GENERALISED PARTICLE FILTERS WITH GAUSSIAN MEASURES

Dan Crisan and Kai Li

Department of Mathematics, Imperial College London
180 Queen's Gate, London, SW7 2AZ, United Kingdom
{d.crisan, kai.li08}@imperial.ac.uk

ABSTRACT

The stochastic filtering problem deals with the estimation of the posterior distribution of the current state of a signal process $X = \{X_t\}_{t\geq 0}$ given the information supplied by an associate process $Y = \{Y_t\}_{t\geq 0}$. The scope and range of its applications includes the control of engineering systems, global data assimilation in meteorology, volatility estimation in financial markets, computer vision and vehicle tracking. A massive scientific and computational effort is dedicated to the development of viable tools for approximating the solution of the filtering problem. Classical PDE methods can be successful, particularly if the state space has low dimensions. In higher dimensions, a class of numerical methods called particle filters have proved the most successful methods to-date. These methods produce an approximations of the posterior distribution by using the empirical distribution of a cloud of particles that explore the signal's state space. We discuss here a more general class of numerical methods which involve *generalised* particles, that is, particles that evolve through larger spaces. Such generalised particles include Gaussian measures, wavelets, and finite elements in addition to the classical particle methods. We will construct the approximating particle system under the Gaussian measure framework and prove the corresponding convergence result.

1. THE FILTERING FRAMEWORK

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space on which we have defined a process $X = \{X_t\}_{t\geq 0}$ called the signal and an associate process $Y = \{Y_t\}_{t\geq 0}$ called the observation. The process X is the solution of a d-dimensional stochastic differential equation driven by a p-dimensional Brownian motion V, that is:

$$X_t = X_0 + \int_0^t f(X_s)ds + \int_0^t \sigma(X_s)dV_s. \tag{1}$$

where $f : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times p}$ are bounded and globally Lipschitz. We denote the distribution of X_0 by π_0. The process Y satisfies the evolution equation

$$Y_t = Y_0 + \int_0^t h(X_s)ds + W_t, \tag{2}$$

where $h : \mathbb{R}^d \to \mathbb{R}^m$ is a measurable function with linear growth and W is an m-dimensional Brownian motion independent of X. Let $\mathscr{Y} = (\mathscr{Y}_t)_{t\geq 0}$ be the filtration generated by the observation process Y. The σ-field $\mathscr{Y}_t$ is a mathematical model for the information available to us at time t. Mathematically, the solution of the filtering problem is the conditional distribution of the signal process X_t, given the σ-field $\mathscr{Y}_t$, denoted in what follows by π_t. The (random) measure π_t is also called the posterior distribution of the signal. The process $\pi = \{\pi_t\}_{t\geq 0}$ can be viewed as a stochastic process taking values in an infinite dimensional space: the space of probability measures over the state space of the signal.

Under additional assumptions, one can show that π satisfies a non-linear stochastic PDE called the Kushner-Stratonovich equation. Moreover a certain unnormalised version of π denoted by $\rho = \{\rho_t\}_{t\geq 0}$ satisfies a simpler, linear stochastic PDE, called the Duncan-Mortensen-Zakai equation. We have (see [23])

$$\rho_t(\varphi_t) = \pi_0(\varphi_0) + \int_0^t \rho_s(\frac{\partial \varphi_s}{\partial s} + A\varphi_s)ds + \int_0^t \rho_s(\varphi_s h^\top)dY_s \tag{3}$$

where A is the infinitesimal generator of the signal process X defined as

$$A = \sum_{i=1}^d f^i \frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i=1}^d \sum_{j=1}^d (\sigma\sigma^\top)_{ij} \frac{\partial^2}{\partial x_i \partial x_j}. \tag{4}$$

and $\varphi_t(x) = \varphi(t,x) : [0,\infty) \times \mathbb{R}^d \to \mathbb{R}^d$ is a bounded Borel measurable function continuously differentiable in the time variable and twice differentiable in the space variable with all derivatives bounded. By Kallianpur-Striebel formula (see, e.g. Chapter 8 in [11]) we have

$$\pi_t\varphi = \frac{\rho_t(\varphi)}{\rho_t(\mathbf{1})} \qquad \mathbb{P}-a.s., \tag{5}$$

where $\mathbf{1}$ is the constant function 1.

