An Honest Look at the Weighted Particle Filter

Michael A. Kouritzin [a,*]

[a]*Department of Mathematical and Statistical Sciences*

University of Alberta, Edmonton, AB T6G 2G1 Canada

Abstract

The classical particle filter, introduced in 1993, approximates the normalized filter directly. It has two defiencies, over resampling and the inability to distinguish models, the former of which was overcome but the later is fundamental. Conversely, the weighted particle filter, motivated by the unnormalized filter development, does not employ resampling and facilitates Bayes' factor model selection but often suffers particle spread, where the majority of particles do not track the underlying signal. Still, resampling introduces noise and there are situations where the weighted particle filter does perform well. Herein, the weighted particle filter is analyzed in a simple discrete-time setting and rate-of-convergence baseline results are established that can be compared to results for other particle filters. Moreover, an example illustrating failure of the weighted particle filter is given.

Key words: Unnormalized Filter, Particle Filter, Law of Large Numbers, Functional Central Limit Theorem, Functional Law of the Iterated Logarithms, Large Deviations Principle

2010 MSC: primary 60B05; secondary 60B10

1 Introduction

Nonlinear filtering deals with estimating the current state of a non-observable signal X based on the history of a distorted, corrupted partial observation process Y living on the same probability space $(\Omega, \mathcal{F}, P)$ as X. For many practical problems the signal is a time-homogeneous discrete-time Markov process $\{X_n,\ n = 0, 1, 2, ...\}$, living on some complete, separable metric space (E, ρ), with initial distribution π_0 and transition probability kernel K. The observation process takes the form ($Y_0 = 0$ and) $Y_n = h(X_{n-1}) + V_n$ for $n \in \mathbb{N}$, where $\{V_n\}_{n=1}^{\infty}$ are independent random vectors with common *strictly positive, bounded* density g and the sensor function h is a measurable mapping from E to $\mathbb{R}^d$. Then, the objective of filtering is to compute the conditional probabilities $\pi_n(A) = P\left(X_n \in A \middle| \mathcal{F}_n^Y\right)$, $n = 1, 2, ...$, for all Borel sets A or, equivalently, the conditional expectations $\pi_n(f) = E^P\left(f(X_n) \middle| \mathcal{F}_n^Y\right)$ for all bounded, measurable $f : E \to \mathbb{R}$, where $\mathcal{F}_n^Y \doteq \sigma\{Y_l, l = 1, ..., n\}$ is the information obtained from the back observations. Since π_n only depends upon the joint distribution of (X, Y) there is no loss of generality in taking $\Omega = (E \times \mathbb{R}^d)^\infty, \mathcal{F} = \mathcal{B}((E \times \mathbb{R}^d)^\infty)$ in the sequel. While there are well-known mathematical formulae for π_n, these formulae are, with few exceptions like the Kalman and Benes filters, fundamentally infinitely dimensional and hence not computer implementable. Still, there are many ways to approximate these conditional distributions π_n in a computer workable manner. Particle filters are one of the most popular ways to approximate these distributions.

Hammersley and Morton [9] appear to be the first to have thought of using particle filters. However, the ideas were formulated more clearly in the pioneering work of Handschin and Mayne [11] as well as Handschin [10]. Still, Gordon, Salmond and Smith [8] were probably the first to derive a general particle filter to approximate the filter distributions π_n directly.

* Corresponding author.
Email address: michaelk@ualberta.ca (Michael A. Kouritzin).

Their so-called Bootstrap algorithm was later improved in Del Moral, Kouritzin and Miclo [7]. Alternatively, one can approximate the unnormalized filter distributions introduced (in the continuous setting) by Zakai [15], which was first done (also in the continuous setting) by Crisan and Lyons [5] as well as Crisan, Gaines and Lyons [6]. However, the actual weighted method was introduced by Kurtz and Xiong [13], [14] and illustrated by Ballantyne, Chan and Kouritzin [1]. Various authors have added resampling techniques to improve performance but it is this weighted method that we explain and study below in the discrete-time setting.

Suppose hereafter that $\mathcal{F}^Z_{-1} \doteq \{\emptyset, \Omega\}$, $\mathcal{F}^Z_n \doteq \sigma\{Z^k_l, k \in \mathcal{K}, l \leq n\}$ when $n \in \mathbb{N}_0$ and $\mathcal{F}^Z_\infty \doteq \sigma\{Z^k_l, k \in \mathcal{K}, l < \infty\}$ for random variables $\{Z^k_n, k \in \mathcal{K}, n \in \{0,1,2,...\}\}$ on $(\Omega, \mathcal{F})$. (This is consistent with $\mathcal{F}^Y_n$ defined above if $\mathcal{K}$ just has one element.) One of best ways of constructing particle filters is to transfer all of the information contained in the observations to a likelihood process by way of measure change. In this reference probability method, a new fictitious probability measure Q is introduced under which the signal, observation process $\{(X_n, Y_{n+1}),\ n = 0,1,2,...\}$ has the same (process) distribution as the signal, noise process $\{(X_n, V_{n+1}),\ n = 0,1,2,...\}$ does under P. In particular, this means that the observations become i.i.d. random vectors with strictly-positive, bounded density g that are independent of X under measure Q. All the observation information is absorbed into the likelihood ratio process $\{L_n,\ n = 1,2,...\}$ transforming Q back to P, which in our case has the form

$$\frac{dP}{dQ}\Big|_{\mathcal{F}^X_\infty \vee \mathcal{F}^Y_n} = L_n = \prod_{j=1}^n W_j, \quad W_j = \alpha_j(X_{j-1}), \tag{1}$$

and the weight function has the form

$$\alpha_j(x) = \frac{g\left(Y_j - h\left(x\right)\right)}{g\left(Y_j\right)}, \tag{2}$$

so $L_n = W_n L_{n-1}$ and $L_0 = 1$. The following result constructs the real probability P from the fictitious one Q.

