Data reduction for Particle Filters

Christian Musso
ONERA/DTIM
29 av. de la division Leclerc,
92322 cedex Châtillon, France
christian.musso@onera.fr

Nadia Oudjane
EDF R&D/OSIRIS
1 av. du Général De Gaulle,
92140 cedex, Clamart, France
nadia.oudjane@edf.fr

Abstract

In this paper, we are interested in nonlinear filtering approximations. Approximate filters (such as the Extended Kalman Filter or Particle Filters) are known to converge to the optimal filter when the local error (committed at each step of time) vanishes. But this convergence is in general not uniform in time. Error bounds obtained in the general case suggest that the approximation error could grow exponentially with time. This divergent phenomena is actually observed in some simulations. To avoid that divergence of approximate filters with the number of observations, an idea is to reduce the number of observations without losing too much information. This paper proposes an optimal approach to reduce the number of observations for filtering. This new approach is applied to Particle Filtering and tested in the case of the bearing only tracking problem.

1 Introduction

Nonlinear filtering consists in computing recursively the conditionnal distribution of a state process given some noisy and partial observations of that process accumulated along the time. Theoritical analysis of the error committed by approximate filters, show that the error can grow exponentially with the number of observations (see [2], [6] or [7]). In the specific case of the Particle Filter, Monte Carlo errors are committed at each observation step. Those local errors are transported along the optimal filter evolution via a multiplicative coefficient which can grow exponentially with the observations number. Practically, this divergent phenomena is actually observed in some simulations. To avoid that exponential growth of the error with the observations number, an idea is to reduce the number of observations without losing too much information.
In this paper, we propose an optimal approach to reduce the number of observations based on the Gauss-Legendre quadrature. The original set of observations is replaced by a smaller set of observations, each new observation being more precise than the original ones. This approach has been originally developed in [8] and [3] to reduce the size of maximum likelihood problems. This paper extends the application of this approach to the nonlinear filtering problem, and more specifically to Particle Filtering.
In the next section, we state the basic results needed for the application of the quadrature approach for the approximation of discrete sums. In the third section, we recall the nonlinear filtering framework and our motivations. The fourth section consists in applying the quadrature approach to reduce the observations number. Section 5 presents an application of this approach to Particle Filtering in the case of the bearing only tracking problem.

2 Gauss-Legendre quadrature for discrete sums

The Gauss-Legendre quadrature is a numerical method used to approximate the integral of a function f by a weighted sum of values taken by f. Here we are interested in approximating a discrete sum of a large number of f values by a weighted sum of a small number of f values. In this section, we will state the general setting and the main results needed to apply the Gauss-Legendre quadrature to the case of discrete sums approximation.

2.1 General description

Let $t_1 < \cdots < t_n$ be n positive reals. Let us denote by μ_n the discrete measure such that, for all $s \in \mathbb{R}$,

$$\mu_n(ds) = \sum_{t=t_1}^{t_n} \delta_t(ds) \ ,$$

where δ_t denotes the Dirac measure on t, taking the value one if $ds = \{t\}$ and zero otherwise.
Let us denote by $L^2(\mu_n)$ the set of real valued functions f defined on $\mathbb{R}$ such that

$$\sum_{t=t_1}^{t_n} f^2(t) < \infty \ .$$

$L^2(\mu_n)$ is endowed with the inner product $\langle \cdot, \cdot \rangle$ defined by

$$\langle f, g \rangle = \sum_{t=t_1}^{t_n} f(t)\, g(t) \ , \qquad (1)$$

for any $f, g \in L^2(\mu_n)$.

For any $1 \leq m \leq n$, let us denote by Ψ_m, the Legendre Polynomial of degree m such that

$$\langle \Psi_m, t^j \rangle = 0 \ , \tag{2}$$

for any integer $0 \leq j \leq m-1$. Ψ_m can be computed recursively using the Graham-Schmidt formula (8). Let us denote by $T_1 \leq \cdots \leq T_m$ the roots of Ψ_m. One can prove that for any $i \neq j$ with $1 \leq i, j \leq m$, T_i is a real such that

$$t_1 \leq T_i \leq t_n \quad \text{and} \quad T_i \neq T_j \ .$$

If $m = n$, then $T_i = t_i$, for any $1 \leq i \leq m = n$.
The roots of Ψ_m can be computed approximately by the Laguerre algorithm (see (9)).