It is possible to explicitly describe π if the system is linear. Outside the linear context, it is known that only a few very exceptional examples have explicitly described posterior distributions (for example, the Benes filter). Therefore, typically the process π or (ρ) can only be approximated numerically. There are a wide variety of approaches to estimate π, some more successful than others (see e.g. [3], [4], [7] and [15] and the references therein, see also Cahpters 8 and 9 in [2]).

Particle methods ([2, 8]) are one of the most effective and versatile methods for solving the filtering problem numerically, and their development has been intensified in the last two decades (see Chapter 7 in [7]). These methods approximate the process π_t (or ρ_t) with discrete random measures (sums of Dirac measures) of the form

$$\alpha_n = \sum_i a_i(t)\delta_{v_i(t)}$$

The work of D. Crisan was partially supported by the EPSRC Grant No: EP/H0005500/1.

The work of K. Li was partially supported by the Department of Mathematics, Imperial College London.

© EURASIP, 2011 - ISSN 2076-1465

with stochastic weights $a_1(t), a_2(t), \ldots$, and corresponding stochastic positions $v_1(t), v_2(t), \ldots$, where $v_j(t) \in \mathbb{R}^d$. The measure α_n can be thought of as the empirical distribution of particles whose positions are given by the processes v_i and carrying the weights a_i.

2. GENERALISED PARTICLE FILTERS

In this section, we introduce the concept of *generalised particle filters*. In this case, the positions of the (generalised) particles are in (possibly) larger spaces than the state space of the signal process X.

The reason for introducing the generalised particle is twofold. Firstly, the particles involved in the classical particle filter carry information about their position and their weight. One can interpret the system of particles as a *quantization* of the posterior distribution π, respectively, of the unnormalised conditional distribution of the signal. This limited information may be wasteful. Indeed, it may be the case that if we allow more information to be carried by the particle then perhaps we will need a smaller number. Therefore we may be able to reduce the overall computational effort.

The second reason is that we can integrate within the framework of generalised particle framework a wide variety of numerical methods including

- *Classical Particle Filters*: as explained above, in this case the particles carry information about their weight and position.
- *Gaussian Mixtures*: the particles are in this case characterised by Gaussian measures. They are parameterised by their weights, mean values and the corresponding the covariance matrix.
- *Wavelets*: an orthonormal wavelets series with proper selected dilation and translation parameters is chosen to characterise particles. The transition centres are viewed as positions; and the weights of the particles are the inner products of the wavelets and a certain chosen density function.
- *Finite Elements*: the shape functions of a finite element are considered as the positions of the generalised particles, and the nodal variables should act as the generalised weights.

In what follows, we will discuss the second class of generalised particles, i.e., Gaussian mixtures. The classic particle filters use a mixture of Dirac measures to construct the particle approximation; several attempts have been made to generalise this idea. Kotecha and Djurić ([12]) first introduced the so called Gaussian particle filters, where they used a single Gaussian to approximate the posterior distribution. They shortly improved their initial work and built the approximations by weighted Gaussian mixtures. Up to now, most of the existing work has been closely related to the extended (or ensemble) Kalman filter, because of the Gaussian nature of Kalman filter, and hence this method may provide a way to improve the asymptotic behaviour for the ensemble Kalman filter (see discussions in [14]). The majority of the previous work is in the discrete time framework. Reich ([18]) recently took a Gaussian mixture to generalise the ensemble Kalman filter and designed a new algorithm based on continuous time formulation. See Flament et al ([10]), Van der Merwe and Wan ([22]), and Carmi et al ([5]) for more related work. However, there has been no existing rigorous mathematical study on the convergence results of such Gaussian mixture methods. Therefore, it is of great interests to fill this gap; and this is the aim of the present work.

In the following section, a new algorithm using Gaussian mixtures to approximate the continuous-time nonlinear filtering problem is presented. Compared with most of the existing related work, which mainly considered the problem from an ad-hoc perspective and relied on the numerical implementation; this new algorithm is investigated mathematically and a rigorous proof of its convergence rate is given.

2.1 The Approximating Algorithm

Let $\Delta = \{0 = \delta_0 < \delta_1 < \cdots < \delta_N = T\}$ be an equidistant partition of the interval $[0,T]$ with equal length, with $\delta_i = i\delta, \quad i = 1, \ldots, N$; and $N = \frac{T}{\delta}$. We introduce the following algorithm involving mixtures of Gaussian measures.