Theorem 1 *Suppose* $\{X_n,\ n=0,1,...\}$ *and* $\{Y_n,\ n=1,2,...\}$ *are independent processes and* $\{Y_n\}$ *are i.i.d. with strictly-positive, bounded density* g *on* $\mathbb{R}^d$ *with some probability measure* Q*, and* $V_n = Y_n - h(X_{n-1})$ *for all* $n = 1,2,...$ *Then, there exists a probability measure* P *such that (1) holds,* $\{V_n,\ n=1,2,...\}$ *are i.i.d. on* $(\Omega,\mathcal{F},P)$ *with density* g *and* $\{X_n\}$ *is independent of* $\{V_n\}$ *with the same law as on* $(\Omega,\mathcal{F},Q)$*.*

This is basically a discrete version of Girsanov's theorem. We give the proof for completeness, even though the ideas are well known.

Proof. Define P_n by Radon-Nykodym derivative

$$\frac{dP_n}{dQ} = L_n = \prod_{m=1}^{n} \frac{g\left(Y_m - h\left(X_{m-1}\right)\right)}{g\left(Y_m\right)} \tag{3}$$

and let $1 \le j_1 < j_2 < \cdots < j_k \le n$, $0 \le i_1 < i_2 < \cdots < i_l$. Then, by the independence of X and Y under Q we have for $f_r \in B(\mathbb{R}^d)$ and $\phi_p \in B(E)$

$$\begin{aligned}
& E^{P_n}\left[\prod_{r=1}^{k} f_r(V_{j_r}) \prod_{p=1}^{l} \phi_p(X_{i_p})\right] \qquad (4)\\
&= E^Q\left[\prod_{m=1}^{n} \frac{g\left(Y_m - h\left(X_{m-1}\right)\right)}{g\left(Y_m\right)} \prod_{r=1}^{k} f_r(Y_{j_r} - h(X_{j_r-1})) \prod_{p=1}^{l} \phi_p(X_{i_p})\right] \\
&= E^Q\left[\prod_{p=1}^{l} \phi_p(X_{i_p}) \int_{\mathbb{R}^d} g_1\left(y_1 - h\left(X_0\right)\right) dy_1 \cdots \int_{\mathbb{R}^d} g_n\left(y_n - h\left(X_{n-1}\right)\right) dy_n\right] \\
&= E^Q\left[\prod_{p=1}^{l} \phi_p(X_{i_p})\right] \int_{\mathbb{R}^d} g_1\left(v_1\right) dv_1 \cdots \int_{\mathbb{R}^d} g_n\left(v_n\right) dv_n \\
&= E^Q\left[\prod_{p=1}^{l} \phi_p(X_{i_p})\right] \prod_{r=1}^{k} \int_{\mathbb{R}^d} f_r(v_{j_r}) g\left(v_{j_r}\right) dv_{j_r},
\end{aligned}$$

$$\text{where } g_i = \begin{cases} g f_r & \text{if } i = j_r \\ g & \text{if } i \notin \{j_1, ..., j_k\} \end{cases}. \tag{5}$$

The $\{P_n\}$ are consistent by (4). The result follows by Kolmogorov's consistency (see e.g. Dudley Theorem 12.1.2). $\square$

We define the unnormalized filter as

$$\sigma_n(f) = E^Q\left(L_n f(X_n)\middle|\mathcal{F}_n^Y\right) \tag{6}$$

so $\sigma_0 = \pi_0$, as $L_0 = 1$ and $\mathcal{F}_0^Y = \{\emptyset, \Omega\}$. Two nice features about σ_n are that: 1) $\sigma_n(1)$ provides the Bayes factor that $\{Y_m\}_{m=1}^n$ satisfies $Y_m = h(X_{m-1}) + V_m$ over $Y_m = V_m$, with $\{V_m\}$ being i.i.d. with density g, since

$$\sigma_n(1) = \frac{E^Q\left[\prod_{m=1}^{n} g\left(Y_m - h\left(X_{m-1}\right)\right)|\mathcal{F}_n^Y\right]}{E^Q\left[\prod_{m=1}^{n} g\left(Y_m\right)|\mathcal{F}_n^Y\right]}$$

is the ratio of marginal likelihoods for these two models. 2) $\pi_n(f) = \frac{\sigma_n(f)}{\sigma_n(1)}$ by Bayes rule since

$$\begin{aligned} E^Q[\pi_n(f)\sigma_n(1)1_A] &= E^Q[E^P(f(X_n)|\mathcal{F}_n^Y)E^Q(L_n|\mathcal{F}_n^Y)1_A] \\ &= E^Q[E^Q(L_n E^P(f(X_n)1_A|\mathcal{F}_n^Y)|\mathcal{F}_n^Y)] \\ &= E^Q[L_n E^P(f(X_n)1_A|\mathcal{F}_n^Y)] \\ &= E^P[f(X_n)1_A] \\ &= E^Q[L_n f(X_n)1_A] = E^Q[\sigma_n(f)1_A] \end{aligned} \tag{7}$$

for any $A \in \mathcal{F}_n^Y$. Therefore, one can construct particle system approximations (σ_n^N in this note) to σ_n and then produce filter approximations to π_n as $\pi_n^N(f) = \frac{\sigma_n^N(f)}{\sigma_n^N(1)}$. Moreover, Bayes factor for model selection is obtained by taking the ratio $B_{12} = \frac{\sigma_n^{(1)}(1)}{\sigma_n^{(2)}(1)}$ of unnormalized filter total mass for two models. In particular, if $\sigma_n^{(1)}(f) = E^Q\left(L_n f\left(X_n^{(1)}\right)\middle|\mathcal{F}_n^Y\right)$ and $\sigma_n^{(2)}(f) = E^Q\left(L_n f\left(X_n^{(2)}\right)\middle|\mathcal{F}_n^Y\right)$ for two different signal models $X^{(1)}$ and $X^{(2)}$, then B_{12} provides the Bayes factor for $\{Y_m = h(X_{m-1}^{(1)}) + V_m\}_{m=1}^n$ over $\{Y_m = h(X_{m-1}^{(2)}) + V_m\}_{m=1}^n$. Hence, we can also do Bayesian model selection by approximating the unnormalized filter for each candidate model (see Kouritzin and Zeng [12]).