For any $1 \leq i \leq m$, let us introduce Φ_i the Lagrange polynomial of degree $m-1$ such that

$$\Phi_i(t) = \begin{cases} 1 & \text{if} \quad t = T_i \\ \\ 0 & \text{if} \quad t = T_j \neq T_i \ . \end{cases} \tag{3}$$

Just as in the usual continuous case (see [10]), one can prove that the following quadrature formula holds.

Proposition 2.1. *Let $t_1 < \cdots < t_n$ be n positive reals and $1 \leq m \leq n$ be an integer. Let f be a real valued function defined on $\mathbb{R}$, having its derivatives continuous up to the order m, then the following approximation identity holds*

$$\sum_{t=t_1}^{t_n} f(t) = \sum_{i=1}^{m} \|\Phi_i\|_2^2 f(T_i) \ + \ E_m(f) \ , \tag{4}$$

with

$$E_m(f) = \frac{1}{(2m)!} \|\Psi_m\|_2^2 f^{(2m)}(\xi) \ , \tag{5}$$

where $t_1 < \xi < t_n$.

If f is a polynomial with a degree $p \leq 2m-1$, then $E_m(f) = 0$ and any sum of n values of f is equal to a weighted sum of $m \leq n$ values of f.

Φ_i's norm can be approximated by the following lemma.

Lemma 2.2. *For all $1 \leq i \leq m$, let Φ_i be the Lagrange polynomial of degree $m-1$ defined by (3). Then*

$$\sum_{i=1}^{m} \|\Phi_i\|_2^2 = n \ , \quad \textit{and} \quad \|\Phi_i\|_2^2 = n \int_0^1 \tilde{\Phi}_i(t) dt + o(\frac{1}{n}) \ , \tag{6}$$

where, for $i = 1, \cdots, m$, $\tilde{\Phi}_i$ are the Lagrange Polynomials associated to the integral inner product on $[0,1]$.

This lemma says approximately that $\|\Phi_i\|_2^2$ is of order n/m.

PROOF OF LEMMA 2.2. Since Φ_i is a polynomial of degree $m-1 \leq 2m-1$, $E_m(\Phi_i) = 0$, hence applying Proposition 2.1 (the Gauss-Legendre quadrature) with Φ_i yields to the following equality,

$$\sum_{t=t_1}^{t_n} \Phi_i(t) = \sum_{j=1}^{m} \|\Phi_i\|_2^2 \Phi_i(T_j) = \|\Phi_i\|_2^2 \ . \tag{7}$$

Let Q denote the polynomial defined, for all $t \in \mathbb{R}$, by

$$Q(t) = -1 + \sum_{i=1}^{m} \Phi_i(t) \ .$$

Q has degree $m-1$ and is equal to zero on m points $T_1, \cdots, T_m$. Hence $Q \equiv 0$. This result, together with (7) yields

$$\sum_{i=1}^{m} \|\Phi_i\|_2^2 = \sum_{i=1}^{m} \sum_{t=t_1}^{t_n} \Phi_i(t) = \sum_{t=t_1}^{t_n} \sum_{i=1}^{m} \Phi_i(t) = n \ .$$

The second part of the proof is obtained by contraction of the interval $[t_0, t_n]$ into $[0,1]$.

