Large Deviations for Interacting Particle Systems. Applications to Non Linear Filtering

P. Del Moral[*], A. Guionnet[†]

Abstract

The non linear filtering problem consists in computing the conditional distributions of a Markov signal process given its noisy observations. The dynamical structure of such distributions can be modelled by a measure valued dynamical Markov process. Several random particle approximations were recently suggested to approximate recursively in time the so-called non linear filtering equations. We present an interacting particle system approach and we develop large deviations principles for the empirical measures of the particle systems. We end this paper extending the results to an interacting particle system approach which includes branchings.

Keywords : Large deviations, Interacting random processes, Filtering, Stochastic approximation.
code A.M.S : 60F10, 60H10, 60G35, 93E11, 62L20.

1 Introduction

1.1 Background and motivations

The non linear filtering problem consists in computing the conditional distribution of a signal given its noisy observation. Roughly speaking, a basic model for non linear filtering problems is to assume that the signal is a time inhomogeneous Markov process X with observations Y described by

$$Y_n = h_n(X_n) + V_n \tag{1}$$

[*]UMR C55830, CNRS, Bat.1R1, Univ. Paul Sabatier, 31062 Toulouse, delmoral@cict.fr.
[†]URA 743, CNRS, Bat. 425, Université de Paris Sud, 91405 Orsay, France, guionnet@stats.matups.fr

where h_n are some continuous functions and V_n a noise, which we will assume independent of the signal X_n.

It was proven in a general setting by Kunita [25] and Stettner [31] that the law of X_n given the observations $(Y_p, p \leq n)$ obeys the so-called non linear filtering equations. Basically, if η_n denotes the law of X_n given $(Y_p, p \leq n)$, these equations are of the form

$$\eta_n = \phi(n, \eta_{n-1}) \qquad \forall n \geq 1 \qquad \eta_0 = \eta \tag{2}$$

where η is the law of the initial signal and $\phi(n, .)$ an application on the space of probability measures on the state space of the signal. $\phi(n, .)$ depends on the observations $(Y_p, p \leq n)$, on the laws of $(V_p, p \leq n)$ and on the transition probability kernels of the signal process (see Lemma 3.1 for its complete description).

The study of equations of type (2) is far from being straightforward. Such equations also occur in Statistical Physics and Biology (see [12],[35] and references therein). In these frameworks, the dynamical system (2) usually describes the time evolution of the density profiles of McKean-Vlasov stochastic processes with mean field drift functions. It was proposed by McKean and Vlasov to approximate the corresponding equations by mean field interacting particle systems. A crucial practical advantage of this situation is that the dynamical structure of the non linear stochastic process can be used in the design of an interacting particle system in which the mean field drift is replaced by a natural interaction function. Such models are called in Physics Masters equations and/or weakly interacting particle systems. They are now well understood (see [2], [11], [12],[21], [35], [36] and references therein). Under rather general assumptions, it was shown that the particle density profile (that is the random empirical measures of the particle systems) converges towards the solution of (2) as the number of particles is going to infinity. As a consequence, propagation of chaos occurs. To specify the rate of this convergence, large deviations properties and fluctuations were studied.

Among the most exciting developments in Non Linear Filtering Theory are those centered around the recently established connection with interacting and branching particle systems. In non linear filtering problems the dynamical system (2) describes the time evolution of the conditional distribution of the internal states in dynamical systems when partial observations are made. In contrast to the situation described above the conditional distributions cannot be viewed as the law of a finite dimensional stochastic process which incorporates a mean field drift. We therefore have to find a new strategy to define an interacting particle system which will approximate the desired distributions. In the last decade several different stochastic particle approximations were suggested to approximate the so-called non linear filtering equation. The evolution of this rapidly developing area of research may be seen quite directly through the following chain of papers [6],[8],[9],[10],[13], [15] and [20]. In [16] and [17], general particle systems which include branching and non linear interactions were described. The laws of the empirical measures of these systems were shown to converge weakly to the desired conditional distribution as the number of particles is growing.

Several practical problems which have been solved using these methods are given in [5], [6], [13], [14], signal processing and GPS/INS integration.

In the current work we develop large deviations for some of the particle approximations studied in [16] and [17] in which the interaction function only depends on the empirical measure of the system. Such results provide the exact speed of convergence of the algorithms we consider until a finite given time. The study of their long time behavior is a rather different subject which we will hopefully investigate in another paper.

1.2 Description of the model; Statement of some results

To describe precisely our model, let us introduce some notations. The signal X_n at time n will take its values into a Polish space E. E is endowed with a Borel σ-field $\mathbf{B}(E)$. We denote by $\mathbf{M}_1(E)$ the space of all probability measures on E furnished with the weak topology. We recall that the weak topology is generated by the bounded continuous functions. We will denote $\mathcal{C}_b(E)$ the space of these functions.

The particle system $(\Omega, F_n, (\xi_n)_{n\geq 0}, P)$ under study will be a Markov process with state space E^N, where $N \geq 1$ is the size of the system. The N-tuple of elements of E, i.e. the points of the set E^N, are called particle systems and will be mostly denoted by the letters x, z.
Our dynamical system is then described by

$$P(\xi_0 \in dx) = \prod_{p=1}^{N} \eta_0(dx^p) \qquad P(\xi_n \in dx / \, \xi_{n-1} = z) = \prod_{p=1}^{N} \phi\left(n, \frac{1}{N}\sum_{q=1}^{N} \delta_{z^q}\right)(dx^p) \tag{3}$$

where $dx \stackrel{\text{def}}{=} dx^1 \times \ldots \times dx^N$. Our goal is to prove large deviation principles for the law of the empirical distribution of the N-particle system (3)

$$\eta^N(\xi_{[0,n]}) = \frac{1}{N}\sum_{i=1}^{N} \delta_{(\xi_0^i, \ldots, \xi_n^i)} \tag{4}$$

which is a random measure on the path space $\Sigma_n = E^{n+1}$. Our results will then basically be stated under the following form.

Theorem 1.1 *If the functions $(\phi(n,.))_{n\geq 1}$ are "good enough" then we have that*

$$\eta^N(\xi_{[0,n]}) \xrightarrow[N\to\infty]{} \eta_{[0,n]} \stackrel{\text{def}}{=} \eta_0 \otimes \ldots \otimes \eta_n \qquad a.s. \tag{5}$$

In addition

1. *The law $(Q_N^n)_{N\geq 1}$ of $\eta^N(\xi_{[0,n]})$ satisfies a* **large deviation principle** *(LDP) with rate function I_n which reads*
   (a) *$I_n : \Sigma_n \to [0,\infty]$ has compact level sets, that is $\{I_n \leq M\}$ is a compact subset of Σ_n.*
   (b) *For any closed set $C \subset \mathbf{M}_1(\Sigma_n)$*
   $$\limsup_{N\to\infty} \frac{1}{N} \log Q_N^n(C) \leq -\inf_C I_n. \tag{6}$$
   (c) *For any open set $A \subset \mathbf{M}_1(\Sigma_n)$*
   $$\liminf_{N\to\infty} \frac{1}{N} \log Q_N^n(A) \geq -\inf_A I_n. \tag{7}$$
2. *I_n has a unique minimizer which is $\eta_{[0,n]}$.*

The LDP precise the rate of the convergence (5). Let us give an explicit upper bound for this rate. To do so, let us introduce a metric d on $\mathbf{M}_1(\Sigma_n)$ compatible with the weak topology :

$$d(\mu,\nu) \stackrel{\text{def}}{=} \sum_{m\geq 1} 2^{-(m+1)} \, |\mu f_m - \nu f_m| \tag{8}$$

where $(f_m)_{m\geq 1}$ is a suitable sequence of uniformly continuous functions uniformly bounded by 1 (theorem 6.6 pp. 47, Parthasarathy [29]). We consider the function $\Phi_n : \mathbf{M}_1(\Sigma_n) \to \mathbf{M}_1(\Sigma_n)$ so that

$$\Phi_n(\mu) = \eta_0 \otimes \phi(1, T_0\mu) \otimes \ldots \otimes \phi(n, T_{n-1}\mu) \tag{9}$$

where $T_k\mu$, $0 \leq k \leq n$, is the k-marginal of μ. Clearly,

$$d(\Phi_n(\mu), \mu) = 0 \Longleftrightarrow \mu = \eta_{[0,n]}.$$

We will see that, thanks to the large deviation results proved in section 2, under appropriate assumptions, for any $\epsilon > 0$ there exists $N(\epsilon) \geq 1$ so that

$$\forall N \geq N(\epsilon) \qquad P\left(d(\eta^N(\xi_{[0,n]}), \Phi_n\left(\eta^N(\xi_{[0,n]})\right)) > \epsilon\right) \leq e^{-\frac{N\epsilon^2}{4}}.$$

The crucial point is now to specify the assumptions needed on the $\phi(n,.)$'s for such result to hold. Throughout this paper, we shall weaken these hypotheses as much as we can in order to include as many examples encountered in non linear filtering problems as possible.