2 Notation and Unnormalized Filter

For any finite measure η and integrable function f, we define

$$\eta f = \int_E f(x)\,\eta(dx)$$
$$\eta K(dx) = \int_E K(z,dx)\,\eta(dz) \text{ and } K^n(y,dx) = \int_E K^{n-1}(z,dx)K(y,dz) \text{ for } n = 2,3,...$$
$$Kf(x) = \int_E f(z)\,K(x,dz)\,.$$

Since $Q\left(X_{n+1} \in A \middle| \mathcal{F}_n^X\right) = K(X_n, A)$, one has $E^Q\left[f(X_n)\middle|\mathcal{F}_{n-1}^X\right] = E^P\left[f(X_n)\middle|\mathcal{F}_{n-1}^X\right] = Kf(X_{n-1})$. Now, we let $B(E)$, $B(E)_+$, $\overline{C}(E)$ and $\overline{C}(E)_+$ denote the bounded measurable, non-negative bounded, continuous bounded, and non-negative continuous bounded functions respectively and define $|f|_\infty = \sup_{x\in E}|f(x)|$. It follows that $Kf \in B(E)_+$ if $f \in B(E)_+$. We also let $\mathcal{M}(E)$ ($\mathcal{P}(E)$) denote the finite (probability) measures on E with weak convergence topology, defined for $\{\mu_n\}, \mu \in \mathcal{M}(E)$ by $\mu_n \Rightarrow \mu$ if and only if $\mu_n(f) \to \mu(f)$ for all $\overline{C}(E)$. It follows from Blount and Kouritzin [3] that there is countable collection $\{f_i\}_{i=1}^\infty \subset \overline{C}(E)_+$ that is closed under multiplication and satisfies the property that $\mu_n(f_i) \to \mu(f_i)$ for all i implies that $\mu_n(f) \to \mu(f)$ for all $\overline{C}(E)$ and, thereby, that $\mu_n \Rightarrow \mu$. For the sake of completeness, one such possible collection is given by

$$\{f_i\}_{i=1}^\infty = \left\{\prod_{j=1}^{l}(1-\rho(\cdot,x_j)) \vee 0 : l \in \{0,1,2,...\}, x_j \in \{y_k\}_{k=1}^\infty\right\}, \tag{8}$$

for some dense collection $\{y_k\}_{k=1}^\infty \subset E$. Here, the product over zero functions is taken to be the constant function 1. These $\{f_i\}_{i=1}^\infty$ are actually Lipschitz continuous as well.

Now, $\sigma_0 = \pi_0$ and, using (1,6), we have the following recursion for σ_n:

$$\begin{aligned}\sigma_n(f) &= E^Q\left[L_n f(X_n)\middle|\mathcal{F}_n^Y\right] \qquad (9)\\ &= E^Q\left[W_n L_{n-1} E^Q\left[f(X_n)\middle|\mathcal{F}_n^Y \vee \mathcal{F}_{n-1}^X\right]\middle|\mathcal{F}_n^Y\right] \text{ by the tower property}\end{aligned}$$

$$
\begin{aligned}
&= E^Q\left[W_n L_{n-1} Kf\left(X_{n-1}\right)\middle|\mathcal{F}_n^Y\right] \text{ by } Q-\text{independence of } X, Y \\
&= E^Q\left[L_{n-1} A_n f\left(X_{n-1}\right)\middle|\mathcal{F}_{n-1}^Y\right] \text{ since } \{Y_n\} \text{ is iid, independent of } X \\
&= \sigma_{n-1}\left(A_n f\right) \ \forall n = 1, 2, ...,
\end{aligned}
$$

where the (random) operator A_n is defined as

$$
A_n f\left(x\right) = \begin{cases} \frac{g(Y_n - h(x))}{g(Y_n)} Kf\left(x\right) = \alpha_n(x) Kf\left(x\right) & n = 1, 2, ... \\ f(x) & n = 0 \end{cases}. \tag{10}
$$

Applying this recursion (9) repeatedly, we have that

$$
\sigma_n\left(f\right) = \pi_0\left(A_{1,n} f\right), \text{ where } A_{i,n} f\left(x\right) = \begin{cases} A_i\left(A_{i+1} \cdots \left(A_n f\right)\right)\left(x\right) \ \forall i \le n \\ f(x) & i = n+1 \end{cases}. \tag{11}
$$

This immediately implies (see (7)) that

$$
\pi_n(f) = \frac{\sigma_{n-1}\left(A_n f\right)}{\sigma_{n-1}\left(A_n 1\right)} = \frac{\pi_0\left(A_{1,n} f\right)}{\pi_0\left(A_{1,n} 1\right)}.
$$

Now, it will be helpful in computing variances in the sequel to define the following.

Definition 1 *The observation co-variability and variability functions are*

$$
\lambda(x, \xi) = \int \frac{g(y - h(x)) g(y - h(\xi))}{g(y)} dy \text{ and } \overline{\lambda}(x) = \lambda(x, x).
$$

Example 1 *Suppose that g is $\mathcal{N}(m, \sigma)$. Then, it follows easily that*

$$
\begin{aligned}
\lambda(x, \xi) &= \frac{1}{\sqrt{2\pi}\sigma} \int \exp\left(\frac{(y-m)^2 - (y - h(x) - m)^2 - (y - h(\xi) - m)^2}{2\sigma^2}\right) dy \\
&= \exp\left(\frac{h(x) h(\xi)}{\sigma^2}\right)
\end{aligned} \tag{12}
$$

so $\overline{\lambda}(x) = \exp\left(\frac{h^2(x)}{\sigma^2}\right)$.