Initialization: At time zero, the particle system consists of n Gaussian measures all with equal weights $1/n$, initial positions $v_j^n(0)$, and initial covariance matrices $\omega_j^n(0)$, for $j = 1, \ldots, n$; and the associated Gaussian measure is denoted by $\Gamma_{v_j^n(0),\omega_j^n(0)}$. The approximation of ρ_0 has the form

$$\rho_0^n \triangleq \frac{1}{n} \sum_{j=1}^{n} \Gamma_{v_j^n(0),\omega_j^n(0)}, \tag{6}$$

Recursion: During the interval $t \in [i\delta, (i+1)\delta)$, the approximation ρ^n of the unnormalised conditional distribution ρ will take the form

$$\rho_t^n \triangleq \frac{1}{n} \sum_j a_j^n(t) \Gamma_{v_j^n(t),\omega_j^n(t)}, \tag{7}$$

where $v_j^n(t)$ denotes the mean and $\omega_j^n(t)$ denotes the covariance matrix of the Gaussian measure $\Gamma_{v_j^n(t),\omega_j^n(t)}$, and $a_j^n(t)$ is the weight of the particle. Obviously, each particle is characterised by the triple process $(a_j^n, v_j^n, \omega_j^n)$ which is assumed to evolve as

$$\begin{cases} a_j^n(t) = 1 + \int_{i\delta}^t a_j^n(s) h(v_j^n(s)) dY_s, \\ v_j^n(t) = v_j^n(i\delta) + \int_{i\delta}^t f\left(v_j^n(s)\right) ds \\ \qquad\qquad + (1-\alpha) \int_{i\delta}^t \sigma\left(v_j^n(s)\right) dV_s^{(j)}, \\ \omega_j^n(t) = \alpha \int_{i\delta}^t (\sigma\sigma^\top)\left(v_j^n(s)\right) ds, \end{cases} \tag{8}$$

where $\{V^{(j)}\}_{j=1}^n$ are mutually independent d-dimensional Brownian motions and independent of Y. The parameter α is a real number in the interval $[0,1]$. For $\alpha = 0$ we recover the classic particle approximation (see, for example, Chapter 9 in [2]); for $\alpha = 1$ the mean of the Gaussian measures evolve deterministically (the stochastic term is eliminated).

At the end of the interval $[i\delta, (i+1)\delta)$, each particle branches into a number of offsprings. We denote by $o_j^{n,(i+1)\delta}$ the number of offspring produced by the j-th random variable at time $(i+1)\delta$. We set

$$o_j^{n,(i+1)\delta} = \begin{cases} \left[n a_j^{n,(i+1)\delta}\right] & \text{with prob.} \quad 1 - \{n a_j^{n,(i+1)\delta}\} \\ \left[n a_j^{n,(i+1)\delta}\right] + 1 & \text{with prob.} \quad \{n a_j^{n,(i+1)\delta}\} \end{cases} \tag{9}$$

Then from (9), we have

$$\mathbb{E}\left[o_{j'}^{n,(i+1)\delta}|\mathscr{F}_{(i+1)\delta-}\right] = na_{j'}^{n,(i+1)\delta} \tag{10}$$

and therefore no bias is introduced at branching times $\mathbb{E}\left[\rho^n_{(i+1)\delta}|\mathscr{F}_{(i+1)\delta-}\right] = \rho^n_{(i+1)\delta-}$ The offsprings of the j-th particle will be Dirac measures. We can think of them as being degenerate Gaussian measures with means that are independent samples from the parent particle $\Gamma_{v_j^n(t),\omega_j^n(t)}$ and null covariance matrices.

After branching all the particles are re-indexed from 1 to n and all of the unnormalised weights are re-initialised back to 1; and the particles evolve following (8) again.

The recursion is repeated N times until we reach the terminal time T, where we obtain the approximation ρ_T^n of ρ_T.