The paper has the following structure:

We will derive in section 2 large deviation principles for the particle system (3) keeping in mind to weaken the assumptions on the functions $\phi(n,.)$'s as much as we can in order to apply it to non linear filtering particle systems. In subsection 2.1, we consider the empirical measure on path space. In subsection 2.2, we considerably weaken the hypotheses needed in the latter and get large deviation principles for the time marginals.

The applications of the large deviation results obtained in section 2 to non linear filtering problems will be explored in section 3.

We end this paper with some generalizations of the former large deviation results to some random particle approximations which includes branchings and interactions.

2 LDP for Interacting Particle Systems

Our interest is in large deviation results for the laws of the empirical measures associated to our interacting particle systems (3). The study of the minimizers of the rate functions governing these large deviations will in turn provide convergence of the empirical trajectories towards $\eta_{[0,n]}$ and an exponential speed for this convergence. A general formulation for studying such problems was given by Varadhan in [38] and by Azencott in [1]. More recent developments can be found in Deuschel-Stroock [19] and Dembo-Zeitouni [18].

Our developments will be mainly based on Laplace method, Gärtner-Ellis and Baldi theorem.

2.1 LDP for the Empirical Measures on the Path Space

In this section, we focus on the empirical trajectories $\eta^N(\xi_{[0,n]}) \in \mathbf{M}_1(\Sigma_n)$ defined by (4) until a given finite time $n \in I\!\!N$. With these notations, we can rewrite the law Q_n^N of $\eta^N(\xi_{[0,n]})$ for our particle system (3) as the probability so that, for any $F \in \mathcal{C}_b(\mathbf{M}_1(\Sigma_n))$,

$$Q_n^N F = \int_{\Sigma_n^N} F(\eta^N(x))\ \phi(n, T_{n-1}\eta^N(x))^{\otimes N}(dx_n) \dots \phi(1, T_0\eta^N(x))^{\otimes N}(dx_1)\ \eta_0^{\otimes N}(dx_0)$$

where $\phi(k,\mu)^{\otimes N}$, $1 \le k \le n$, $\mu \in \mathbf{M}_1(E)$, is the N-fold product of the measure $\phi(k,\mu)$.

To prove large deviations for $\{Q_n^N, N \ge 1\}$ we will always assume that

(A) : For any time $n \geq 1$ there exists a probability of reference $\lambda_n \in \mathbf{M}_1(E)$ such that

$$\forall \mu \in \mathbf{M}_1(E) \qquad \phi(n,\mu) \sim \lambda_n.$$

This condition might seem difficult to check in general but in fact covers many typical examples of non linear filtering problems (see section 3). It is obvious that the situation becomes considerably more involved when dispensing with this assumption. As we will see in section 2.2, it turns out that a continuity assumption on the functions $\phi(n,\boldsymbol{.})$ is sufficient to obtain an inductive LDP for the time marginals.

The main simplification due to assumption (A) is that each law Q_n^N is equivalent to the distribution $R_n^N \in \mathbf{M}_1(\mathbf{M}_1(E))$ given by

$$R_n^N F = \int_{\Sigma_n^N} F(\eta^N(x))\ \lambda_0^{\otimes N}(dx_0)\ldots\lambda_n^{\otimes N}(dx_n)$$

for any $F \in \mathcal{C}_b(\mathbf{M}_1(\Sigma_n))$, with the convention $\lambda_0 = \eta_0$. We notice that the latter formula can be written in the form

$$R_n^N F = \int_{\Sigma_n^N} F(\eta^N(x))\ R_n^{\otimes N}(dx) \quad \text{with} \quad R_n = \lambda_0 \otimes \ldots \otimes \lambda_n$$

It is also easily seen that

$$\frac{dQ_n^N}{dR_n^N} = \exp(NF_n) \qquad R_n^N - \text{a.s.} \tag{10}$$

where $F_n : \mathbf{M}_1(\Sigma_n) \to I\!\!R$ is the function defined by

$$F_n(\mu) = \sum_{k=1}^{n} \int_E \log \frac{d\phi(k, T_{k-1}\mu)}{d\lambda_k}\ dT_k\mu = \int_{\Sigma_n} \log \frac{d\Phi_n(\mu)}{dR_n}\ d\mu \tag{11}$$

with the notation of (9). In a first stage for analysis it is reasonable to suppose that

(B) For any time $n \geq 1$ the function $(\mu,\nu) \to \int \log \dfrac{d\phi(n,\nu)}{d\lambda_n}\ d\mu$ is bounded continuous.

If $I(\mu|\nu)$ denotes the relative entropy of μ with respect to ν, that is the function

$$I(\mu|\nu) = \begin{cases} \int \log \frac{d\mu}{d\nu}\ d\mu & \text{if } \mu << \nu \\ +\infty & \text{otherwise,} \end{cases},$$

Sanov's theorem and Varadhan's lemma yields

Theorem 2.1 *Under (A) and (B), $\{Q_n^N,\ N \geq 1\}$ satisfies a LDP with good rate function*

$$I_n(\mu) = I(\mu|\Phi_n(\mu)).$$

$\eta_{[0,n]}$ *is the unique minimizer of* I_n.

Indeed, F_n is bounded continuous under (B) so that $\{Q_n^N,\ N \geq 1\}$ satisfies a LDP with good rate function $I_n = I(.|R_n) - F_n$ according to Sanov's theorem and Varadhan's lemma (see [19] for instance). From the definition of $I(\ |R_n)$ and F_n, it is obvious that I_n is also given by $I_n(\mu) = I(\mu|\Phi_n(\mu))$ from which it is easily seen that

$$I_n(\mu) = 0 \Longleftrightarrow \mu = \eta_{[0,n]}.$$

■

At this point it is appropriate to address a deficiency in the preceding result. An approximatively equivalent condition of (B) is given by the two following assumptions

(C0) For any time $n \geq 1$,

$$(x, \nu) \to \log \frac{d\phi(n,\nu)}{d\lambda_n}(x)$$

is uniformly continuous w.r.t. x and continuous w.r.t. ν.

(C) For any time $n \geq 1$ there exists a non negative real number a_n so that

$$\forall (x,\nu) \in E \times \mathbf{M}_1(E) \qquad a_n^{-1} \leq \frac{d\phi(n,\nu)}{d\lambda_n}(x) \leq a_n$$

A clear disadvantage of condition (C) is that it is in general not satisfied when E is not compact and in particular in many non linear filtering problems. Yet we are going to see that this condition can be relaxed considerably. The relevance of the foregoing results will be illustrated in section 3 when applied to non linear filtering problems.
It is now convenient to introduce some additional notations. For any $M > 0$ we note $\psi^M : I\!R \to I\!R$ the cut-off function given by

$$\psi^M(x) = \begin{cases} x & \text{if } |x| \leq M, \\ \text{sign}(x)\, M & \text{if } |x| > M, \end{cases}$$

and $F_n^M : \mathbf{M}_1(\Sigma_n) \to I\!R$ the function

$$F_n^M(\mu) = \sum_{k=1}^{n} \int_E \psi^M \left(\log \frac{d\phi(k, T_{k-1}\mu)}{d\lambda_k} \right) \, dT_k\mu.$$

Under (C0) , F_n^M is bounded continuous (beware here that this statement requires the uniform continuity (and not only the continuity) property of hypothesis (C0). Next conditions relax assumption (C).
(C1) For any time $n \geq 0$ and $\epsilon > 0$ there exists a function $L_{n;\epsilon}$, $L_{n;\epsilon}(M)$ goes to infinity when M goes to infinity, so that

$$R_n^N \left(e^{-NF_n}\, 1\!\!I_{(|F_n - F_n^M| > \epsilon)} \right) \leq e^{-NL_{n,\epsilon}(M)}.$$

(C2) For any time $n \geq 0$ and for any $A \in \mathbf{B}(\mathbf{M}_1(\Sigma_n))$ we have

$$\lim_{M\to\infty} \inf_A \left(I(\,.\,|R_n) - F_n^M \right) = \inf_A \left(I(\,.\,|R_n) - F_n \right).$$

Let us describe the main result of this section

Theorem 2.2 *Assume that the functions $\phi(n,\,.\,)$, $n \geq 1$, satisfy conditions (A), (C0), (C1) and (C2). Then, for any $n \geq 0$, $\{Q_n^N : N \geq 1\}$ satisfies a LDP with good rate function I_n.*

The proof is based on the ideas of Azencott and Varadhan and amounts to replace the functions F_n (which are a priori nor bounded nor continuous) by the functions F_n^M to get the LDP up to a small error ϵ in the rate function by (C1) and then pass to the limit $M \to \infty$ by (C2) to let finally $\epsilon \downarrow 0$. We leave the details to the reader.