Example 2 *Suppose that g is Laplace with $g(x) = \frac{1}{2}e^{-|x|}$ and the sensor function h is non-negative. Then, when $h(x) \leq h(\xi)$ we have that*

$$\lambda(x,\xi) = \frac{e^{-h(x)-h(\xi)}}{2}\left(\int_{-\infty}^{0} e^{y}dy + \int_{0}^{h(x)} e^{3y}dy\right) + \frac{e^{h(x)-h(\xi)}}{2}\int_{h(x)}^{h(\xi)} e^{y}dy + \frac{e^{h(x)+h(\xi)}}{2}\int_{h(\xi)}^{\infty} e^{-y}dy \quad (13)$$
$$= \frac{e^{-h(x)-h(\xi)}}{2}\left[\frac{2}{3} + \frac{1}{3}e^{3h(x)}\right] + \frac{e^{h(x)-h(\xi)}}{2}\left[e^{h(\xi)} - e^{h(x)}\right] + \frac{e^{h(x)+h(\xi)}}{2}e^{-h(\xi)}$$
$$= \frac{1}{3}e^{-h(x)-h(\xi)} - \frac{1}{3}e^{2h(x)-h(\xi)} + e^{h(x)}.$$

Hence, we have by symmetry that

$$\lambda(x,\xi) = \frac{\left[e^{-h(x)-h(\xi)} - e^{2h(x)\wedge h(\xi) - h(x)\vee h(\xi)}\right]}{3} + e^{h(x)\wedge h(\xi)} \text{ and } \overline{\lambda}(x) = \frac{e^{-2h(x)} - e^{h(x)}}{3} + e^{h(x)}.$$

Notwithstanding the previous calculations, the observation variability function is often difficult to find in closed form. Fortunately, we only use the closed form in our example of Section 4. Our one and two variable *filter* kernels involve the observation variability function and the the Markov kernel K:

$$\overline{K}_\lambda(x,dz) = \overline{\lambda}(x)K(x,dz), \quad (14)$$
$$K_\lambda(x,\xi,dz,d\zeta) = \lambda(x,\xi)K(x,dz)K(\xi,d\zeta). \quad (15)$$

To ease notation, we define the n^{th} step filter kernels

$$\overline{K}^n_\lambda(x,dz) = \int_E \overline{K}^{n-1}_\lambda(y,dz)\overline{K}_\lambda(x,dy) \quad (16)$$
$$K^n_\lambda(x,\xi,dz,d\zeta) = \int_E\int_E K^{n-1}_\lambda(y,\theta,dz,d\zeta)K_\lambda(x,\xi,dy,d\theta) \quad (17)$$

for $n = 2, 3, ...$ The next lemma establishes that the variance of the unnormalized filter is:

$$E^Q[(\sigma_n(f) - E^Q(\sigma_n(f)))^2] = \pi_0 \times \pi_0\left(K^n_\lambda(f\times f)\right) - (\pi_0(K^n f))^2. \quad (18)$$

Lemma 2 *Suppose $f \in B(E)_+$. Then, $E^Q[\sigma_n(f)] = E^Q[L_n f(X_n)] = \pi_0(K^n f)$, $E^Q[(L_n f(X_n))^2] = \pi_0\left(\overline{K}^n_\lambda(f^2)\right)$ and $E^Q[\sigma^2_n(f)] = \pi_0 \times \pi_0\left(K^n_\lambda(f\times f)\right)$.*

Remark 1 *If λ is bounded by B say, then $\pi_0\left(\overline{K}_\lambda^n(f^2)\right) \le |f|_\infty^2 B^n$ and $E^Q[(L_n f(X_n))^2] < \infty$. Moreover, it follows by Example 1 that λ is bounded if the observation noise is Gaussian and the observation function h is bounded. Similarly, it follows by Example 2 that λ is bounded if the observation noise is Laplace and the observation function h is bounded.*

Proof. Taking expectations over Y, we have that

$$E^Q\left[\frac{g\left(Y_j-h\left(x\right)\right)}{g\left(Y_j\right)}\right]=\int g\left(y-h\left(x\right)\right)dy=1,\quad E^Q\left[\frac{(g\left(Y_j-h\left(x\right)\right))(g\left(Y_j-h\left(\xi\right)\right))}{g^2\left(Y_j\right)}\right]=\lambda(x,\xi).$$

Hence, we find by (11) and (10) that

$$\begin{aligned} E^Q\left[\sigma_n\left(f\right)\right] &= \int E^Q\left[\frac{g\left(Y_1-h\left(x\right)\right)}{g\left(Y_1\right)}KA_{2,n}f\left(x\right)\right]\pi_0\left(dx\right) \text{ by Fubini's theorem} \qquad (19)\\ &= \int E^Q[A_{2,n}f\left(z\right)]K\pi_0(dz) \text{ by Fubini, fact } \{Y_n\} \text{ is i.i.d.}\\ &= \int f\left(y\right)\left(K^n\pi_0\right)\left(dy\right) \text{ by recursion.} \end{aligned}$$

Therefore, the required first moments are finite since f is bounded. Next, one has by fact $\{Y_i\}$ is i.i.d. and independent of X that

$$E^Q[(L_nf(X_n))^2]=E^Q\left[\prod_{l=1}^{n}\overline{\lambda}(X_{l-1})f^2(X_n)\right]=E^Q\left[\prod_{l=1}^{n-1}\overline{\lambda}(X_{l-1})\overline{K}_\lambda f^2(X_{n-1})\right]=\pi_0\overline{K}_\lambda^n f^2. \quad (20)$$

Finally, one has by fact $\{Y_i\}$ is i.i.d. and independent of X, (11) and (10) again that

$$E^Q\left[\sigma_n^2\left(f\right)\right]=\int\int E^Q[A_{1,n}f(x)A_{1,n}f(\xi)]\,\pi_0(dx)\pi_0(d\xi) \qquad (21)$$

and

$$\begin{aligned} E^Q[A_{i,n}f(x)A_{i,n}f(\xi)] &= \int\int E^Q[A_{i+1,n}f(z)A_{i+1,n}f(\zeta)]K_\lambda(x,\xi,dz,d\zeta) \qquad (22)\\ &= K_\lambda^{n+1-i}(f\times f)(x,\xi) \text{ by recursion} \end{aligned}$$

for $i=1,2,...,n$. Therefore, we have by substitution of (22) into (21) that $E^Q\left[\sigma_n^2\left(f\right)\right]=\pi_0\times\pi_0\left(K_\lambda^n(f\times f)\right)$, which completes the proof. □

Notice that the variance of the optimal filter propagates using kernal K_λ. We will show that the classical weighted particle filter is propagated using $\overline{K}_\lambda$.

The following result follows by bilinearity and symmetry.