□

2.2 Computations of Legendre Polynomials roots

The Graham-Schmidt orthogonalization process applied to the basis of polynomials of degree m, $(1, t, \cdots, t^m)$, leads to the following recurrence equation, to compute Legendre polynomials,

$$\Psi_{k+1}(t) = (t - \beta_k)\Psi_k(t) - \gamma_k \Psi_{k-1}(t) \ , \quad \text{with} \tag{8}$$

$$\beta_k = \frac{\langle t\Psi_k, \Psi_k \rangle}{\|\Psi_k\|_2^2} \quad \text{and} \quad \gamma_k = \frac{\langle t\Psi_k, \Psi_{k-1} \rangle}{\|\Psi_k\|_2^2} \ ,$$

where the initial condition is given by

$$\Psi_1(t) = t - \frac{1}{n} \sum_{t=t_1}^{t_n} t \quad \text{and} \quad \Psi_0(t) = 1 \ .$$

Recall that all roots of Ψ_m are simple. One of the most efficient method to compute Ψ_m's roots is the following Laguerre algorithm.

Algorithm 2.3. *(Laguerre algorithm)*
Set $k = m$ and $P_m = \Psi_m$.

(i) Repeat the following recurrence until convergence to τ_∞:

$$\tau_{p+1} = \tau_p - \frac{kP_k(\tau_p)}{P_k'(\tau_p) \pm \sqrt{H_k(\tau_p)}} \ , \quad \textit{where} \tag{9}$$

$$H_k(t) = (k-1)\left[(k-1)(P_k'(t))^2 - kP_k(t)P_k''(t)\right] \ .$$

The sign before the square root is chosen such that $|\tau_{p+1} - \tau_p|$ is the smallest.

(ii) If $k \geq 1$, set $T_{(k)} = \tau_\infty$, $P_{k-1}(t) = P_k(t)/(t - T_{(k)})$ and $k = k-1$, and go back to (i).

(iii) If $k = 0$, stop.

The roots $T_1 < \cdots < T_m$ are obtained by ordering $\{T_{(1)}, \cdots, T_{(k)}\}$.

In the case of the Legendre polynomial with degree $m = 5$ and when $t_k = k$ for $1 \leq k \leq n$, the roots can be computed exactly with the following formula :

$$T_3 = \frac{n+1}{2}\,, \quad T_1 = T_3 - a\,, \quad T_2 = T_3 - b\,, \qquad (10)$$

$$T_4 = T_3 + b\,, \quad \text{and} \quad T_5 = T_3 + a\,, \qquad (11)$$

with a and b being the positive real such that

$$a^2 = \frac{1}{36}\left(5(n^2-7) + 2\left[\frac{10n^4 - 95n^2 + 1228}{7}\right]^{1/2}\right)$$

$$b^2 = \frac{1}{36}\left(5(n^2-7) - 2\left[\frac{10n^4 - 95n^2 + 1228}{7}\right]^{1/2}\right)\,.$$

3 The filtering problem: motivations

3.1 The filtering framework

The filtering problem consists in computing recursively the conditional law of a signal (called the state process) conditionally to noisy and partial observations accumulated along the time.
In this article, we focus on the following specific filtering problem. The state process $(X_t)_{t\geq 0}$ is a deterministic continuous time process in $\mathbb{R}^d$ and the observation process $(Y_{t_k})_{1\leq k\leq n}$ is a random sequence in $\mathbb{R}$ verifying the following system,

$$\begin{cases} X_{t_0} \sim \pi_0 \\ \dot{X}_t = f(X_t) \qquad \text{for} \quad t \geq 0 \\ Y_{t_k} = h(X_{t_k}) + \sigma\varepsilon_k \quad \text{for} \quad 1 \leq k \leq n\,, \end{cases} \qquad (12)$$

where the following properties are verified:

(i) π_0 is a given probability distribution on $\mathbb{R}^d$ (modelling the error on X_{t_0});

(ii) $(\varepsilon_1, \cdots, \varepsilon_n)$ are i.i.d. standard and centered Gaussian random variables.