3. CONVERGENCE RESULTS

The approximating sequence $\rho^n = \{\rho_t^n; t \geq 0\}$ constructed using the procedure described in the previous section satisfies the evolution equation described in the following procedure (for simplicity of notations we consider one-dimensional case hereafter):

Proposition 3.1 *The measure-valued process $\rho^n = \{\rho_t^n : t \geq 0\}$ satisfies the following evolution equation:*

$$\begin{aligned}\rho_t^n(\varphi) = \pi_0^n(\varphi) &+ \int_0^t \rho_s^n(A\varphi)ds + \int_0^t \rho_s^n(h\varphi)dY_s \\ &+ \int_0^t R_s^1(\varphi)ds + \int_0^t R_s^2(\varphi)dY_s + \frac{1}{n}\sum_{j=1}^n \int_0^t R_s^3(\varphi)dV_s^{(j)} \\ &+ \alpha N_t + M_{[t/\delta]}^{n,\varphi},\end{aligned} \tag{11}$$

for $\varphi \in C_b^m(\mathbb{R})$ and $t \in [0,T]$ with $m \geq 5$. In (11), the terms $R_t^1(\varphi)$, $R_t^2(\varphi)$, and $R_t^3(\varphi)$ are of order δ, N_t is the following martingale:

$$N_t = \frac{1}{n}\sum_{j=1}^n \int_0^t a_j^n(s)\varphi'(v_j^n(s))\sigma(v_j^n(s))dV_s^{(j)};$$

and $M^{n,\varphi} = \{M_i^{n,\varphi}, i > 0 \text{ and } i \in \mathbb{N}\}$ is the discrete process

$$\begin{aligned}M_i^{n,\varphi} = \frac{1}{n}\sum_{k=1}^i \sum_{j=1}^n \Bigg[& o_j^{n,k\delta}\varphi(X_{j'}^n(k\delta)) \\ & - a_j^n(k\delta-)\int_{\mathbb{R}} \varphi(y)\frac{e^{-\frac{(y-v_j^n(k\delta-))^2}{2\omega_j^n(k\delta-)}}}{\sqrt{2\pi\omega_j^n(k\delta-)}}dy\Bigg]\end{aligned} \tag{12}$$

where $X_{j'}^n(k\delta) \sim N(v_{j'}^n(k\delta), \omega_{j'}^n(k\delta))$ is a Gaussian random variable.

Consider the time inhomogeneous Zakai equation (3), and fix $s > 0$. Define

$$\tilde{\varphi}(t,x) \equiv P_{s-t}\varphi(x), \qquad t \in [0,s]$$

where $(P_r)_{r\geq 0}$ is the Markov semigroup whose infinitesimal generator is the differential operator A (as defined in (4)). Then the Zakai equation (3) becomes (see [20])

$$\rho_t(\varphi) = \rho_0(P_s\varphi) + \int_0^t \rho_r(hP_{s-r}\varphi)dY_r. \tag{13}$$

Similarly for $\rho_t^n(\varphi)$ we rewrite (11) for $t \in [0,s]$ and get

$$\begin{aligned}\rho_t^n(\varphi) = &\rho_0^n(P_s\varphi) + \int_0^t \rho_r^n(hP_{s-r}\varphi)dY_r \\ &+ \int_0^t R_r^1(P_{s-r}\varphi)dr + \int_0^t R_r^2(P_{s-r}\varphi)dY_r + M_{[t/\delta]}^{n,\varphi} \\ &+ \frac{1}{n}\sum_{j=1}^n \int_0^t R_r^3(P_{s-r}\varphi)dV_r^{(j)} + N_t;\end{aligned} \tag{14}$$

and the error of the approximation has the representation

$$\begin{aligned}(\rho_t^n - \rho_t)(\varphi) = &(\rho_0^n - \rho_0)(P_s\varphi) + M_{[t/\delta]}^{n,\varphi} \\ &+ \int_0^t R_r^1(P_{s-r}\varphi)dr + \int_0^t R_r^2(P_{s-r}\varphi)dY_r \\ &+ \frac{1}{n}\sum_{j=1}^n \int_0^t R_r^3(P_{s-r}\varphi)dV_r^{(j)} + N_t \\ &+ \int_0^t (\rho_r^n - \rho_r)(hP_{s-r}\varphi)dY_r\end{aligned} \tag{15}$$

where $M_i^{n,\varphi}$ and N_t are the same as in Proposition 3.1, except that φ replaced by $P_{s-t}\varphi$. In the following, we will use the norm $\|\cdot\|_{m,\infty}$ ($m \geq 0$) defined as

$$\|\varphi\|_{m,\infty} = \sum_{|\eta|\leq m} \sup_{x\in\mathbb{R}^d} |D_\eta\varphi(x)|,$$

where $\eta = (\eta^1, \ldots, \eta^d)$ is a multi-index and $D_\eta = (\partial_1)^{\eta_1}\cdots(\partial_d)^{\eta_d}$. In this section, we assume $d = 1$ for simplicity of notations. By computing the bounds on the first six terms (excluding the final term) on the right hand side of (15), one obtains the rate of convergence of the approximation in terms of the three parameters n, δ and α. In what follows we will assume, without loss of generality that $\delta < 1$.