Corollary 3 *Suppose* $f, g \in B(E)_+$ *and* $\pi_0 \times \pi_0 \left(K_\lambda^n(f \times f)\right), \pi_0 \times \pi_0 \left(K_\lambda^n(g \times g)\right) < \infty$. *Then,* $E^Q\left[\sigma_n\left(f\right)\sigma_n\left(g\right)\right] = \pi_0 \times \pi_0\left(K_\lambda^n(f \times g)\right)$.

3 Weighted Particle System

In the sequel, we will fix an observation path, set $Q^Y(\cdot) = Q(\cdot|\mathcal{F}_\infty^Y)$ and let $E^Y[Z]$ denote expectation with respect to Q^Y.

The weighted particle systems does not resample. In this case, the conditional expectation $\sigma_n(f) = E^Q[L_n f(X_n)|\mathcal{F}_n^Y]$ with respect to fictitious probability Q is replaced with independent sample average to arrive at

$$\sigma_n^N\left(f\right) = \frac{1}{N}\sum_{k=1}^{N} L_n^k f\left(\mathfrak{X}_n^k\right), \tag{23}$$

where the *particles* $\left\{\mathfrak{X}^k\right\}_{k=1}^{\infty}$ are independent (π_0, K)-Markov processes that are independent of Y and the *weights* $L_n^k = \prod_{j=1}^{n} W_j^k$ with $W_j^k = \alpha_j(\mathfrak{X}_{j-1}^k) = \frac{g\left(Y_j - h\left(\mathfrak{X}_{j-1}^k\right)\right)}{g(Y_j)}$. Then, $\sigma_0^N\left(f\right) = \frac{1}{N}\sum_{k=1}^{N} f\left(\mathfrak{X}_0^k\right)$. By enlarging the space, we can take the particles $\{\mathfrak{X}^k\}$ on the same space $(\Omega, \mathcal{F}, Q)$ as the signal and observations. Note that $(\mathfrak{X}^k, Y) \overset{D}{=} (X, Y)$ for all k. For convenience, we define the single particle measures

$$\beta_n^k = L_n^k \delta_{\mathfrak{X}_n^k} \ \forall n = 0, 1, ... \ \text{ and } \ \beta_{-1}^k = \pi_0 \ \forall \ k = 1, 2, ..., N.$$

Then, using (11) and (10), one has the following measure-valued evolution

$$
\begin{aligned}
\beta_n^k(f) &= L_{n-1}^k \frac{g\left(Y_n - h\left(\mathfrak{X}_{n-1}^k\right)\right)}{g(Y_n)} Kf(\mathfrak{X}_{n-1}^k) + L_n^k\left(f\left(\mathfrak{X}_n^k\right) - E^Y\left[f(\mathfrak{X}_n^k)\middle|\mathcal{F}_{n-1}^{\mathfrak{X}}\right]\right) \\
&= \beta_{n-1}^k(A_n f) + L_n^k\left(f\left(\mathfrak{X}_n^k\right) - E^Y\left[f\left(\mathfrak{X}_n^k\right)\middle|\mathcal{F}_{n-1}^{\mathfrak{X}}\right]\right) \\
&= \beta_0^k(A_{1,n}f) + \sum_{l=1}^{n} L_l^k\left(A_{l+1,n}f\left(\mathfrak{X}_l^k\right) - E^Y\left[A_{l+1,n}f\left(\mathfrak{X}_l^k\right)\middle|\mathcal{F}_{l-1}^{\mathfrak{X}}\right]\right) \text{ by recursion} \\
&= \pi_0(A_{1,n}f) + M_n^{\beta^k}(f) \\
&= \sigma_n(f) + M_n^{\beta^k}(f),
\end{aligned} \tag{24}
$$

where it follows by the fourth and second equalities in (24) that

$$
\begin{aligned}
M_n^{\beta^k}(f) &= \sum_{l=0}^{n}\left[L_l^k\left(A_{l+1,n}f\left(\mathfrak{X}_l^k\right) - E^Y\left[A_{l+1,n}f\left(\mathfrak{X}_l^k\right)\middle|\mathcal{F}_{l-1}^{\mathfrak{X}}\right]\right)\right] \\
&= \sum_{l=0}^{n}[\beta_l^k(A_{l+1,n}f) - \beta_{l-1}^k(A_{l,n}f)].
\end{aligned} \tag{25}
$$

It follows from the fact that $L_l^k \in \mathcal{F}_\infty^Y \vee \mathcal{F}_{l-1}^{\mathfrak{X}^k}$ that M^{β^k} is a $\{\mathcal{F}_\infty^Y \vee \mathcal{F}_n^{\mathfrak{X}}\}_{n\geq -1}$-martingale and

$$
E^Y[\beta_n^k(f)] = E^Y[\sigma_n(f)] + E^Y[M_n^{\beta^k}(f)] = \sigma_n(f) = \pi_0(A_{1,n}f). \tag{26}
$$

Averaging (24) and (25) over the particles, one has that

$$
\sigma_n^N(f) = \sigma_n(f) + M_n^N(f), \tag{27}
$$

where

$$
M_n^N(f) = \sum_{l=0}^{n}[\sigma_l^N(A_{l+1,n}f) - \sigma_{l-1}^N(A_{l,n}f)].
$$

$\{M_n^N(f),\ n = 0, 1, ...\}$ is a zero-mean $\{\mathcal{F}_\infty^Y \vee \mathcal{F}_n^{\mathfrak{X}}\}_{n\geq -1}$-martingale in n as well as the average $\frac{1}{N}\sum_{k=1}^{N} M_n^{\beta^k}(f)$ of i.i.d. zero-mean random variables both with respect to Q^Y.