(iii) h is a real valued function defined on $\mathbb{R}^d$;

(iv) For all $k \in \{1, \cdots, d\}$, let us denote by g_k the real valued function defined on $\mathbb{R}^+$ such that

$$g_k(t) = \frac{\partial h}{\partial x_k}(X(t))\,, \quad \text{for all} \quad t \in \mathbb{R}^+\,, \qquad (13)$$

assume that for all $k \in \{1, \cdots, d\}$, g_k has its derivatives continuous up to the order $2m$;

(v) Let us denote by g the real valued function defined on $\mathbb{R}^+$ such that

$$g(t) = h(X(t))\,, \quad \text{for all} \quad t \in \mathbb{R}^+\,, \qquad (14)$$

(iv) implies that g has its derivatives continuous up to the order $2m$.

Remark 3.1. *Notice that there is no dynamical noise in the filtering system (20). However, the main results of this paper presented in Proposition 4.1 are still valid in the case of a small dynamical noise. In the general case, the first assertion of Proposition 4.1 is still valid, but the second one should be analyzed more specifically.*

For $1 \leq k \leq n$, the conditional distribution of X_{t_k} given the past observations up to the current time t_k, $(Y_{t_1}, \cdots, Y_{t_k})$ is denoted by π_k. For all $x \in \mathbb{R}^d$,

$$\pi_k(dx) = \mathbb{P}(X_{t_k} \in dx \,|\, Y_{t_1}, \cdots, Y_{t_k})\,.$$

The sequence of probability measures $(\pi_k)_{0\leq k\leq n}$ is called the optimal filter. The evolution of the optimal filter, between two steps of time t_k and t_{k+1}, consists into two steps:

(i) the **prediction** step according to the dynamics of the state produces the prediction filter $\pi_{k+1|k}$;

(ii) the **correction** step taking the measurements into account produces the new optimal filter $\pi_{k+1|k+1}$.

3.2 Particle filtering and exponential growth of the error with time

When f and h are linear functions of the state, it is well-known that the Kalman Filter (KF) computes rapidly and exactly the optimal filter. Out of this specific case, some approximations are required. The Extended Kalman Filter (EKF) consists in applying the KF algorithm in linearizing locally f and h around $\hat{X}_{t_k}$, the current estimate of π_k's expectation. When f or h are highly nonlinear, particle filtering ([4]) can be a good way to approximate the optimal filter. The idea is to produce, at each time step, a sample of independent random variables, called a particle system, (approximately) distributed according to the optimal filter.

The evolution of the particle filter, π_k^N (where N denotes the size of the particle system) between two steps of time t_k and t_{k+1}, consists into two steps:

(i) **Sampled prediction**: we generate N i.i.d. random variables according to the weighted discrete probability distribution $\pi_k^N = \sum_{i=1}^N \omega_k^i\, \delta_{X^i_{k|k-1}}$. The more likely particles are selected, so that the particle system concentrates in regions of interest of the state space. Then each selected particle evolves independently according to the dynamics. This produces a new particle system $(X^1_{k+1|k}, \cdots, X^N_{k+1|k})$.

(ii) **Correction**: each particle is weighted according to its likelihood w.r.t. to the current observation $Y_{t_{k+1}}$, which produces the weights $(\omega^1_{k+1}, \cdots \omega^N_{k+1})$ and the new estimate $\pi^N_{k+1} = \sum_{i=1}^N \omega^i_{k+1}\, \delta_{X^i_{k+1|k}}$.

In the case where the dynamical noise is small or nonexistent (which is the case of system (20)), the variety of the

particle system decreases at each time step because of the accumulation of repetitions in the sample. To insure the diversity of the particle system as time progresses, a "regularization step" has been added to this algorithm. The resulting filters, are called Kernel Filters [5] or Regularized Particle Filters (RPF) [9]. The first stage of the sampled prediction is changed. Instead of simply resampling the particle system, an artificial noise is added.