Conditions (C1) and (C2) are hard to work with. In practice we will check the following more elegant conditions

(D1) For any time $n \geq 1$, there exists constants $c_n < \infty$, $\alpha_n > 1$ such that

$$R_n^N \left(e^{\alpha_n N F_n} \right) \leq e^{c_n N} \tag{12}$$

and, for every $\epsilon > 0$ there exists a function $L_{n,\epsilon}$, $L_{n;\epsilon}(M)$ goes to infinity when M goes to infinity, so that

$$R_n^N \left(|F_n - F_n^M| > \epsilon \right) \leq e^{-N L_{n,\epsilon}(M)}. \tag{13}$$

(D2) For any time $n \geq 1$, there exists constants $\delta_n > 0$ $C_n < \infty$, $D_n < \infty$ and a function ϵ_n, $\epsilon_n(M)$ is going to zero when M is going to infinity, such that for any $\mu \in \mathbf{M}_1(\Sigma_n)$ and $M \in I\!R \cup \{\infty\}$

$$\begin{array}{rcl} I(\mu|R_n) - F_n^M(\mu) & \geq & \delta_n \, I(\mu|R_n) - C_n \\ |F_n(\mu) - F_n^M(\mu)| & \leq & \epsilon_n(M) \, (I(\mu|R_n) + D_n) \end{array}$$

Using Hölder's inequality it can be checked directly that $(D1) \Rightarrow (C1)$. On the other hand, for any $A \in \mathbf{B}(\mathbf{M}_1(\Sigma_n))$ and any integer number L we have

$$\inf_A \{I(.|R_n) - F_n^M\} = \inf \left\{ \inf_{A\cap\{I(.|R_n)\leq L\}} \{I(.|R_n) - F_n^M\}; \inf_{A\cap\{I(.|R_n)\geq L\}} \{I(.|R_n) - F_n^M\} \right\}. \tag{14}$$

Let us first assume that $\inf_A I_n < \infty$. Then, the second assumption of (D2) yields

$$\left| \inf_{A\cap\{I(.|R_n)\leq L\}} \{I(.|R_n) - F_n^M\} - \inf_{A\cap\{I(.|R_n)\leq L\}} \{I(.|R_n) - F_n\} \right| \leq \epsilon_n(M)(L + D_n). \tag{15}$$

On the other hand, the first assumption shows that, uniformly in M,

$$\inf_{A\cap\{I(.|R_n)\geq L\}}\{I(.|R_n)-F_n^M\}\geq \delta_n L - C_n.$$

Thus, for L and M large enough, it is clear that (14) implies

$$\inf_A\{I(.|R_n)-F_n^M\}=\inf_{A\cap\{I(.|R_n)\leq L\}}\{I(.|R_n)-F_n^M\},\qquad \inf_A I_n=\inf_{A\cap\{I(.|R_n)\leq L\}} I_n$$

and therefore

$$\left|\inf_A\{I(.|R_n)-F_n^M\}-\inf_A\{I(.|R_n)-F_n\}\right|\leq \epsilon_n(M)(L+D_n).$$

Also, if $\inf_A I_n=+\infty$, $\inf_{A\cap\{I(.|R_n)\leq L\}} I_n=+\infty$ for any integer number L and therefore (15) gives, for any integer number L,

$$\inf_A\{I(.|R_n)-F_n^M\}=\inf_{A\cap\{I(.|R_n)\geq L\}}\{I(.|R_n)-F_n^M\}\geq \delta_n L - C_n.$$

Letting L going to infinity implies

$$\inf_A\{I(.|R_n)-F_n^M\}=+\infty=\inf_A I_n$$

which completes the proof of $(D2)\Rightarrow(C2)$.
We end this section with an example of how the preceding theorem can be applied. This corollary will be one of the key tools used in most of the applications of our results to non linear filtering problems (cf section 3). It is quite remarkable that the weakening of condition (B) is compensated by an exponential moment condition.

Corollary 2.3 *Suppose the functions $\phi(n,.)$, $n\geq 1$, satisfy (A) and (C0) and that for any $1\leq k\leq n$, $x\in E$ and $\mu\in\mathbf{M}_1(E)$*

$$\left|\log\frac{d\phi(k,\mu)}{d\lambda_k}(x)\right|\leq\varphi(x)+\mu(\psi)\tag{16}$$

for some non negative and $\mathbf{B}(E)$-measurable functions φ and ψ. In addition, assume that there exists constants $\alpha,\beta\in]1,\infty]$ and $\epsilon>0$ such that $\frac{1}{\alpha}+\frac{1}{\beta}<1$ and for any $1\leq k\leq n$

$$\int\exp\left(\alpha\varphi^{1+\epsilon}\right)d\lambda_k\vee\int\exp\left(\beta\psi^{1+\epsilon}\right)d\lambda_k<\infty\tag{17}$$

Then, $\{Q_n^N:\ N\geq 1\}$ satisfies the LDP with good rate function I_n.

Proof:
Recalling our discussion preceding the corollary we only have to check that (17) implies

(D1) and (D2). For any $n \geq 1$, choose and fix constants $\alpha, \beta > 1$ and $\epsilon > 0$ so that (17) is satisfied. Define $p, q > 1$ and $\delta > 1$ by

$$\frac{1}{\delta} = \frac{1}{\alpha} + \frac{1}{\beta} \qquad \frac{1}{p} = 1 - \frac{\delta}{\beta} \qquad \frac{1}{q} = 1 - \frac{1}{p}$$

Let us now state some useful bounds which are needed in the sequel. From (16) we find that for any $\mu \in \mathbf{M}_1(\Sigma_n)$ and $n \geq 1$

$$|F_n(\mu)| \vee |F_n^M(\mu)| \leq \sum_{k=0}^{n} T_k\mu(\phi + \psi) \tag{18}$$

and

$$|F_n(\mu) - F_n^M(\mu)| \leq \left(\frac{2}{M}\right)^{\epsilon} \sum_{k=0}^{n} T_k\mu(\phi^{1+\epsilon} + \psi^{1+\epsilon}) \tag{19}$$

The last inequality is a clear consequence of Hölder's inequality and the fact that $(a+b)^{1+\epsilon} \leq 2^{\epsilon}(a^{1+\epsilon} + b^{1+\epsilon})$ for any $a, b \geq 0$ and $|\psi^M(x) - x| \leq |x| \; 1_{|x|>M}$.
Let us now establish the moment condition (12). Using (18) we have

$$\frac{1}{N} \log R_n^N(e^{\delta N F_n}) \leq \sum_{k=0}^{n} \log \int \exp\left(\delta(\varphi + \psi)\right) d\lambda_k$$

using Hölder's inequality this shows that $R_n^N(e^{\delta N F_n}) \leq \exp(NC_n)$ with

$$C_n = \sum_{k=0}^{n} \log \left\{ \int \exp(\alpha\varphi) \, d\lambda_k \vee \int \exp(\beta\psi) \, d\lambda_k \right\}$$

On the other hand we see from (19) that, for any positive δ,

$$\begin{aligned} R_n^N(|F_n - F_n^M| > \delta) \;&\leq\; R_n^N\left(\left\{\mu : \sum_{k=0}^{n} T_k\mu(\phi^{1+\epsilon} + \psi^{1+\epsilon}) > \left(\frac{M}{2}\right)^{\epsilon} \epsilon\right\}\right) \\ &\leq\; \exp(-N\delta 2^{-\epsilon} M^{\epsilon}) \left(\prod_{k=0}^{n} \int \exp(\phi^{1+\epsilon} + \psi^{1+\epsilon}) d\lambda_k\right)^N \end{aligned}$$

Again using Hölder's inequality and recalling that $\alpha > p$ and $\beta > q$ one concludes

$$\frac{1}{N} \log R_n^N(|F_n - F_n^M| > \delta) \leq -L_{n,\delta}(M)$$

with

$$L_{n,\delta}(M) = \delta 2^{-\epsilon} M^{\epsilon} - \sum_{k=0}^{n} \log \left(\int \exp(\alpha\varphi^{1+\epsilon}) \, d\lambda_k \vee \int \exp(\beta\psi^{1+\epsilon}) \, d\lambda_k \right)$$

Now we proceed to the proof of (D2). Taking into consideration the inequality (18) we have for any $\mu \in \mathbf{M}_1(\Sigma_n)$ and $M < \infty$

$$|F_n^M(\mu)| \leq \mu(\theta_n) \qquad \text{with} \qquad \theta_n(x_0, \dots, x_n) = \sum_{k=0}^{n} \varphi(x_k) + \psi(x_k).$$

Using the well known property of the relative entropy

$$I(\mu|\nu) = \sup_{V \in \mathcal{C}_b(\mathcal{X})} \Big(\mu(V) - \nu(e^V)\Big) \tag{20}$$

and the monotone convergence theorem it follows that

$$\delta |F_n^M(\mu)| \leq I(\mu|R_n) + \log \int \exp\left(\delta\theta_n\right) dR_n.$$