Now, for each $n \in \mathbb{N}$, we define the $\mathcal{F}_\infty^Y$-measurable random variance

$$
\gamma_n^W(f) = E^Y[|M_n^{\beta^k}(f)|^2] \tag{28}
$$

when it exists. For comparison purposes, we are also interested in the expectation $E^Q[\gamma_n^W(f)]$ of $\gamma_n^W(f)$. Let

$$A_j^{(2)}f(x)=\begin{cases}\alpha_j^2(x)Kf(x) & j=1,2,...\\ \\ f(x) & j=0\end{cases}\quad \text{and } A_{i,n}^{(2)}f=\begin{cases}A_i^{(2)}\Big(A_{i+1}^{(2)}\cdots\Big(A_n^{(2)}f\Big)\Big) & \forall i\le n\\ \\ f & i=n+1\end{cases}. \tag{29}$$

Lemma 4 *Suppose* $\overline{K}_\lambda g(x)\doteq\int\int g(z,z)\overline{\lambda}(x)K(x,dz)$ *for* $g\in B(E\times E)$. *Then,*

$$\begin{aligned}\gamma_n^W(f)=&\,\pi_0((A_{1,n}f)^2)-(\pi_0(A_{1,n}f))^2 \\ &+\sum_{l=1}^n\pi_0A_{1,l-1}^{(2)}\left[A_l^{(2)}(A_{l+1,n}f)^2-(A_{l,n}f)^2\right]\ \forall f\in B(E)_+.\end{aligned}\tag{30}$$

and

$$\begin{aligned}E^Q[\gamma_n^W(f)]=&\,\pi_0K_\lambda^n(f\times f)-\pi_0\times\pi_0K_\lambda^n(f\times f) \\ &+\sum_{l=1}^n\pi_0\overline{K}_\lambda^{l-1}[\overline{K}_\lambda-K_\lambda]K_\lambda^{n-l}(f\times f)\ \forall f\in B(E)_+.\end{aligned}\tag{31}$$

Remark 2 *Under our conditions,* $A_{l,n}f$ *are bounded functions for each fixed* $Y_1,...,Y_n$ *and* $f\in B(E)_+$. *Therefore, (30) is an* $\mathbb{R}$*-valued random variable. (31) should be interpretted as 'when it makes sense'. In particular, there is potential for* $\infty-\infty$ *situations. The point of the first sentence in the lemma is to explain how we interpret a single variable kernel applied to a two variable function in (31).*

Proof. Since $(\mathfrak{X}^k,Y)\stackrel{D}{=}(X,Y)$ we can just work with (X,Y) and $\beta_l\doteq L_l\delta_{X_l}$. Let M^β be defined as (25) but with β instead of β^k. By the martingale property

$$E^Y[|M_n^\beta(f)|^2]=\sum_{l=0}^n E^Y[|\beta_l(A_{l+1,n}f)-\beta_{l-1}(A_{l,n}f)|^2]. \tag{32}$$

However, by the second equality in (24) and fact $\beta_{l-1}(A_{l,n}f)$ is $\mathcal{F}_{l-1}^X$-measurable

$$E^Y\left[|\beta_l(A_{l+1,n}f)-\beta_{l-1}(A_{l,n}f)|^2\right] \tag{33}$$

$$
\begin{aligned}
&= E^Y\left[|\beta_l(A_{l+1,n}f)|^2\right] - E^Y\left[|\beta_{l-1}(A_{l,n}f)|^2\right] \\
&= E^Y\left[L_l^2(A_{l+1,n}f)^2(X_l)\right] - E^Y\left[L_{l-1}^2(A_{l,n}f)^2(X_{l-1})\right] \\
&= E^Y\left[L_{l-1}^2\{A_l^{(2)}(A_{l+1,n}f)^2(X_{l-1}) - (A_{l,n}f)^2(X_{l-1})\}\right] \\
&= \pi_0 A_{1,l-1}^{(2)}\left[A_l^{(2)}(A_{l+1,n}f)^2 - (A_{l,n}f)^2\right]
\end{aligned}
$$

in the case $l \geq 1$. When $l = 0$ the above equation becomes

$$
E^Y\left[|\beta_l(A_{l+1,n}f) - \beta_{l-1}(A_{l,n}f)|^2\right] = \pi_0((A_{1,n}f)^2) - (\pi_0(A_{1,n}f))^2, \tag{34}
$$

which completes (30). Now, taking expectations of (33) with respect to i.i.d. $\{Y_l\}$ and using (22) as well as Fubini's theorem, one finds that

$$
\begin{aligned}
E^Q\left[|\beta_l(A_{l+1,n}f) - \beta_{l-1}(A_{l,n}f)|^2\right] &= \pi_0 \overline{K}_\lambda^{l-1}[\overline{K}_\lambda E^Q[(A_{l+1,n}f)^2 - E^Q[(A_{l,n}f)^2]] \\
&= \pi_0 \overline{K}_\lambda^{l-1}[\overline{K}_\lambda - K_\lambda]K_\lambda^{n-l}(f \times f).
\end{aligned} \tag{35}
$$

in the case $l \geq 1$. The case $l = 0$ is a simple calculation. □

This leads us to our first main results of this section, which are a strong law of large numbers, a rate of L^2-convergence and a quenched central limit theorem.

Theorem 5 *For Q-almost all Y, the weighted particle system satisfies:*

SLLN $\sigma_n^N \Rightarrow \sigma_n$ *(i.e. weak convergence) a.s. $[Q^Y]$*

L^2-rates $E^Y\left|\sigma_n^N(f) - \sigma_n(f)\right|^2 = \frac{\gamma_n^W(f)}{N}$ *for Q-almost all Y for all $f \in \overline{C}(E)_+$*

CLT $\sqrt{N}\left(\sigma_n^N(f) - \sigma_n(f)\right) \Rightarrow \mathcal{N}\left(0, \sqrt{\gamma_n^W(f)}\right)$ *for Q-almost all Y for all $f \in \overline{C}(E)_+$.*

Proof. The theorem follows by the fact $\sigma_n^N(f) - \sigma_n(f) = \frac{1}{N}\sum_{k=1}^{N} M_n^{\beta^k}(f)$, (30), the classical strong law of large numbers and the classical central limit theorem. Note that $M_n^{\beta^k}(f)$ has bounded second moment by the remark following (30). Also recall the $\{f_i\} \subset B(E)_+$ defined in (8) and note $\sigma_n^N(f_i) \to \sigma_n(f_i)$ a.s. $[Q^Y]$ for all i implies $\sigma_n^N \Rightarrow \sigma_n$ a.s. $[Q^Y]$. □