Theorem 3.2 *If the coefficients σ, f, and h are bounded and Lipschitz, then for any $T \geq 0$, $m \geq 5$, there exists a constant c_0^T independent of n, δ or α such that for any $\varphi \in C_b^m(\mathbb{R})$, we have for $t \in [0,T]$*

$$\tilde{\mathbb{E}}\left[(\rho_t^n(\varphi) - \rho_t(\varphi))^2\right] \leq c_0^T\|\varphi\|_{m,\infty}^2 c(n,\delta,\alpha), \tag{16}$$

where

$$c(n,\delta,\alpha) = \max\left\{\frac{1}{n},\ \alpha^2\delta^2,\ \frac{1}{\sqrt{\delta n}},\ \frac{\alpha}{n},(1-\alpha)^2\alpha^2\delta^2\right\}$$

The following is a brief discussion of the method of proof of Theorem 3.2. The basis of the result is a variation of a convergence criterion introduced in [16] (Theorem 4.9). Typically the criterion establishes the convergence of finite *signed* measure valued processes provided they satisfy certain evolution equations with terms that can be controlled as the convergence parameter (in our case the number of Gaussian measures comprising the approximation) increases. We use the fact that both the operators P_s and hP_{s-r} in (15) are bounded and linear; we then calculate the upper bounds of the second moments for the first six terms on the right hand side of (15), which are obtained as the terms in the coefficient $c(n,\delta,\alpha)$ (ignoring the constants). The conclusion in Theorem 3.2, namely inequality (16), is then obtained by applying Theorem 4.9 in [16] to equation

(15) and the corresponding second moments.

In what follows, we should determine $c(n,\delta,\alpha)$ to obtain the L^2-convergence rate of the approximation process ρ_t^n.

When $\alpha = 0$ in (8), the component Gaussian measures have null covariance matrices, in other words they are Dirac measures. In this case ρ_n is nothing other than the classic particle filter (see, for example, [2]). In this case several terms in $c(n,\delta,\alpha)$ coming from the covariance term disappear. The rate of convergence $c(n,\delta,0)$ becomes:

$$c(n,\delta,0) = \max\left\{\frac{1}{n},\ \frac{1}{\sqrt{\delta}n}\right\}.$$

Obviously the fastest rate is obtained when δ is a fixed constant independent of n.The L^2-convergence rate will be in this case of order $1/n$, which concides with the results in [2].

When $\alpha \in (0,1]$, the rate of convergence can deteriorate. First of all let us observe that we still need to choose δ to be a fixed constant independent of n. Then the convergence depends on the simpler coefficient $c(n,\alpha)$ given by

$$c(n,\alpha) = \max\left\{\frac{1}{n},\ \alpha^2, (1-\alpha)^2\alpha^2\right\}$$

In this case we need to choose $\alpha = \frac{1}{\sqrt{n}}$ (or of order $1/\sqrt{n}$) to ensure the optimal rate of convergence, which equals to $1/n$.

The following theorem, as a direct corollary of Theorem 3.2, gives the convergence for both ρ_t^n and its normalized version for fixed partition mesh δ and $\alpha = 1/\sqrt{n}$:

Theorem 3.3 *If the coefficients σ, f, and h are bounded and Lipschitz, then for any $T \geq 0$, $m \geq 5$, there exist constants c_1^T and $\tilde{c}_1^T$ independent of n, such that for any $\varphi \in C_b^m(\mathbb{R})$, we have for $t \in [0,T]$*

$$\tilde{\mathbb{E}}\left[(\rho_t^n(\varphi) - \rho_t(\varphi))^2\right] \leq \frac{c_1^T}{n}\|\varphi\|_{m,\infty}^2; \tag{17}$$

and for the normalised conditional distribution $\pi_t(\varphi)$ and its approximation $\pi_t^n(\varphi) \triangleq \rho_t^n(\varphi)/\rho_t^n(\mathbf{1})$:

$$\tilde{\mathbb{E}}\left[|\pi_t^n(\varphi) - \pi_t(\varphi)|\right] \leq \frac{\tilde{c}_1^T}{\sqrt{n}}\|\varphi\|_{m,\infty}. \tag{18}$$

A stronger convergence result can be proved, from which we can see that the convergence are uniform in time t.