In both case, the IPF and the RPF are proved to converge to the optimal filter, in the weak sense, with rate $1/\sqrt{N}$, but the error is not uniformly bounded in time (except under some strong mixing assumptions on the state process, ([2], [6], [7]) which explains why some divergent behaviors are observed (see [1] for a complete analysis of the asymptotic behavior). The best bounds obtained in a general setting for the error committed by a large class of approximate filters, π'_n, (such as Particle Filters) suggest that this error could grow exponentially with time. More precisely,

$$\sup_{\|\phi\|=1} \mathbb{E}[\,|\langle \pi_n - \pi'_n, \phi\rangle| \mid Y_{t_1}, \cdots, Y_{t_n}\,] \leq \sum_{k=1}^{n} C^{n-k+1}\delta\ , \tag{15}$$

where δ is a bound for the local error and C is a positive real. Note that in the case of the Particle Filter, the Monte Carlo method implies that $\delta = \dfrac{B}{\sqrt{N}}$. Hence, each local error committed at each time of observation t_k propagates along the optimal filter evolution from t_k to t_n with a multiplicative coefficient which grows exponentially with the number of observations $n-k+1$.

In this paper, we propose to reduce the number of observations to avoid the exponential growth of the error with time. This approach, developed in the following section, can be used for any kind of approximate filters. In Section 5, devoted to simulations, we focus on Particle Filters.

4 Optimal data reduction

Reducing the number of observations could be a good approach to avoid the divergence of approximate filters with the number of observations but, in the same time, it could be a disastrous approach in terms of loss of information. An optimal way to reduce the number of observations without losing too much information is to use the Gauss-Legendre quadrature introduced in Section 2. This approach has been originally developed in [8] and [3] to reduce the size of maximum likelihood problems. Here we extend the application of this approach to the nonlinear filtering problem.
Let $1 \leq m \leq n$ be the new number of observations. The new set of observations denoted by $(\tilde{Y}_1, \cdots, \tilde{Y}_m)$ will be given as a linear combination of the original set of observations $(Y_{t_1}, \cdots, Y_{t_n})$,

$$\tilde{Y}_i = \frac{1}{\|\Phi_i\|^2} \sum_{t=t_1}^{t_n} \Phi_i(t)\, Y_t\ , \tag{16}$$

where Φ_i is the Lagrange polynomial of degree $m-1$ defined in (3).

The following proposition will help us to change the initial filtering problem (20) for a quasi-equivalent filtering problem with a number $1 \leq m \leq n$ of observations.

Proposition 4.1. *The following assertions holds.*

(i) The new set of observations $(\tilde{Y}_1, \cdots, \tilde{Y}_m)$ defined by (16) is related to the state process by the following observation equation, for all $1 \leq i \leq m$,

$$\tilde{Y}_i = h(X_{T_i}) + \frac{\sigma}{\|\Phi_i\|_2}\varepsilon_i\ , \tag{17}$$

where $(\varepsilon_1, \cdots, \varepsilon_m)$ are independent standard Gaussian random variables, ε_i having its mean equal to $E_m(\Phi_i g) \approx 0$.

(ii) The Fisher matrix associated to observations $(Y_{t_1}, \cdots, Y_{t_n})$ denoted by I is approximately equal to the Fisher matrix associated to the new set of observations $(\tilde{Y}_1, \cdots, \tilde{Y}_m)$ denoted by $\tilde{I}$ i.e.

$$I = \tilde{I} + [E_m(g_k g_l)]_{1\leq k,l\leq d}\ , \tag{18}$$

where $[E_m(g_k g_l)]_{1\leq k,l\leq d}$ denotes the $d\times d$ matrix with elements $E_m(g_k g_l) \approx 0$.

Hence the new set of observations $(\tilde{Y}_1, \cdots, \tilde{Y}_m)$ is quasi-exhaustive (no loss of information). The initial filtering problem (20) can be replaced without losing too much information by the following filtering problem where the number of obervations has been reduced $m \leq n$,

$$\begin{cases} X_{t_0} & \sim & \pi_0 & & \\ \dot{X}_t & = & f(X_t) & \text{for} & t \geq 0 \\ \tilde{Y}_i & = & h(X_{T_i}) + \tilde{\sigma}_i\varepsilon_i & \text{with} & \tilde{\sigma}_i = \dfrac{\sigma}{\|\Phi_i\|_2}\ , \end{cases} \tag{19}$$

where the observation noises $(\varepsilon_1, \cdots, \varepsilon_m)$ are approximately i.i.d. standard and centered Gaussian random variables.