Since $N(\sigma_n^N(f) - \sigma_n(f))$ is a sum of i.i.d. random variables with finite second moment (with respect to Q^Y), we can use the classical probability results on this weighted particle filter, which will characterize how the filter improves as we increase the number of particles. To do this, we assign the $C([0,1])$-valued random variables:

$$\tilde{\sigma}_n^N(f,t) = \sum_{k=1}^{\lfloor Nt \rfloor} (\beta_n^k(f) - \sigma_n(f)) + \left(t - \frac{\lfloor Nt \rfloor}{N}\right) (\beta_n^{\lfloor Nt \rfloor + 1}(f) - \sigma_n(f)) \ \forall t \in [0,1] \tag{36}$$

for $f \in B(E)_+$, let W denote a standard Brownian motion and define compact subset of $C([0,1])$

$$\mathcal{S} = \left\{ g : g(t) = \int_0^t \phi(u) du \ \forall t \in [0,1], \ \int_0^t \phi^2(u) du \leq 1 \right\}. \tag{37}$$

We also define the cummulant generating and rate functions:

$$\Lambda(\lambda) \doteq \log E^Y \left[\exp(\lambda \beta_n^1(f)) \right] \text{ and } I(x) \doteq \sup_{\lambda \in \mathbb{R}} \{ \lambda x - \Lambda(\lambda) \}. \tag{38}$$

Theorem 6 *For Q-almost all Y and all $f \in \overline{C}(E)_+$, the weighted particle system satisfies:*

Donsker's CLT: $\frac{1}{\sqrt{N \gamma_n^W(f)}} \tilde{\sigma}_n^N(f, \cdot) \Rightarrow W$ *in* $C([0,1])$.
Strassen's LIL: $\frac{1}{\sqrt{N \gamma_n^W(f) \log\log(N)}} \tilde{\sigma}_n^N(f, \cdot) \to\to \mathcal{S}$ *a.s.* $[Q^Y]$.
Cramer's LDP: $\limsup_{N \to \infty} \frac{1}{N} \log Q^Y(\sigma_n^N(f) \in F) \leq - \inf_{x \in F} I(x)$ *for all closed* F *and* $\liminf_{N \to \infty} \frac{1}{N} \log Q^Y(\sigma_n^N(f) \in G) \geq - \inf_{x \in G} I(x)$ *for all open* G.

Remark 3 *We used the notation $\to\to \mathcal{S}$ to mean that the set of cluster points as $N \to \infty$ is equal to $\mathcal{S}$.*

It appears from these (quenched) results that weighted particle filter works perfectly and one need only supply enough particles to acheive the desired performance. However, there is a catch, which is explained in the following section.

4 Weighted Particle Filter Failure

The weighted particle filter can basically fail due to particle spread. We show in this section that this problem can be so bad that adding more particles still can not solve it. This failure is best explained by comparing K_λ and $\overline{K}_\lambda$ in a setting where explicit calculations are manageable.

Example 3 *Suppose* $h(x) = x$ *and* $g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ *so* $\lambda(x, \xi) = \exp(x\xi)$ *by Example 1. Moreover, let* $K(x, dz) = \frac{1}{\sqrt{\pi}} e^{-(x-z)^2} dz$ *so* $\overline{K}_\lambda(x, dz) = \dfrac{1}{\sqrt{\pi}} e^{2xz-z^2} dz$ *and we have*

$$X_k = X_{k-1} + W_k \tag{39}$$
$$Y_k = X_{k-1} + V_k \tag{40}$$

with $\{(W_k, V_k)\}_{k=1}^\infty$ *being i.i.d.* $\mathcal{N}\left(0, \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \end{bmatrix}\right)$, *so the filter could be solved by the Kalman filter if* X_0 *is independent and Gaussian. However, one might still use a particle filter if one wants to do model selection or if the initial condition is not Gaussian.*

Using $\overline{K}_\lambda^{l+1}(x, dz) = \int \overline{K}_\lambda^l(\zeta, dz) \overline{K}_\lambda(x, d\zeta)$, *we find that*

$$\begin{aligned} \overline{K}_\lambda^2(x, dz) &= \frac{1}{\pi}\left[\int \exp\left(2\zeta z - z^2 + 2x\zeta - \zeta^2\right) d\zeta\right] dz \\ &= \frac{1}{\pi}\int \exp\left(-(\zeta - x - z)^2\right) d\zeta \exp\left((x+z)^2 - z^2\right) dz \\ &= \frac{1}{\sqrt{\pi}} \exp\left(x^2 + 2xz\right) dz \text{ for any } x, z \in \mathbb{R} \end{aligned}$$

and

$$\overline{K}_\lambda^3(x, dz) = \frac{1}{\pi}\left[\int \exp\left(\zeta^2 + 2\zeta z + 2x\zeta - \zeta^2\right) d\zeta\right] dz = \infty \text{ for any } x, z \in \mathbb{R}. \tag{41}$$

On the other hand,

$$K_\lambda(x,\xi,dz,d\zeta)=\frac{1}{\pi}e^{x\xi-x^2-z^2-\xi^2-\zeta^2+2xz+2\xi\zeta}dzd\zeta. \tag{42}$$

Hence, using $K_\lambda^{l+1}(x,\xi,dz,d\zeta)=\int\int K_\lambda^l(y,\theta,dz,d\zeta)\,K_\lambda(x,\xi,dy,d\theta)$ *, we find that*

$$\begin{aligned}
&\frac{K_\lambda^2(x,\xi,dz,d\zeta)}{dzd\zeta} \qquad (43)\\
&=\frac{\exp\left(x\xi-x^2-\xi^2-z^2-\zeta^2\right)}{\pi^2}\\
&\times\int\int\exp\left(y\theta+2\left[xy+\xi\theta+yz+\theta\zeta-y^2-\theta^2\right]\right)dyd\theta\\
&=\frac{1}{\pi^2}\exp\left(\frac{19x\xi+16xz+16\xi\zeta+4x\zeta+4z\xi+4z\zeta-7x^2-7\xi^2-7z^2-7\zeta^2}{15}\right)\\
&\times\int\int\exp\left(-\begin{bmatrix}y-\frac{8z+8x+2\zeta+2\xi}{15} & \theta-\frac{8\zeta+8\xi+2z+2x}{15}\end{bmatrix}\begin{bmatrix}2 & -\frac{1}{2}\\ -\frac{1}{2} & 2\end{bmatrix}\begin{bmatrix}y-\frac{8z+8x+2\zeta+2\xi}{15}\\ \theta-\frac{8\zeta+8\xi+2z+2x}{15}\end{bmatrix}\right)dyd\theta\\
&=\frac{\sqrt{15}}{\pi}\exp\left(\frac{19x\xi+16xz+16\xi\zeta+4x\zeta+4z\xi+4z\zeta-7x^2-7\xi^2-7z^2-7\zeta^2}{15}\right).
\end{aligned}$$