Theorem 3.4 *If the coefficients f, σ, and h are bounded and Lipschitz, then for any $T \geq 0$, $m \geq 5$, there exist constants c_2^T and $\tilde{c}_2^T$ independent of n such that for any $\varphi \in C_b^{m+2}(\mathbb{R})$,*

$$\tilde{\mathbb{E}}\left[\sup_{t\in[0,T]}(\rho_t^n(\varphi) - \rho_t(\varphi))^2\right] \leq \frac{c_2^T}{n}\|\varphi\|_{m+2,\infty}^2; \tag{19}$$

and for the normalised conditional distribution π_t^n,

$$\tilde{E}\left[\sup_{t\in[0,T]}|\pi_t^n(\varphi) - \pi_t(\varphi)|\right] \leq \frac{\tilde{c}_2^T}{\sqrt{n}}\|\varphi\|_{m+2,\infty}. \tag{20}$$

Remark 3.5 *The fact that the optimal value for α decreases with n is not surprising. As the number of particles increases, the quantization of the posterior distribution becomes finer and finer. Therefore,* asymptotically*, the position and the weight of the particle provide sufficient information to obtain a good approximation.*

Remark 3.6 *Since the approximations ρ_t^n and π_t^n have smooth densities with respect to the Lebesgue measure, it makes it possible to study various properties the density of ρ_t and that from its approximation ρ_t^n (for example, the position of their maximum value, the decay in time, the properties of their derivatives, etc). This would not be possible under the classic particle filtering framework, where the approximations are linear combinations of Dirac measures, unless a smoothing procedure is applied first.*

4. CONCLUSION AND FURTHER WORK

In this paper we have analysed a class of approximations of the posterior distribution under continuous time framework. The approximation uses mixtures of Gaussian measures to approximate the posterior distribution. The L^2-convergence rate of such approximation is given. This method can be viewed as a natural extension of the classic particle filters, in the sense that the classic one is a special case of the generalised one when $\alpha = 0$. Its asymptotic behaviour (as $n \to \infty$) is similar to the classic particle filters. For $\alpha > 0$, the approximating measure has a density with respect to the Lebesgue measure and this can enable us to study the properties of the density of the posterior measures. However, the Gaussian mixture particle filters reduces the computational efforts by carrying more information on each (generalised) particle; it also enables us to study more properties about ρ_t, especially its density, from its approximation ρ_t^n.

The next step is to study the effect of the 'mollifying' parameter α on the variance of the method, and identify the optimal one. In this way we will keep the computation effort to a minimum.

The use of the method presented here can provide an answer to the issue of *sample degeneracy* (or *sample impoverishment*), see for example [1, 6]. This affects classical particle filters and it is manifested through the fact that, over time the weights of majority of the particles can decrease to zero, with only few becoming very large. As a result the accuracy of the approximation decays over time. Several ad-hoc approaches have been investigated to tackle this problem. An approach called *jittering* was introduced in [6] by adding an additional noise to the samples to reduce the degeneracy. Normally the weights are calculated using the posterior likelihood (see [9]), Thrun et al ([21]) proposed a risk-sensitive particle filters, where they measured the weights by both the likelihood and a specific risk function measuring the risk of not tracking a particular area of the state space. A theoretical study on this method (see [17]) shows that it is a successful alternative approach to mitigate sample impoverishment.

As mentioned in Section 2, other possible tools to help construct the generalised particles include wavelets and finite elements. The key idea of building the approximations is similar to using the Gaussian mixtures. Again we denote the approximation of the ρ_t by ρ_t^n, and aim to make the approximation ρ_t^n satisfy

$$d\rho_t^n(\varphi_t) = \rho_t^n(\frac{\partial\varphi_t}{\partial t} + A\varphi_t)dt + \rho_t^n(\varphi_t h^\top)dY_t + R_t^n(\varphi_t). \tag{21}$$

Comparing (21) with the Zakai equation (3), it can be seen that these two equations are "sufficiently" close to each other provided the remainder term $R_t^n(\varphi_t)$ in (21) is "sufficiently" small, in which case ρ_t^n will converge to ρ_t (see Theorem 4.9 in [16] for the rigorous statement and its proof). This development is part of our future work. Once this is done, we aim to provide a comparative theoretical analysis in order to identify the optimal methods within the class of generalised particle filters for various classes of approximations.

