}^i g_{k_0}(\xi_{k_0}^i)}{\sum_{j=1}^N w_{k_0-1}^j g_{k_0}(\xi_{k_0}^j)}$.
    [Regularization] Compute the empirical covariance matrix S of $\{\xi_{k_0}^i\}_{i=1,\ldots,N}$. Compute A, the Cholesky decomposition of S ($AA^T = S$). Resample $\{\xi_{k_0}^i\}_{i=1,\ldots,N}$ according to the weights $\{w_{k_0}^i\}_{i=1,\ldots,N}$ to obtain a new set of particles $\{\tilde{\xi}_{k_0}^i\}_{i=1,\ldots,N}$. Draw N independent samples η^i from the Epanechnikov kernel K. For $i = 1, \ldots, N$, set $\xi_{k_0}^i = \tilde{\xi}_{k_0}^i + h_{\text{opt}} A \eta^i$ where h_{opt} is the optimal bandwidth. See [13] for details.
  - Else
    [Mutation] For $i = 1, \ldots, N$, $\xi_k^i = F\xi_{k-1}^i$.
    [Weighting] For $i = 1, \ldots, N$, compute the weights $w_k^i = \frac{w_{k-1}^i g_k(y_k|\xi_k^i)}{\sum_{j=1}^N w_{k-1}^j g_k(y_k|\xi_k^j)}$.

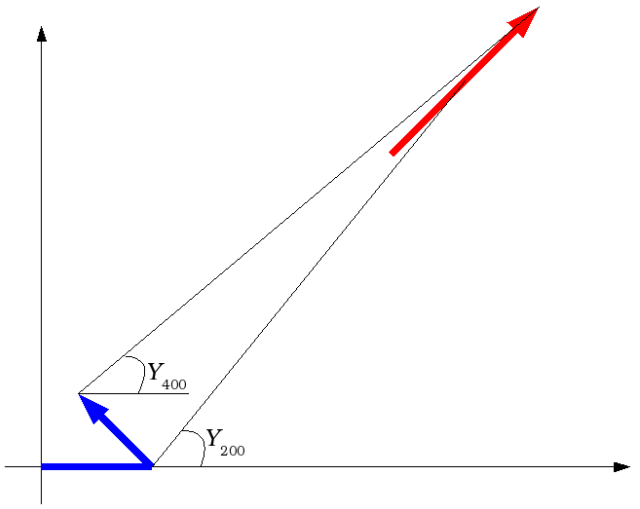

Figure 1. Bearings-only tracking scenario with measurements at time $n = 200$ and $n = 400$. Blue line: observer trajectory. Red line: target trajectory.

C. Results

The number of time steps is $n_{\max} = 400$, and the sampling period in $\Delta = 1$ second. The norm of the velocity of the observer is 20 m/s and the maneuver occurs at time $n_{\max}/2$, so that the trajectory of the observer is composed of two equal legs (see Figure 1). The mean of the initial state distribution is $m_0 = (10\text{km}, 10\text{m/s}, 10\text{km}, 10\text{m/s})^T$, and its covariance matrix is $Q_0 = \text{diag}((5\text{km})^2, (10\text{m/s})^2, (5\text{km})^2, (10\text{m/s})^2)$. The standard deviation of the observation noise is $\sigma_R = 0.3$ degrees. The true initial position is randomly generated from $\mathcal{N}(m_0, Q_0)$.

Let us benchmark two algorithms for the bearings-only tracking problem. The difference between the two is how the decision is taken at each time step about resampling/regularization or not. In the first one (the RPF), a classical empirical entropy-based criterion is used [14]. The entropy of the particles population at time n is defined as

$$\log N + \sum_{i=1}^n w_n^i \log w_n^i. \quad (52)$$

Resampling/regularization is done when the entropy is above 0.1 (this threshold has been optimized empirically). The second is the Laplace-RPF, where the threshold has been determined empirically to $T = 5 \cdot 10^{-4}$.

We have performed 100 independent runs of the scenario. Each run consists in generating new measurements and new trajectories of the particles. In order to keep approximatively the same computing time for the two algorithms, we take 7000 particles for the RPF and 5000 particles for the Laplace-RPF. The Laplace-RPF produced 1 divergence while the RPF produced 6 divergences. By divergence, we mean that error of the final state estimate is 4 times greater than the standard deviation given by the posterior Cramér-Rao bound (PCRB) [15]. In Figures 2 and 3, the empirical root mean squared error (RMSE) of the estimated x-position is plotted (mean over 100 independent runs) . These RMSE are computed for the two following cases: one case including the divergent runs and one case excluding the divergent runs. We can see the benefit of the Laplace-RPF in terms of robustness. The bearings-only tracking problem we treat here is difficult as the standard deviation of the measurement noise is small. Note that the Laplace approximation can be introduced in any particle filter, like the progressive correction [16] technique, which is more adapted when the measurement noise is weak.