Moreover, letting

$$a=\frac{44\,(16z+4\zeta+30x)+19\,(16\zeta+4z+30\xi)}{44^2-19^2},b=\frac{19\,(16z+4\zeta+30x)+44\,(16\zeta+4z+30\xi)}{44^2-19^2},$$

we find

$$\begin{aligned}
\frac{K_\lambda^3(x,\xi,dz,d\zeta)}{dzd\zeta}&=\int K_\lambda^2(y,\theta,dz,d\zeta)\,K_\lambda(x,\xi,dy,d\theta) \qquad (44)\\
&=\frac{2\sqrt{7}}{\pi^2}\exp\left(\frac{4z\zeta-7z^2-7\zeta^2}{15}+x\xi-x^2-\xi^2\right)\\
&\times\int\exp\left(\frac{19y\theta+16yz+16\theta\zeta+4y\zeta+4z\theta-22y^2-22\theta^2}{15}+2xy+2\xi\theta\right)dyd\theta\\
&=\frac{\sqrt{15}}{\pi^2}\exp\left(\frac{22a^2+22b^2-19ab+4z\zeta-7z^2-7\zeta^2}{15}+x\xi-x^2-\xi^2\right)\\
&\times\int\exp\left(-\begin{bmatrix}y-a & \theta-b\end{bmatrix}\begin{bmatrix}\frac{22}{15} & -\frac{19}{30}\\ -\frac{19}{30} & \frac{22}{15}\end{bmatrix}\begin{bmatrix}y-a\\ \theta-b\end{bmatrix}\right)dyd\theta
\end{aligned}$$

$$= \frac{\sqrt{105}}{\pi} \exp\left(\frac{22a^2 + 22b^2 - 19ab + 4z\zeta - 7z^2 - 7\zeta^2}{15} + x\xi - x^2 - \xi^2 \right).$$

Substituting $n = 3$ as well as the values for $K^1_\lambda, K^2_\lambda, K^3_\lambda$ and $\overline{K}^1_\lambda, \overline{K}^2_\lambda, \overline{K}^3_\lambda$ into (31), we see the expected weighted particle filter variance $E^Q[\gamma_3^W(f)] = \infty$ (since $\overline{K}^3_\lambda = \infty$) for any non-trivial non-negative f. Therefore, taking expectations in the L^2-rates of Theorem 5, one finds $E^Q[(\sigma_3^N(1) - \sigma_3(1))^2] = \infty$ for all N so the weighted particle filter can not really work as a model selection nor a tracking device.

Notice the variance of the unnormalized filter itself $\sigma_3(f)$, given in (18), is finite as it only involves the kernals $K^1_\lambda, K^2_\lambda, K^3_\lambda$.

This example also illustrates the importance of the observation variability function. The unbounded nature of the observation variability function adversely affects expected variances.

References

[1] Ballantyne, D.J., Chan, H.Y. and Kouritzin, M.A. (2000). A novel branching particle method for tracking, in *Signal and Data Processing of Small Targets2000, Proceedings of SPIE*, **4048** Ed. Oliver E. Drummond 277–287.

[2] Billingsley, P. (1999). *Convergence of Probability Measures.* Wiley, New York.

[3] Blount, D. and Kouritzin, M. A. (2010). On convergence determining and separating classes of functions. *Stoch. Proc. Appl.* **120**, 1898–1907.

[4] H. Cox: On the estimation of state variables and parameters for noisy systems. IEEE Trans. Auto. Contr. **AC-9**, 5-12 (1964).

[5] Crisan, D. and Lyons, T.J. (1997). Nonlinear filtering and measure valued processes, *Prob. Theory Related Fields*, **109**: 217-244.

[6] Crisan, D., Gaines, J., Lyons, T.J. (1998). Convergence of a branching particle method to the solution of the Zakai equation. *SIAM J. Appl. Math.* **58**, 1568-1590.

[7] Del Moral, P., Kouritzin, M.A., and Miclo, L. (2001). On a class of discrete generation interacting particle systems. *Electronic Journal of Probability* **6**: Paper No. 16, 26 p.

[8] Gordon, N.J., Salmond, D.J. and Smith, A.F.M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE Proceedings-F, Radar and Signal Processing* **140**, 107-113.

[9] Hammersley, J. M. and Morton, K. W. (1954). Poor Man's Monte Carlo. *Journal of the Royal Statistical Society.* Series B (Methodological) **16**, 23-38.

[10] J. Handschin, "Monte Carlo techniques for prediction and filtering of non-linear stochastic processes," Automatica, vol. 6, pp. 555563, 1970.

[11] Handschin, J. and Mayne, D. (1969). "Monte Carlo techniques to estimate the conditionnal expectation in multi-stage non-linear filtering," in Int. J. Control, vol. 9, , pp. 547559.

[12] Kouritzin, M.A. and Zeng, Y. (2005). Bayesian Model Selection via Filtering for a Class of Micro-movement Models of Asset Price. *International Journal of Theoretical and Applied Finance* **8**, 97-122.

[13] Kurtz, T.G., Xiong, J. (1999). Particle representations for a class of nonlinear SPDEs. *Stochastic Process. Appl.* **83**, 103-126.

[14] Kurtz, T.G. and Xiong, J. (2000). Numerical solutions for a class of SPDEs with application to filtering, Stochastics in Finite and Infinite Dimension: In Honor of Gopinath Kallianpur. Edited by T. Hida, R. Karandikar, H. Kunita, B. Rajput, S. Watanabe and J. Xiong. *Trends in Mathematics.* Birkhauser.

[15] Zakai, M. (1969). "On the optimal filtering of diffusion processes". *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **11**: 230-243.