Remark 4.2. *Note that Lemma 2.2 shows that the new standard deviation $\tilde{\sigma}_i$ is in general smaller that the original one σ. More precisely, $\sum_{i=1}^{m} \sigma^2/\tilde{\sigma}_i^2 = n$. Hence the new observation set $\tilde{Y}$ is smaller, but in the same time each new observation is more precise than original observations.*
The new system (19) is a very precise approximation of the original system (20) when the signal has a polynomial behavior (m is a tuning parameter depending on the regularity of the signal).

Remark 4.3. *If this data reduction method is used globally on the whole set of observations, then the recursive property of the filter is lost since one has to wait to have all the observations available to compute the estimate. In practice, this approach is used periodically every n observations. The modified filter is updated less frequently than the original filter since one has to wait a given number of observations before updating the estimate, but the modified filter remains in a sense "piecewise recursive".*

PROOF OF PROPOSITION 4.1. In view of equation (16), the vector $\tilde{Y} = (\tilde{Y}_1, \cdots, \tilde{Y}_m)$ is a linear function of the original set of observations $Y = (Y_{t_1}, \cdots, Y_{t_n})$. Since Y is a Gaussian vector (recall that X is deterministic) then $\tilde{Y}$ is also a Gaussian vector. Hence proving equation (17) is equivalent to compute the two first moments of $\tilde{Y}$.

Taking the expectation in (16) yields

$$\mathbb{E}[\tilde{Y}_i] = \frac{1}{\|\Phi_i\|^2} \sum_{t=t_1}^{t_n} \Phi_i(t)\, h(X_t) \ .$$

Since $g = h(X(\cdot))$ has its derivatives continuous up to the order $2m$, then we can apply Proposition 2.1 for the product function $(\Phi_i\, g)$,

$$\begin{aligned} \mathbb{E}[\tilde{Y}_i] &= \frac{1}{\|\Phi_i\|^2} \sum_{j=1}^{m} \|\Phi_j\|^2\, \Phi_i(T_j)\, h(X_{T_j}) + E_m(\Phi_i\, g) \\ &= h(X_{T_i}) + E_m(\Phi_i g) \ . \end{aligned}$$

Since $\mathrm{cov}(Y_t, Y_{t'}) = 0$ for $t \neq t'$ then

$$\begin{aligned} \mathrm{cov}(\tilde{Y}_i, \tilde{Y}_j) &= \frac{1}{\|\Phi_i\|^2\, \|\Phi_j\|^2} \sum_{t,t'=t_1}^{t_n} \Phi_i(t)\, \Phi_j(t')\, \mathrm{cov}(Y_t, Y_{t'}) \\ &= \frac{\sigma^2}{\|\Phi_i\|^2\, \|\Phi_j\|^2} \sum_{t=t_1}^{t_n} \Phi_i(t)\, \Phi_j(t) \end{aligned}$$

Since $(\Phi_i\, \Phi_j)$ is a polynomial of degree $2m-2 \leq 2m-1$, then for $i \neq j$, using Proposition 2.1 yields

$$\sum_{t=t_1}^{t_n} \Phi_i(t)\, \Phi_j(t) = \sum_{k=1}^{m} \|\Phi_k\|^2\, \Phi_j(T_k)\, \Phi_i(T_k) = 0 \ ,$$

which finally gives $\mathrm{cov}(\tilde{Y}_i, \tilde{Y}_j) = \sigma^2/\|\Phi_i\|^2$ if $i = j$ and 0 otherwise.