Thus, we arrive at

$$I(\mu|R_n) - F_n^M(\mu) \geq \left(1 - \frac{1}{\delta}\right) I(\mu|R_n) - \frac{1}{\delta} \sum_{k=0}^{n} \log \int \exp\left(\delta(\varphi + \psi)\right) d\lambda_k.$$

Now, by a further use of Hölder's inequality one gets

$$I(\mu|R_n) - F_n^M(\mu) \geq \left(1 - \frac{1}{\delta}\right) I(\mu|R_n) - C$$

with

$$C = \sum_{k=0}^{n} \log \left(\int \exp\left(\alpha\varphi\right) d\lambda_k \vee \int \exp\left(\beta\psi\right) d\lambda_k \right).$$

By a method similar to that used above one can also establish that for any $\mu \in \mathbf{M}_1(\Sigma_n)$

$$|F_n(\mu) - F_n^M(\mu)| \leq \left(\frac{2}{M}\right)^{\epsilon} \left(I(\mu|R_n) + D\right)$$

with

$$D = \sum_{k=0}^{n} \log \left(\int \exp\left(\alpha\varphi^{1+\epsilon}\right) d\lambda_k \vee \int \exp\left(\beta\psi^{1+\epsilon}\right) d\lambda_k \right).$$

This ends the proof. ∎

Before closing this section we examine how theorem 2.2 makes it possible to estimate in a simple way the probability of the events $B_{n,\epsilon} := \{\mu : \; d(\mu, \Phi_n(\mu)) < \epsilon\}$, $\epsilon > 0$. Under (C0), $\mu \to \Phi_n(\mu)$ is continuous so that $B_{n,\epsilon}$ is open for the weak topology. Recalling that $d(\mu, \Phi_n(\mu))) = 0$ iff $\mu = \eta_{[0,n]}$, we see that this event is an open neighborhood of $\eta_{[0,n]}$. If we denote $\|\cdot\|_{TV}$ denotes the total variation norm then using the well known inequalities

$$d(\mu,\Phi_n(\mu)) \leq \|\mu-\Phi_n(\mu)\|_{TV} \leq (2\; I(\mu|\Phi_n(\mu)))^{1/2}$$

and the large deviation upper bound one concludes that

$$\limsup_{N\to\infty} \frac{1}{N}\log Q_n^N(\{\mu:\; d(\mu,\Phi_n(\mu))\geq\epsilon\}) \leq -\frac{\epsilon^2}{2}$$

It follows that there exists $N(\epsilon)\geq 1$ such that for any $N\geq N(\epsilon)$

$$Q_n^N\,(\{\mu\;:\; d(\mu,\Phi_n\,(\mu))<\epsilon) \geq 1-\exp-(\frac{N\,\epsilon^2}{4})$$

2.2 LDP for the Particles Density Profiles

The large deviations results presented in section 2.1 rely entirely on the existence of a family of reference distributions $\{\lambda_n:\; n\geq 1\}$ satisfying condition (A) and therefore does not apply to some filtering problems (see section 3). To remove this assumption we shall be dealing with the law P_n^N, $n\geq 0$, $N\geq 1$, of the particle density profiles

$$\eta^N(\xi_n) \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^N \delta_{\xi_n^i}.$$

We also can relax the continuity assumption (C0) into

(CW) : For any time $n\geq 1$, $\phi(n,.)$ is continuous.

To insure an exponential tightness property we shall propose the following assumption which is motivated for its applications in non linear filtering problems.
If, for any Markov transition M and any $\mu\in\mathbf{M}_1(E)$ we denote μM the probability so that for any $f\in\mathcal{C}_b(E)$,

$$\mu M f = \int f(y)M(dy,x)\mu(dx),$$

this hypothesis reads

(ET) : For any $n\geq 1$, $\epsilon>0$ and for any Markov transition M on E, there exists a Markov kernel $\widetilde{M}$ and $0<\delta\leq\epsilon$ such that

$$\mu\widetilde{M}(A^c)<\delta \Longrightarrow \phi(n,\mu)M(A^c)<\epsilon$$

for any $\mu\in\mathbf{M}_1(E)$ and for any compact set $A\subset E$.

Remark 2.4: Observe that condition (ET) holds if, for any $n \geq 1$, there exists a Markov transition K_n on E and a non negative constant $c(n)$ so that

$$\phi(n,\mu)(A^c) \leq c(n)\,\mu K_n(A^c)$$

for any compact set $A \subset E$.

Proposition 2.5 *Assume that condition (ET) holds. Then, for any $L > 0$ and $n \geq 0$ we can find a compact set $K_L \subset \mathbf{M}_1(E)$ so that*

$$P\left(\eta^N(\xi_n) \in K_L^c\right) \leq 4\, e^{-N\,L} \tag{21}$$

Proof:

For every sequence of real numbers $m = (m_l)_{l\geq 1}$ satisfying $\lim_{l\to\infty} m_l = \infty$ and $m_l \geq 1$ and, every sequence of compact subsets $A = (A_l)_{l\geq 1}$ of E we shall denote $C(A,m)$ the compact subset of $\mathbf{M}_1(E)$ given by

$$C(A,m) = \cap_{l\geq 1} G(A_l, m_l) \qquad G(A_l,m_l) = \left\{\nu \in \mathbf{M}_1(E)\ :\ \nu(A_l^c) \leq \frac{1}{m_l}\right\} \qquad \forall l \geq 1$$

We shall need a modification of Azencott and Stroock lemma (see for instance lemma 6.13 pp. 125 [34]).

Lemma 2.6 *For any $L < \infty$ and for any sequence of real numbers $m = (m_l)_{l\geq 1}$ satisfying $lim_{l\to\infty} m_l = \infty$ and $m_l \geq 1$, $l \geq 1$, there exists a sequence of real numbers $\widetilde{m} = (\widetilde{m}_l)_{l\geq 1}$ so that $\widetilde{m}_l \geq m_l$ for any $l \geq 1$, and*

$$\mu M \in C(A,\widetilde{m}) \Longrightarrow \mu^{\otimes N}\left(\frac{1}{N}\sum_{i=1}^N \delta_{x^i} M \in C(A,m)^c\right) \leq 2\ \exp\left(-NL\right)$$

for any $\mu \in \mathbf{M}_1(E)$, any Markov transition M on E, any sequence of compact sets $A = (A_l)_{l\geq 1}$ and any $N \geq 1$.

Using this lemma we see that for any

- $N \geq 1$,
- Markov kernel M on E,
- sequence of compact sets $A = (A_l)_{l\geq 1}$,
- sequence of real numbers $m(0) = (m_l(0))_{l\geq 1}$ satisfying

$$\lim_{l\to\infty} m_l(0) = \infty \qquad \text{and} \qquad m_l(0) \geq 1 \qquad \forall l \geq 1$$

we have

$$\begin{aligned}
& P\left(\eta^N(\xi_n)M \in C(A,m(0))^c\right) \\
\leq\ & P\left(\{\eta^N(\xi_n)M \in C(A,m(0))^c\} \cap \{\phi(n,\eta^N(\xi_{n-1}))M \in C(A,\overline{m}(0))\}\right) \\
& +P\left(\phi(n,\eta^N(\xi_{n-1}))M \in C(A,\overline{m}(0))^c\right) \\
\leq\ & 2\,e^{-N\,(n+1)\,L} + P\left(\phi(n,\eta^N(\xi_{n-1}))M \in C(A,\overline{m}(0))^c\right)
\end{aligned}$$

for any sequence of real numbers $\overline{m}(0) = (\overline{m}_l(0))_{l\geq 1}$ chosen according to lemma 2.6. Hence, using (ET) there exists a sequence of real numbers $m(1) = (m_l(1))_{l\geq 1}$ satisfying $m_l(1) \geq \overline{m}_l(0)$ for any $l \geq 1$ and a Markov transition $M_!$ so that

$$P\left(\eta^N(\xi_n)M \in C(A,m(0))^c\right) \leq 2\,e^{-N\,(n+1)\,L} + P\left(\eta^N(\xi_{n-1})M_1 \in C(A,m(1))^c\right) \quad (22)$$

Using repeatedly (22) one concludes

$$P\left(\eta^N(\xi_n) \in C(A,m(0))^c\right) \leq 2\sum_{k=0}^{n-1} e^{-NL(n+1-k)} + P\left(\eta^N(\xi_0)M_n \in C(A,m(n))^c\right) \quad (23)$$

for some sequence of real numbers $m(n) = (m_l(n))_{l\geq 1}$ and a Markov transition M_n. Since E is separable and complete $\eta_0 M_1 \ldots M_n$ is tight. Then, for any sequence or real numbers $\widetilde{m} = (\widetilde{m}_l)_{l\geq 1}$ satisfying $lim_{l\to\infty}\widetilde{m}_l = \infty$ one can choose the sequence of compact sets $A = (A_l)_{l\geq 1}$ so that

$$\eta_0 M_n \in C(A,\widetilde{m})$$

Using lemma 2.6 it follows that

$$P\left(\eta^N(\xi_0)M_n \in C(A,m(n))^c\right) \leq 2\,e^{-NL}$$

which, plugged in (23), gives a compact set $K_L = C(A,m(0)) \subset \mathbf{M}_1(E)$ such that

$$P\left(\eta^N(\xi_n) \in C(A,m(0))^c\right) \leq 4\,e^{-N\,L}.$$

■

We are now in position to state the main result of this section

Theorem 2.7 *Assume that conditions (CW) and (ET) hold. Then, for any $n \geq 1$, $\{P_n^N : N \geq 1\}$ obeys a large deviation principle with good rate function H_n given by*

$$\left\{\begin{array}{lcll} H_n(\mu) & = & \inf_{\nu}\{I\left(\mu|\phi(n,\nu)\right) + H_{n-1}(\nu)\} & \quad n \geq 1 \\ H_0(\mu) & = & I(\mu|\eta_0) & \end{array}\right.$$

In addition $H_n(\mu) = 0$ iff $\mu = \eta_n$, for any $n \geq 1$.