In this simulation, the asymptotic variance of the weights is computed online to monitor weight degeneracy. We have considered a small measurement noise, so that the sampling density does not overlap the likelihood. In that case, the empirical entropy criterion (or any criterion based on the empirical variance of the weigths, like the effective sample size [17]) is inaccurate, and can lead to inappropriate resampling decisions. We have observed that the Laplace-RPF resamples more frequently the particle population than the standard RPF. We have increased the resampling rate of the RPF by reducing the threshold on the entropy. In that case, the number of divergences of the RPF decreases but the RMSE becomes significantly higher. Note that before the maneuver, the RMSE of the Laplace-RPF is high. This is due to the fact that, during the transition period where the target is not observable, the Laplace-RPF resamples at each time step. After the maneuver, the Laplace-RPF tends to resample less frequently as time grows. Finally, we can conclude that the improvement of the Laplace-RPF is due to an appropriate choice of resampling times.

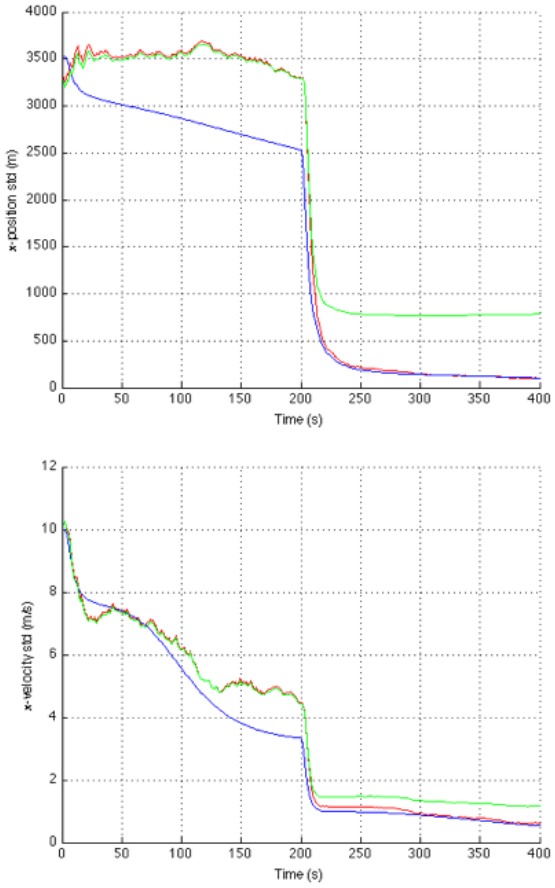

Figure 2. Comparison of the RPF error (x-position and x-velocity) and the PCRB. Green line: RPF standard deviation including divergent runs. Red line: RPF standard deviation excluding divergent runs. Blue line: PCRB. $N = 7000$ particles.

VI. Conclusion

In Bayesian statistics, the Laplace approximation is, under some regularity conditions, an accurate method for approximating integrals that involve the posterior density. In this article, we introduce the Laplace approximation in particle filtering. This method allows an accurate and fast computation of the asymptotic variance of the particle approximation and the asymptotic variance of the weights. When these variances are high, which is typically the case when the prior density does not overlap the likelihood function, one may fear weight degeneracy. In that case, the Laplace approximation can be used to compute them, while empirical methods give poor results.

The simulations show that the Laplace method improves the robustness of particle filtering. Another potential application of this technique is the accurate computation of the confidence region of the approximated posterior expectation.

Appendix A

Sequential importance sampling (SIS)

- $k = 0$
  [Mutation] For $i = 1, \ldots, N$, sample independently $\xi_0^i \sim q_0(x_0)dx_0$.
  [Weighting] For $i = 1, \ldots, N$, compute the weights $w_0^i = \frac{g_0(\xi_0^i)}{\sum_{j=1}^N g_0(\xi_0^j)}$.
- $k = 1, \ldots, n$
  [Mutation] For $i = 1, \ldots, N$, sample $\xi_k^i \sim q_k(x_k|\xi_{k-1}^i)dx_k$.
  [Weighting] For $i = 1, \ldots, N$, compute the weights $w_k^i = \frac{w_{k-1}^i g_k(y_k|\xi_k^i)}{\sum_{j=1}^N w_{k-1}^j g_k(y_k|\xi_k^j)}$.