Let us consider now the proof of equation (18). Note that, for all $1 \leq k \leq d$, g_k has all its derivatives continuous up to the order $2m$, hence applying Proposition 2.1 to the $\mathbb{R}^d$ valued function $\frac{\partial h}{\partial X}(X(\cdot)) = (g_1, \cdots, g_d)'$ yields

$$\begin{aligned} I &= \frac{1}{\sigma^2} \sum_{t=t_1}^{t_n} \left(\frac{\partial h}{\partial X_t}\right)_{t=t} \left(\frac{\partial h}{\partial X_t}\right)'_{t=t} \\ &= \frac{1}{\sigma^2} \sum_{i=1}^{m} \|\Phi_i\|^2 \left(\frac{\partial h}{\partial X_t}\right)_{t=T_i} \left(\frac{\partial h}{\partial X_t}\right)'_{t=T_i} \\ &\quad + [E_m(g_k g_l)]_{1\leq k,l\leq d} \ . \end{aligned}$$

On the other hand, using definition (16), one computes

$$\tilde{I} = \sum_{i=1}^{m} \frac{\|\Phi_i\|^2}{\sigma^2} \left(\frac{\partial h}{\partial X_t}\right)_{t=T_i} \left(\frac{\partial h}{\partial X_t}\right)'_{t=T_i} \ ,$$

which ends the proof □

5 Simulations

5.1 Bearing only: problem formulation

Let us consider the following bearing-only tracking problem. An observer measures angles from a target moving in a plan every second from time 1 to time n. The goal is to estimate the target state $X_t = (x(t), \dot{x}(t), y(t), \dot{y}(t))'$, where $(x(t), y(t))'$ are the cartesian coordinates of the target in the plan, at time t. The motion of the target is assumed to be uniform.

The observer has to maneuver in order to estimate the state of the target. Let us denote by $(x_o(t), y_o(t))'$, the cartesian coordinates of the observer in the plan.

$$\left\{ \begin{array}{lcll} X_0 & \sim & \pi_0 & \\ \dot{X}_t & = & FX_t & \text{for} \quad t \geq 0 \\ Y_{t_k} & = & \mathrm{Arctan}\left(\frac{x(t_k)-x_o(t_k)}{y(t_k)-y_o(t_k)}\right) + \sigma\varepsilon_k & \text{for} \quad 1 \leq k \leq n \ , \end{array} \right. \tag{20}$$

where $\varepsilon_1, \cdots, \varepsilon_n$ are i.i.d. standard Gaussian random variables.

The trajectory of the observer is composed by two equal legs (the maneuver is done at time $n/2$). The data reduction method is applied for four sets of observations from time 1 to $n/4$, $n/4$ to $n/2$, $n/2$ to $3n/4$ and $3n/4$ to n. For each piece of length $n/4$ we compute five quadrature points ($m = 5$) according to Proposition 4.1. Hence the modified filter will be "picewise recursive" every $n/4$ seconds (see figure 1).

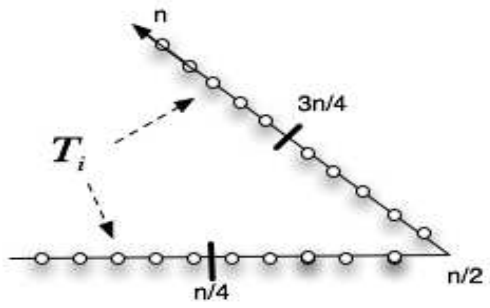

Figure 1. Data reduction configuration

5.2 Results

The parameters of the simulation are the following:

- The initial observer position $(x_o(1), y_o(1)) = (0, 0)$, the observer velocity norm is equal to $20m/s$, the legs angle is equal to 45^o;
- The initial target position $(x(1), y(1)) = (10km, 10km)$, the target velocity $(\dot{x}(t), \dot{y}(t)) = (10m/s, 10m/s)$;

- The number of observations $n = 800$;
- The initial law of the state X_0 is Gaussian with covariance matrix Σ=diag$[(5km, 10m/s, 5km, 10m/s)^2)]$ and with mean generated randomly according to $\mathcal{N}(X_0, \Sigma)$;
- The standard deviation of the observation noise $\sigma = 1^o$.