Proof:
Hereafter we shall denote for $n \geq 1$, $\nu_1, \cdots, \nu_n \in \mathbf{M}_1(\mathbf{E})^n$, $\bar{\nu}_n = \nu_n \otimes \cdots \otimes \nu_1$ and $\Psi(n, \bar{\nu}_{n-1}) = \phi(n, \nu_{n-1}) \otimes \cdots \otimes \phi(1, \nu_1) \otimes \eta_0 \in \mathbf{M}_1(\mathbf{E}^n)$.

Let us first show that H_n is a good rate function. Indeed, it is clearly non negative. Moreover, by induction one finds that

$$H_n(\mu) = \inf_{\nu_{n-1}, \cdots \nu_1 \in \mathbf{M}_1(\mathbf{E})^{n-1}} \{I(\mu \otimes \bar{\nu}_{n-1} | \Psi(n, \bar{\nu}_{n-1}))\}.$$

But, $(\mu, \bar{\nu}_{n-1}) \to I(\mu \otimes \bar{\nu}_{n-1} | \Psi(n, \bar{\nu}_{n-1}))$ is a lower semi-continuous function since it can be obtained as a supremum of continuous functions by the formula

$$I(\mu \otimes \bar{\nu}_{n-1} | \Phi(n, \bar{\nu}_{n-1})) = \sup_{V \in C_b(\mathbf{E}^n)} \{\int V d\mu \otimes \bar{\nu}_{n-1} - \log \int e^V d\Psi(n, \bar{\nu}_{n-1})\}$$

where we have observed that the last term in the above supremum is continuous by assumption (CW). Hence, I is lower semi-continuous and since H_n is obtained by contraction over the last marginal, so is H_n. The fact that H_n has compact level sets with be a consequence of the exponential tightness assumed here and the proof of the weak large deviation principle below (see Dembo-Zeitouni, Lemma 1.2.18). Finally, it is clear that H_n admits η_n as a unique minimizer since $I(\mu \otimes \bar{\nu}_{n-1} | \Psi(n, \bar{\nu}_{n-1}))$ vanishes only at $\eta_n \otimes \cdots \otimes \eta_0$.

Let us now turn to the proof of the large deviation upper bound. We shall obtain in fact a LDUB for the process $(\eta^N(\xi_n), \cdots, \eta^N(\xi_0)) \in \mathbf{M}_1(\mathbf{E})^{n+1}$ with rate function $I_{n+1}(\mu) = I(\mu | \Psi(n+1, \bar{\mu}_n))$. To this end we proceed by induction on the parameter $n \geq 0$. Consider first the case $n = 0$. The particle system $\xi_0 = (\xi_0^1, \ldots, \xi_0^N)$ consists of N i.i.d. variables with common law η_0. Thus Sanov's theorem tells us that the family $\{P_0^N : N \geq 1\}$ obeys a LDUB with rate function $H_0(\mu) = I(\mu | \eta_0)$. Assume that $\{P_{n-1}^N : N \geq 1\}$ obeys a LDUB with rate function I_{n-1} for some $n \geq 1$.

To prove the result at time n, observe first that Proposition 2.5 insures that the law of $(\eta^N(\xi_n), \cdots, \eta^N(\xi_0))$ is exponentially tight so that it is sufficient to prove the weak large deviation upper bound.

Applying Chebyshev's inequality, we conclude that for any $\mu_1, \cdots, \mu_{n-1} \in \mathbf{M}_1(E)^n$, any $V \in \mathbf{C}_b(\mathbf{E})$, any $\delta > 0$,

$$\begin{aligned} & P\left(1_{d(\eta^N(\xi_j), \mu_j) \leq \delta, 1 \leq j \leq n}\right) \\ = \;& P\left(1_{d(\eta^N(\xi_j), \mu_j) \leq \delta, 1 \leq j \leq n-1} \phi(n, \eta^N(\xi_{n-1}))^{\otimes N}\left(e^{N \int V d\eta^N(\xi_n) - N \int V d\eta^N(\xi_n)}\right)\right) \\ \leq \;& e^{-N \int V d\mu_n + N\epsilon(\delta)} P\left(1_{d(\eta^N(\xi_j), \mu_j) \leq \delta, 1 \leq j \leq n-1} e^{N \log \phi(n, \eta^N(\xi_{n-1})(e^V)}\right) \end{aligned} \tag{24}$$

where $\epsilon(\delta)$ is a function going to zero with δ.

Now, for any bounded continuous function V on E, any $\mu_0, \cdots, \mu_{n-1} \in \mathbf{M}_1(\mathbf{E})$,

$$\begin{aligned}
&\limsup_{\delta\to 0}\limsup_{N\to\infty}\frac{1}{N}\log P\left(1_{d(\eta^N(\xi_j),\mu_j)\le\delta, 1\le j\le n-1} e^{N\int V d\eta^N(\xi_n)}\right)\\
=\;& \limsup_{\delta\to 0}\limsup_{N\to\infty}\frac{1}{N}\log P\left(1_{d(\eta^N(\xi_j),\mu_j)\le\delta, 1\le j\le n-1} e^{N\log\phi(n,\eta^N(\xi_{n-1}))(e^V)}\right)\\
\le\;& -\liminf_{\delta\to 0}\inf\{I_{n-1}(\bar\nu_{n-1}) - \log\phi(n,\nu_{n-1})(e^V), d(\nu_j,\mu_j)\le\delta, 1\le j\le n-1\}\\
=\;& -I_{n-1}(\bar\mu_{n-1}) - \log\phi(n,\mu_{n-1})(e^V) \qquad (25)
\end{aligned}$$

where we have used in the last line our induction hypothesis and the fact that $\mu \to \log\phi(n,\mu)(e^V)$ is continuous to apply Laplace's method. Taking the limit N going to infinity and then δ to zero, we find, thanks to our induction hypothesis, (24) and (25), that

$$\limsup_{\delta\to 0}\limsup_{N\to\infty}\frac{1}{N}\log P\left(1_{d(\eta^N(\xi_j),\mu_j)\le\delta, 1\le j\le n}\right) \le -\int V d\mu_n + \log\phi(n,\mu_{n-1})(e^V) - I_{n-1}(\bar\mu_{n-1})$$

Taking the supremum over V and recalling that

$$I(\mu_n|\phi(n,\mu_{n-1})) = \sup_{V\in C_b(E)}\{\int V d\mu_n - \log\phi(n,\mu_{n-1})(e^V)\}$$

gives the weak large deviation upper bound with rate function
$I_{n-1}(\bar\mu_{n-1}) + I(\mu_n|\phi(n,\mu_{n-1})) = I_n(\bar\mu_n)$.

We finally obtain the large deviation upper bound for the law of $\eta^N(\xi_n)$ for $n \in \mathbb{N}$ by the contraction principle.

Let us now turn to the lower bound. Again, we proceed by induction. At time $n = 0$, Sanov's theorem shows that the law of $\eta^N(\xi_0)$ satisfies a large deviation lower bound with rate function $H_0(\mu) = I(\mu|\eta_0)$. Let us assume that the result is true at time $n-1$, for some $n \ge 1$. Let $\mu \in \mathbf{M}_1(\mathbf{E})$ be given. We need to prove that for any $\delta > 0$,

$$\liminf_{N\to\infty}\frac{1}{N}\log P\left(d(\eta^N(\xi_n),\mu) < \delta\right) \quad\ge\quad -H_n(\mu).$$

We can of course restrict ourselves to μ such that $H_n(\mu) < \infty$. Let $\epsilon > 0$ be given and $\nu \in \mathbf{M}_1(\mathbf{E})$ such that

$$H_n(\mu) \ge I(\mu|\phi(n,\nu)) + H_{n-1}(\nu) - \epsilon.$$

A fortiori, since $H_n(\mu) < \infty$ and $H_{n-1}(\nu) \geq 0$, $I(\mu|\phi(n,\nu)) \leq H_n(\mu) < \infty$ so that μ is absolutely continuous with respect to $\phi(n,\nu)$, $d\mu = e^V d\phi(n,\nu)$ with a measurable function V such that $\int e^V d\phi(n,\nu) = 1$.