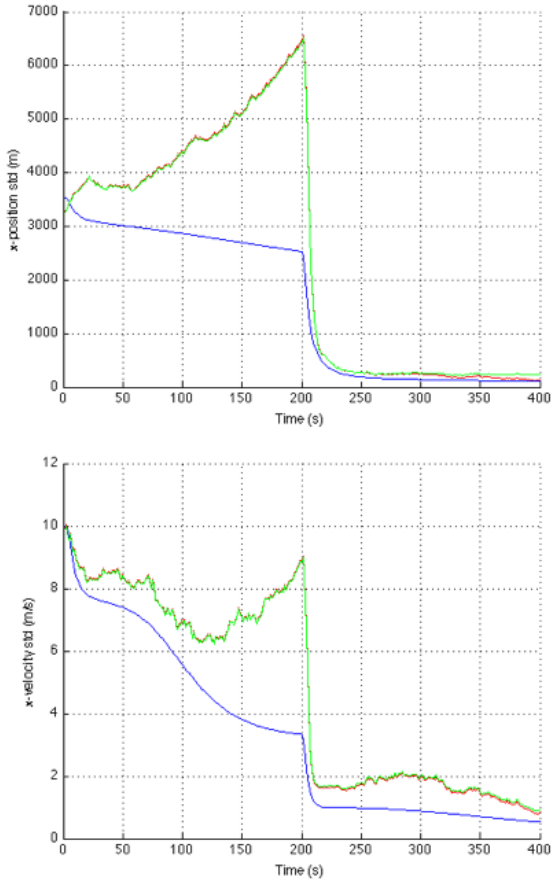

Figure 3. Comparison of the Laplace-RPF error (x-position and x-velocity) with the PCRB. Green line: Laplace-RPF standard deviation including divergent runs. Red line: Laplace-RPF standard deviation excluding divergent runs. Blue line : PCRB. $N = 5000$ particles.

APPENDIX B

Sequential importance sampling with resampling (SIR)

- $k = 0$
  [Mutation] For $i = 1, \ldots, N$, sample independently $\xi_0^i \sim q_0(x_0)dx_0$.
  [Weighting] For $i = 1, \ldots, N$, compute the weights $w_0^i = \frac{g_0(\xi_0^i)}{\sum_{j=1}^N g_0(\xi_0^j)}$.
- $k = 1, \ldots, n$
  [Mutation] For $i = 1, \ldots, N$, select a particle $\hat{\xi}_{k-1}^i$ among the population $\xi_{k-1}^1, \ldots, \xi_{k-1}^N$ according to the weights $w_{k-1}^1, \ldots, w_{k-1}^N$ and sample $\xi_k^i \sim q_k(x_k|\hat{\xi}_{k-1}^i)dx_k$.
  [Weighting] For $i = 1, \ldots, N$, compute the weights $w_k^i = \frac{g_k(\xi_k^i)}{\sum_{j=1}^N g_k(\xi_k^j)}$.

REFERENCES

[1] P. Bui Quang, C. Musso, and F. Le Gland, "An insight into the issue of dimensionality in particle filtering," in *Proceedings of the 13th International Conference on Information Fusion*, 2010.

[2] N. Gordon, D. Salmond, and A. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," in *IEE Proceedings-F*, vol. 140, no. 2, 1993, pp. 107–113.

[3] M. 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*, vol. 50, no. 2, pp. 174–188, 2002.

[4] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, 2000.

[5] O. Cappé, S. Godsill, and E. Moulines, "An Overview of Existing Methods and Recent Advances in Sequential Monte Carlo," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 899–924, 2007.

[6] P. Del Moral, *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer, 2004.

[7] R. E. Kass, L. Tierney, and J. B. Kadane, "Asymptotics in Bayesian Computation," in *Bayesian Statistics 3*, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, Eds. Oxford University Press, 1988, pp. 261–278.

[8] N. Reid, "Saddlepoint methods and statistical inference," *Statistical Science*, vol. 3, no. 2, pp. 213–227, 1988.

[9] H. E. Daniels, "Saddlepoint approximations in statistics," *The Annals of Mathematical Statistics*, vol. 25, no. 4, pp. 631–650, 1954.

[10] R. E. Kass, L. Tierney, and J. B. Kadane, "The validity of posterior expansions based on Laplace's method," in *Bayesian and Likelihood Methods in Statistics and Econometrics*, S. Geisser, J. S. Hodges, S. J. Press, and A. Zellner, Eds. Elsevier Science, 1990, pp. 473–488.

[11] L. Tierney and J. B. Kadane, "Accurate Approximations for Posterior Moments and Marginal Densities," *Journal of the Americain Statistical Association*, vol. 81, no. 393, pp. 82–86, 1986.

[12] L. Tierney, R. E. Kass, and J. B. Kadane, "Fully Exponential Laplace Approximations to Expectations and Variances of Nonpositive Functions," *Journal of the American Statistical Association*, vol. 84, no. 407, pp. 710–716, 1989.

[13] C. Musso, N. Oudjane, and F. Le Gland, "Improving regularised particle filters," in *Sequential Monte Carlo methods in practice*, A. Doucet, N. de Freitas, and N. Gordon, Eds. Springer, 2001.

[14] D. T. Pham, "Stochastic Methods for Sequential Data Assimilation in Strongly Nonlinear Systems," *Monthly Weather Review*, vol. 129, pp. 1194–1207, 2001.

[15] P. Tichavskỳ, C. H. Muravchik, and A. Nehorai, "Posterior Cramer-Rao bounds for discrete-time nonlinear filtering," *IEEE Transactions on Signal Processing*, vol. 46, no. 5, pp. 1386–1396, 1998.

[16] N. Oudjane and C. Musso, "Progressive correction for regularized particle filters," in *Proceedings of the 3rd International Conference on Information Fusion*, 2000.

[17] A. Kong, "A Note on Importance Sampling using Standardized Weights," University of Chicago, Dept. of Statistics, Tech. Rep. 348, 1992.

ACKNOWLEDGEMENT

The authors are grateful to Mathilde Chorewoman for having initiated interesting discussions around this article.