We have applied the RPF and the RPF using the data reduction method (Data Reduction Particle Filter (DRPF)) for this tracking problem. We have run each filter on 50 Monte Carlo trials and computed the square root of the mean integrated error (SMISE) for each filter by averaging errors among all Monte Carlo trial. The number of particles used is 5000 for the RPF and 30000 for the DRPF, so that both filters perform as well (see figures 2 and 3). But the computing cost for the RPF is three times greater than the computing cost of the DRPF (the data reduction method allows to reduce the observations number from 800 to 20). The filters error are compared to the Posterior Cramer Rao lower Band (PCRB) for the x-position the x-velocity. We can see on figures 2 and 3 that both filters are efficient. Notice that with fewer particles (such that the computing time is the same for both filters) the RPF would have shown some divergences.

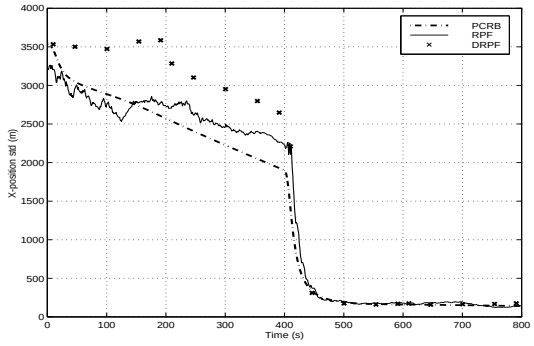

Figure 2. SMISE of the x-position estimate w.r.t. the time

6 Conclusion

We have proposed a method which allows data reduction in filtering problems and keeps approximately the amount of information unchanged. This approach improves the Particle Filter in two ways:

- the computing time is drastically reduced which allows to increase the number of particles;
- the error due to the Monte Carlo approximation is reduced and its propagation along the time is limited.

The new filter has shown good performances in simulations when applied to the bearing only tracking problem.

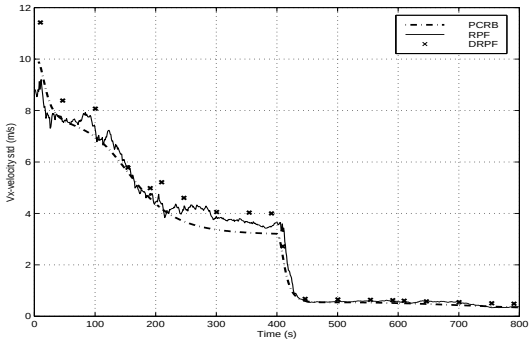

Figure 3. SMISE of the x-velocity estimate w.r.t. the time

Acknowledgements The authors are grateful to David Choreman for having initiated interesting discussions around this article.

References

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

[2] P. Del Moral and A. Guionnet. On the stability of measure valued processes. Applications to nonlinear filtering and interacting particle systems. Publication du Laboratoire de Statistique et Probabilités 3-98, Université Paul Sabatier, Toulouse, 1998.

[3] P. Din Tuan and C. Musso. Une méthode rapide et efficace pour la trajectographie passive par mesures de deux azimuts. In *7 ème colloque GRETSI sur le traitement du signal et des images, Juan-les-Pins*, September 1991.

[4] N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non–Gaussian Bayesian state estimation. *IEE Proceedings, Part F*, 140(2):107–113, Apr. 1993.

[5] M. Hürzeler and H. R. Künsch. Monte Carlo approximations for general state space models. *Journal of Computational and Graphical Statistics*, 7(2):175–193, June 1998.

[6] F. Le Gland and N. Oudjane. A robustification approach to stability and to uniform particle approximation of nonlinear filters: the example of pseudo-mixing signals. *Stochastic processes and their applications*, 106(2):279–316, 2003.

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

[8] C. Musso. *Méthodes rapides d'estimation en trajectographie par mesures d'azimuts*. PhD thesis, Université Joseph Fourier-Grenoble 1, 1993.

[9] C. Musso and N. Oudjane. Regularisation schemes for branching particle systems as a numerical solving method of the nonlinear filtering problem. In *Proceedings of the Irish Signals Systems Conference, Dublin*, June 1998.

[10] J. Stoer and R. Burlisch. *Introduction to numerical analysis*. Springer-Verlag, 1980.