Let us assume that V is bounded continuous. Then, for any $\delta' > 0$,

$$\begin{aligned}
P\left(d(\eta^N(\xi_n),\mu) < \delta\right) &\geq P\left(\{d(\eta^N(\xi_n),\mu) < \delta\} \cap \{d(\eta^N(\xi_{n-1}),\nu) < \delta'\}\right) \\
&= P\left(1_{\{d(\eta^N(\xi_{n-1}),\nu)<\delta'\}}\phi(n,\eta^N(\xi_{n-1}))^{\otimes N}\left(d(\eta^N(\xi_n),\mu) < \delta\right)\right) \\
&\geq e^{-N\int V d\mu + N\log\int e^V d\phi(n,\nu) - N\eta} \\
&\quad \times P\left(1_{\{d(\eta^N(\xi_{n-1}),\nu)<\delta'\}}\phi_V(n,\eta^N(\xi_{n-1}))^{\otimes N}\left(d(\eta^N(\xi_n),\mu) < \delta\right)\right)
\end{aligned} \tag{26}$$

where the last line is true for any $\eta > 0$ provided δ, δ' are small enough (since we assumed V bounded continuous) and we have used the notation

$$d\phi_V(n,\nu)(\xi) = \frac{e^{V(\xi)} d\phi(n,\nu)(\xi)}{\int e^V d\phi(n,\nu)}.$$

Now, for any given ν, the law of large numbers asserts that $\eta^N(\xi_n)$ converges towards $\phi_V(n,\nu)$ almost surely under $\phi_V(n,\nu)^{\otimes N}$. Since $\nu \to \phi(n,\nu)$ is continuous, $\phi_V(n,\nu')$ goes to $\phi_V(n,\nu) = \mu$ as ν' goes to ν. Hence, for any ν' such that $\{d(\nu',\nu) \leq \delta'\}$ with δ' small enough,

$$\lim_{N\to\infty} \phi_V(n,\nu')^{\otimes N}\left(d(\eta^N(\xi_n),\mu) < \delta\right) = 1.$$

We deduce by bounded convergence theorem that for δ' small enough,

$$\lim_{N\to\infty} \frac{P\left(1_{\{d(\eta^N(\xi_{n-1}),\nu)\leq\delta'\}}\phi(n,\eta^N(\xi_{n-1}))^{\otimes N}\left(d(\eta^N(\xi_n),\mu) \leq \delta\right)\right)}{P\left(d(\eta^N(\xi_{n-1}),\nu) < \delta'\right)} = 1$$

Now, by our induction hypothesis,

$$\liminf_{N\to\infty} \frac{1}{N}\log P\left(d(\eta^N(\xi_{n-1}),\nu) < \delta'\right) \geq -H_{n-1}(\nu)$$

so that we obtain finally

$$\begin{aligned}
\liminf_{N\to\infty} \frac{1}{N}\log P\left(d(\eta^N(\xi_{n-1}),\nu) \leq \delta'\right) &\geq -H_{n-1}(\nu) - N\int V d\mu + N\log\int e^V d\phi(n,\nu) \\
&= -H_{n-1}(\nu) - I(\mu|\phi(n,\nu)) \geq -H_n(\mu) - \epsilon.
\end{aligned}$$

To deal with the general case, we remark that in exercise 6.2.20 of Dembo-Zeitouni, it is shown that the continuous bounded away from zero and infinity densities are dense in the level sets of the relative entropy

$$E_L = \{f \in L^1(\phi(n,\nu)), f \geq 0 \quad \phi(n,\nu) a.s., I(f.d\mu|\phi(n,\nu)) = \int f\log f d\phi(n,\nu) \leq L\}$$

for any $L \in I\!R^+$. Let $\mu_p = e^{V_p} d\phi(n,\nu)$, with bounded continuous functions V_p, be a sequence in $E_{H_n(\mu)-H_{n-1}(\nu)+\epsilon}$ which converges to μ. For any $\delta > 0$, for p large enough,

$$\{\alpha \in \mathbf{M}_1(\mathbf{E}) : d(\alpha,\mu_p) < \frac{\delta}{2}\} \subset \{\alpha \in \mathbf{M}_1(\mathbf{E}) : d(\alpha,\mu) < \delta\}$$

so that the previous result shows that

$$\begin{aligned}
& \liminf_{N\to\infty} \frac{1}{N} \log P\left(\{d(\eta^N(\xi_{n-1}),\nu) \le \delta'\} \cap \{d(\eta^N(\xi_n),\mu) \le \delta\}\right) \\
\ge\ & \liminf_{N\to\infty} \frac{1}{N} \log P\left(\{d(\eta^N(\xi_{n-1}),\nu) \le \delta'\} \cap \{d(\eta^N(\xi_n),\mu_p) \le \frac{\delta}{2}\}\right) \\
\ge\ & -H_{n-1}(\nu) - I(\mu_p|\phi(n,\nu)) \\
\ge\ & -H_n(\mu) - \epsilon
\end{aligned}$$

where we have used in the last line the fact that $\mu_p \in E_{H_n(\mu)-H_{n-1}(\nu)+\epsilon}$. We can now let ϵ going to zero to obtain the large deviation lower bound with rate H_n for the law of $\eta^N(\xi_n)$.

∎

3 Applications to Non Linear Filtering

3.1 Introduction

The basic model for the general Non Linear Filtering problem consists of a time inhomogeneous Markov process X and a non linear observation Y with observation noise V. Namely, let (X,Y) be the Markov process taking values in $E \times I\!R^d$, $d \ge 1$, and defined by the system:

$$\mathcal{F}(X/Y) \qquad \begin{cases} X = (X_n)_{n\ge 0} \\ Y_n = h_n(X_n) + V_n \qquad n \ge 0 \end{cases} \tag{27}$$

where E is a locally compact and separable metric space, $h_n : E \to I\!R^d$, $d \ge 1$, are bounded continuous functions and V_n are independent random variables with continuous and positive density g_n with respect to Lebesgue measure. The signal process X that we consider is assumed to be a non-inhomogeneous and E-valued Markov process with Feller transition probability kernel K_n, $n \ge 1$, and initial probability measure ν, on E. We will assume the observation noise V and X are independent.

The classical filtering problem is concerned with estimating the distribution of X_n conditionally to the observations up to time n. Namely,

$$\pi_n(f) \stackrel{\text{def}}{=} E(f(X_n)/Y_0, \ldots, Y_n) \tag{28}$$

for all $f \in \mathcal{C}_b(E)$. This problem has been extensively studied in the literature and, with the notable exception of the linear-Gaussian situation or wider classes of models (Bènes filters [3]) optimal filters have no finitely recursive solution (Chaleyat-Maurel/Michel [7]). Although Kalman filtering ([23],[26]) is a popular tool in handling estimation problems its optimality heavily depends on linearity. When used for non linear filtering (Extended Kalman Filter) its performance relies on and is limited by the linearizations performed on the concerned model. The interacting particle systems approach developed hereafter can be seen as a non linear filtering method which discards linearizations. More precisely these techniques use the non linear system model itself in order to solve the filtering problem. The problem of assessing the distributions (28) is of course related to that of recursively computing the conditional distributions π_n, $n \geq 0$, which provides all statistical informations about the states variables X_n obtainable from the observations $(Y_0, \ldots, Y_n)$, $n \geq 0$. For a detailed discussion of the filtering problem the reader is referred to the pioneering paper of Stratonovich [33] and to the more rigorous studies of Shiryaev [32] and Kallianpur-Striebel [24]. More recent developments can be found in Ocone [27] and Pardoux [28]. Some collateral readings such as Kunita [25], Stettner [31], Michel [7] will be helpful in appreciating the relevance of our approximations.

3.2 Formulation of the Non Linear Filtering Problem

Let us introduce the filtering model in such a way that the techniques of section 2 can be applied. To this end it is convenient to study the distribution of the state process X_n conditionally on the observation up to time $(n-1)$. Namely,

$$\eta_n f \stackrel{\text{def}}{=} E\left(f(X_n)/Y_0, \ldots, Y_{n-1}\right) \qquad \forall n \geq 0 \quad \forall f \in \mathcal{C}_b(E)$$

with the convention $\eta_0 = \nu$. The following result shows that the dynamics structure of the conditional distributions η_n, $n \geq 0$, can be viewed as a special case of (2).