REFERENCES

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

[2] A. Bain and D. Crisan, *Fundamentals of Stochastic Filtering*, Stochastic Modelling and Applied Probability, vol. 60, Springer, 2008.

[3] D. Brigo, B. Hanzon, and F. Le Gland, "A differential geometric approach to nonlinear filtering: the projection filter," *IEEE Trans. Automatic Control*, vol. 43(2), pp. 247-252, Feb. 1998.

[4] Z. Cai, F. Le Gland, and H. Zhang, "An adaptive local grid refinement method for nonlinear filtering," *Technical Report 2679, INRIA*, 1995.

[5] A. Carmi, F. Septier, and S.J. Godsill, "The Gaussian mixture MCMC particle algorithm for dynamic cluster tracking," in *Proc. The 12th International conference on information fusion*, Seattle, WA, USA, July 6-9 2009, pp. 1179-1186.

[6] J Carpenter, P. Clifford, P. Fearnhead, "Improved particle filter for nonlinear problems," *IEE Proceedings on Radar, Sonar and Navigation*, Vol. 146(1), pp. 2-7, Feb.1999.

[7] D. Crisan, and B. Rozovsky, editors, *The Oxford Handbook of Nonlinear Filtering*, Oxford University Press, 2011.

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

[9] A. Doucet, J.F.G. de Freitas, and N.J. Gordon, editors, *Sequential Monte Carlo Methods in Practice*, Springer, 2001.

[10] M. Flament, G. Fleury and M.E. Davoust "Particle filter and Gaussian-mixture filter efficiency evaluation for terrain-aided navigation," in *Proc. EUSIPCO 2004*, Vienna, Austria, September 6-10, 2004, pp. 605-608.

[11] G. Kallianpur, *Stochastic Filtering Theory*, New York, Springer, 1980.

[12] J.H. Kotecha and P.M. Djurić, "Gaussian particle filtering," *IEEE Trans. Signal Processing*, vol. 51, pp. 2593-2602, Oct. 2003.

[13] H.J. Kushner, "Approximations to optimal nonlinear flters," *IEEE Trans. Automatic Control,* vol. 12(5), pp. 546-556, Oct. 1967.

[14] F. Le Gland, V. Monbet and V.D. Tran, "Large sample asymptotics for the ensamble Kalman filter," *Handbook on Nonlinear Filtering*, Oxford University Press, 2010.

[15] S. Lototsky, R. Mikulevicius, and B. Rozovsky, "Nonlinear filtering revisited: A spectral approach," *SIAM J. Control Optim.*, vol. 35(2), pp.435-461, 1997.

[16] O. Obanubi, "Particle filters with random resampling times," *PhD Thesis, Imperial College London*, Dec. 2010.

[17] U. Orguner and F. Gustafsson, "Risk sensitive particle filters for mitigating sample impoverishment," *IEEE Trans. Signal Processing*, vol. 56(10), pp. 5001-5012, Oct. 2008.

[18] S. Reich, "A Gaussian mixture ensemble transform filter," submitted, *eprint arXiv:1102.3089*, Feb. 2011.

[19] C. Rogers and D. Williams, *Diffusions, Markov Processes and Martingales: Volume I Foundations*, second ed., Cambridge, UK, Cambridge University Press, 2000.

[20] J. Szpirglas, "Sur l'équivalence d'équations différentielles stochastiques à valeurs measures intervenant dans le filtrageMarkovien non linéaire" [French], *Ann. Inst. H. Poincaré Sect. B (N.S.)*, vol 14(1), pp. 33-59, 1978.

[21] S. Thrun, J. Langford, and V. Verma, "Risk sensitive particle filters," in *Advances in Neural Information Processing Systems 14*, MIT Press, 2002.

[22] R. Van der Merwe and E. Wan, "Gaussian mixture sigma-point particle filters for sequential probabilistic inference in dynamic state-space models," in *Proc. ICASSP 2003*, pp. 701-704.

[23] M. Zakai, "On the optimal filtering of diffusion process," *Zeit. Wahrsch.*, vol. 11, pp. 230-243, 1969.