Lemma 3.1 (Kunita [25],Stettner [31])
Given a fixed observation record Y=y, $(\eta_n)_{n\geq 0}$ is solution of the $\mathbf{M}_1(E)$-valued dynamical system

$$\eta_{n+1} = \phi_{n+1}(y_n, \eta_n) \qquad n \geq 0 \qquad \eta_0 = \nu \tag{29}$$

where $y_n \in \mathbb{R}^d$ is the current observation and ϕ_n is the continuous function given by

$$\phi_{n+1}(y_n, \eta) = \psi_n(y_n, \eta) K_{n+1} \tag{30}$$

with

$$\psi_n(y_n,\eta)f \stackrel{\text{def}}{=} \frac{\displaystyle\int f(x)\,\overline{g}_n(y_n,x)\,\eta(dx)}{\displaystyle\int \overline{g}_n(y_n,z)\,\eta(dz)} \qquad \forall f\in\mathcal{C}_b(E),\ \eta\in\mathbf{M}_1(E)$$

The equation (29) is usually called the non linear filtering equation. Even if it looks innocent, it can rarely be solved analytically and its solving requires extensive calculations. To obtain a computionnally feasible solution some kind of approximation is needed.

We observe that the recursion (30) involves two separate mechanisms. Namely, the first one

$$\mu(dx)\mapsto \frac{\overline{g}_n(Y_n,x)}{\displaystyle\int \overline{g}_n(Y_n,z)\,\mu(dz)}\mu(dx)$$

updates the distribution given the current observation. The second one

$$\mu\mapsto\mu K_n$$

does not depends on the current observation. It is usually called the prediction.

3.3 Interacting Particle Systems Approximations

Recalling the description (3), and using the fact that

$$\phi_{n+1}\left(Y_n,\frac{1}{N}\sum_{i=1}^N\delta_{z^i}\right)=\left(\sum_{i=1}^N\frac{\overline{g}_n(Y_n,z^i)}{\sum_{j=1}^N\ \overline{g}_n(Y_n,z^i)}\ \delta_{z^i}\right)K_{n+1} \tag{31}$$

the N-particle system associated to (29) is defined by

$$P_Y(\xi_{n+1}\in dx/\,\xi_n=z)=\prod_{j=1}^N\sum_{i=1}^N\frac{\overline{g}_n(Y_n,z^i)}{\sum_{j=1}^N\ \overline{g}_n(Y_n,z^i)}\ K_{n+1}(z^i,dx^j)$$

Thus, we see that the particles move according the following rules

1. **Updating:** When the observation $Y_n=y_n$ is received, each particle examines the system of particles $\xi_n=(\xi_n^1,\ldots,\xi_n^N)$ and chooses randomly a site ξ_n^i with probability
$$\frac{\overline{g}_n(y_n,\xi_n^i)}{\sum_{j=1}^N\overline{g}_n(y_n,\xi_n^j)}$$
2. **Prediction:** After the updating mechanism each particle evolves according the transition probability kernel of the signal process.

This particle approximation of the non linear filtering equation belongs to the class of algorithms called genetic algorithms. These algorithms are based on the genetic mechanisms which guide natural evolution: exploration/mutation and updating/selection. They were introduced by J.H. Holland [22] to handle global optimization problems on a finite set.

3.4 Large Deviation for Interacting Particle Systems

In [17] we proposed some exponential bounds to prove that for a fixed observation record Y=y, for every $f \in \mathcal{C}_b(E)$ and for every $n \geq 0$, $\eta^N(\xi_n)f$ converges $P-a.s.$ to $\eta_n f$ as the size N of the systems is growing.
Our aim is now to show how the LDP developed in section 2 can be applied to obtain the exact exponential rate of convergence of our random particle approximation.
We will use the following assumptions

(H0) : For any time $n \geq 1$, K_n is Feller so that $\mu \to \mu K_n$ is continuous for the weak topology. For any time $n \geq 0$, h_n is bounded continuous and g_n is a positive continuous function.

(H1) : For any time $1 \leq k \leq n$ there exists a reference probability measure $\lambda_k \in \mathbf{M}_1(E)$, $\alpha > 1$, $\epsilon > 0$ and a $\mathbf{B}(E)$-measurable function φ so that

$$\delta_x K_k \ \sim \ \lambda_k$$

In addition $z \to \log \frac{d\delta_x K_k}{d\lambda_k}(z)$ is Lipschitz, uniformly on the parameter x such that

$$\left|\log \frac{d\delta_x K_k}{d\lambda_k}(z)\right| \leq \varphi(z) \quad \text{and} \quad \int \exp\left(\alpha\varphi^{1+\epsilon}\right) d\lambda_k < \infty \tag{32}$$

Let the condition (H0) be satisfied. Then, it is not hard to see that there exists positive functions $a_n : I\!R^d \to I\!R_+$ such that

$$a_n(y)^{-1} \leq \overline{g}_n(y,x) \leq a_n(y) \qquad \forall (y,x) \in I\!R^d \times E \tag{33}$$

We conclude that for any observation record $Y = y$, $(\phi_{n+1}(y_n,.), n \geq 0)$ satisfies (C). Moreover, for any $\mathbf{B}(E)$-measurable function $f : E \to I\!R_+$ and for every $n \geq 0$, $y \in I\!R^d$ and $\mu \in \mathbf{M}_1(E)$ we have

$$a_n(y)^{-2}\, \mu K_{n+1} f \leq \phi_{n+1}(y,\mu) f \leq a_n(y)^2\, \mu K_{n+1} f$$

Therefore one concludes easily that under (H0) theorem 2.7 applies without further work. More precisely we have proved the following proposition

Proposition 3.2 *Assume that condition (H0) holds. Then, for any observation record $Y = y$ and $n \geq 0$, the laws $\{P_n^N \ : \ N \geq 1\}$, of the particle density profiles*

$$\eta^N(\xi_n) = \frac{1}{N}\sum_{i=1}^N \delta_{\xi_n^i} \qquad N \geq 1$$

satisfy a LDP with rate function H_n given by

$$\begin{cases} H_n(\mu) &= \sup_{V\in\mathcal{C}_b(E)} \left(\mu(V) + \inf_{\nu\in\mathbf{M}_1(E)} \left(H_{n-1}(\nu) - \log\left(\psi_{n-1}(y_{n-1},\nu)K_n e^V\right)\right)\right) \qquad n \geq 1 \\ H_0(\mu) &= I(\mu|\nu) \end{cases}$$

$H_n(\mu) = 0$ iff $\mu = \eta_n$, for any $n \geq 0$.

Under condition (H1) we now study LDP for the law Q_n^N, $n \geq 0$, $N \geq 1$, of the empirical distribution on path space

$$\eta^N(\xi_{[1,n]}) = \frac{1}{N}\sum_{i=1}^N \delta_{(\xi_1^i,\ldots,\xi_n^i)}$$

We see from (32) and (H0) that for any $1 \leq k \leq n$, $y \in I\!R^d$, $\mu \in \mathbf{M}_1(E)$ and $z \in E$

$$\left|\log \frac{d\phi_k(y,\mu)}{d\lambda_k}(z)\right| \leq \varphi(z)$$

¿From this and corollary 2.3 we have

Proposition 3.3 *Let (H0) and (H1) be satisfied. Then, for any observation record $Y = y$, $\{Q_n^N \ : \ N \geq 1\}$ obeys a LDP with rate function $I_n(\mu) = I(\mu|\Psi_n(\mu))$ where*

$$\Psi_n(\mu) = \nu \otimes \psi_0(y_0, T_0\mu)K_1 \otimes \ldots \otimes \psi_{n-1}(y_{n-1}, T_{n-1}\mu)K_n$$

and $T_k\mu$, $0 \leq k \leq n$ is the k-th marginal of a given probability measure μ.

Remark 3.4: To see the strength of the preceding propositions, let us first quote that (H0) only depends on the functions g_n and h_n ; the LDP results stated in proposition 3.2 does not depend on the form of the Feller signal process X.
In contrast to the latter, the condition (H1) depends on the transitions Markov kernels K_n, $n \geq 1$. As the preceding proposition shows, so long as (H1) holds, the law of the empirical measure on path space satisfies a LDP. Moreover, referring to the remarks preceding section 3, the rate functions H_n, $n \geq 0$, are smaller than the corresponding contractions of the rate functions I_n, $n \geq 0$.

We now turn to some applications of these propositions.

Example 1 *As a typical example of non linear filtering problem assume the functions $h_n : E \to \mathbb{R}^d$, $n \geq 1$, are bounded continuous and the densities g_n given by*

$$g_n(v) = \frac{1}{((2\pi)^d |R_n|)^{1/2}} \ \exp\left(-\frac{1}{2} v' \ R_n^{-1} \ v\right)$$

where R_n is a $d \times d$ symmetric positive matrix. This correspond to the situation where the observations are given by

$$Y_n = h_n(X_n) + V_n \qquad \forall n \geq 1 \tag{34}$$

where $(V_n)_{n\geq 1}$ is a sequence of $\mathbb{R}^d$-valued and independent random variables with Gaussian densities.
After some easy manipulations one gets the bounds (33) with

$$\log a_n(y) = \frac{1}{2} \ \|R_n^{-1}\| \ \|h_n\|^2 + \|R_n^{-1}\| \ \|h_n\| \ \|y\|$$

where $\|h_n\| = \sup_{x\in E} |h_n(x)|$ and $\|R_n^{-1}\|$ is the spectral radius of R_n^{-1}.

Let us now investigate assumption (H1) through the following example

Example 2 *Suppose that $E = \mathbb{R}^m$, $m \geq 1$ and K_n, $n \geq 1$ are given by*

$$K_n(x, dz) = \frac{1}{((2\pi)^m |Q_n|)^{1/2}} \ \exp\left(-\frac{1}{2}(z - b_n(x))' \ Q_n^{-1} \ (z - b_n(x))\right)$$

where Q is a $m \times m$ symmetric non negative matrix and $b_n : \mathbb{R}^m \to \mathbb{R}^m$ is a bounded continuous function. This correspond to the situation where the signal process is given by

$$X_n = b_n(X_n) + W_n \qquad \forall n \geq 1 \tag{35}$$

where $(W_n)_{n\geq 1}$ is a sequence of $\mathbb{R}^m$-valued and independent random variables with Gaussian densities.
It is not difficult to check that (H1) is satisfied with

$$\lambda_n(dz) = \frac{1}{((2\pi)^m |Q_n|)^{1/2}} \ \exp\left(-\frac{1}{2} z' \ Q_n^{-1} \ z\right) dz.$$

Indeed, we then find out that

$$\log \frac{d\delta_x K_n}{d\lambda_n} = const. - b_n(x)' Q_n^{-1} z$$

which insures the Lipschitz property as well as the growth property with

$$\varphi(z) = \frac{1}{2} \|b_n\|^2 \|Q_n^{-1}\| + \|Q_n^{-1}\| \ \|b_n\| \ |z| \qquad \forall z \in \mathbb{R}.$$

Thus, the Gaussian example satisfies (H1). Let us notice that it does not satisfies condition (B) (or even (C)). We discuss this hypothesis below.

Example 3 *Let us suppose that $E = I\!R$ and*

$$K_n(x, dz) = \frac{1}{\sqrt{2\pi}} \, e^{-\frac{1}{2}(z-b_n(x))^2} \, dz \tag{36}$$

where $b_n : I\!R \to I\!R$ is a bounded $\mathbf{B}(E)$-measurable function such that $b_n(0) = 0$ and $b_n(1) = -1$. Then, hypothesis (C) is not satisfied.
Suppose K_n satisfies (C) for some bounded function φ. Clearly there exists an absolutely continuous probability measure with density p_n such that

$$\forall x, z \in I\!R \qquad c_n^{-1} \, p_n(z) \le e^{-\frac{1}{2}(z-b_n(x))^2} \le c_n \, p_n(z)$$

for some positive constant c_n. Using the fact that $b_n(1) = -1$ we obtain

$$\lim_{z\to\infty} p_n(z) e^{\frac{z^2}{2}} = 0$$

On the other hand $b_n(0) = 0$ implies $p_n(z)e^{\frac{z^2}{2}} \ge c_n^{-1}$ which is absurd.

In fact, the failure of hypothesis (C) is linked in general with the non compactness of E.
We have already pointed out that proposition 3.3 is a refinement of proposition 3.2. To be more precise the exponential moment condition (H1) allows us to prove a LDP for the empirical measure on the path space. Now, it is natural to examine some examples where the condition (H1) is not met but still hypothesis (H0) is fulfilled.

Example 4 *Let us suppose that $E = I\!R$ and*

$$K_n(x, dz) = \sqrt{\frac{\epsilon_n(x)}{2\pi}} \; \exp\left(-\frac{1}{2}\epsilon_n(x) \, z^2\right) dz$$

where $\epsilon_n : I\!R \to I\!R$ is a continuous function such that

$$\forall x \in I\!R \qquad \epsilon_n(x) > 0 \qquad \text{and} \qquad \lim_{|x|\to\infty} \epsilon_n(x) = 0.$$

It is not difficult to see that K_n satisfies (H0). On the other hand, let us assume that K_n satisfies (H1) for some function φ. Since $\delta_x K_n$ is absolutely continuous with respect to Lebesgue measure for any $x \in E$, the probability measure λ_n described in (H1) is absolutely continuous with respect to Lebesgue measure. Therefore, there exists a probability density p_n such that

$$\forall x, z \in I\!R \qquad e^{-\varphi(z)} \, p_n(z) \le \sqrt{\epsilon_n(x)} \; \exp\left(-\frac{1}{2}\epsilon_n(x) \, z^2\right) \le e^{\varphi(z)} \, p_n(z).$$

Letting $|x| \to \infty$ one gets $e^{-\varphi(z)} p_n(z) = 0$ for any $z \in I\!R$ which is absurd since we also assumed $\int e^{\varphi^{1+\epsilon}(z)} p_n(z) dz < \infty$.

Our interacting particle system approach is not restricted to non linear filtering problem with Gaussian transitions K_n or with observations corrupted by Gaussian perturbations. As a result we have a great freedom in the design and the physical construction of the non linear filtering model.
It can be argued that in practice it is commonly assumed that the signal X and its noisy observation Y are given by (35) and (34). Under such assumptions the synthesis of the optimal filter is carried out recursively by the well known Kalman-Bucy filter. More precisely, the conditional distributions are Gaussian and the structure of the optimal filter is determined by a recursion relation on the conditional means and on the matrix of errors of observations.
In several practical problems the conditional distribution may have several different modes and the conditional expectation is not meaningful. On the other hand the observed signal process X has no reason to be "linear and Gaussian" and we have to find a more realistic model.
We now present some examples of other kind of densities that can be handled in our framework.

Example 5 *Suppose that $d = 1$ and g_n is a Cauchy density*

$$g_n(v) = \frac{\theta_n}{\pi\,(v^2 + \theta_n^2)} \qquad \theta_n > 0$$

In this situation the weight functions $\overline{g}_n$ is given by

$$\overline{g}_n(y,x) = \frac{y^2 + \theta_n^2}{(y - h_n(x))^2 + \theta_n^2} \qquad \forall (y,x) \in I\!R \times E$$

Notice that

$$\frac{y^2 + \theta_n^2}{y^2 + \theta_n^2 + \|h_n\|^2 + 2|y|\,\|h_n\|} \leq \overline{g}_n(y,x) \leq 1 + \left(\frac{y}{\theta_n}\right)^2$$

It follows that (33) holds with

$$a_n(y) = 1 + \left(\left(\frac{y}{\theta_n}\right)^2 \vee \frac{(|y| + \|h_n\|)^2}{y^2 + \theta_n^2} \right)$$

Example 6 *Suppose $d = 1$ and g_n is a bilateral exponential density*

$$g_n(v) = \frac{1}{2}\,\alpha_n \exp\left(-\alpha_n |v|\right) \qquad \alpha_n > 0$$

In this case the weight functions $\overline{g}_n$ is given by

$$\overline{g}_n(y,x) = \exp\left(\alpha_n\left(|y| - |y - h_n(x)|\right)\right)$$

Observe that

$$-\|h_n\| \leq |y| - |y - h_n(x)| \leq \|h_n\| \qquad \forall (y,x) \in I\!R \times E$$

One concludes that (33) is satisfied with $a_n(y) = \exp\left(\alpha_n \|h_n\|\right)$.

Example 7 *Suppose $E = \mathbb{R}$ and K_n, $n \geq 1$, are given by*

$$K_n(x, dz) = \frac{1}{2}\,\alpha \exp\left(-\alpha|z - b(x)|\right)dz \qquad \alpha > 0 \quad b \in \mathcal{C}_b(\mathbb{R})$$

This corresponds to the situation where the signal process X is given by

$$X_n = b(X_{n-1}) + W_n \qquad n \geq 1$$

where $(W_n)_{n\geq 1}$ is a sequence of $\mathbb{R}^m$-valued and independent random variables with bilateral exponential densities. Note that K_n may be written

$$K_n(x, dz) = \frac{1}{2}\,\alpha \exp\left(\alpha(|z| - |z - b(x)|)\right)\ \lambda_n(dz) \quad \textit{with} \quad \lambda_n(dz) = \frac{1}{2}\,\alpha \exp\left(-\alpha|z|\right)\,dz$$

It follows that (H1) holds since $|\log \frac{\delta_x K_n}{d\lambda_n}(z)|$ has Lipschitz norm $2\alpha + \alpha\|b\|$.

3.5 Large Deviations for Interacting and Branching Particle Systems

The interacting particle system approach described previously is the crudest of the random particle methods introduced in [17]. The goal of this section is to study large deviation of the particle density profiles associated to a branching refinement method. As we shall see this algorithm is a clear extension of the genetic type algorithm described in section 3.3.
The description of the branching particle particle system under study first appears in [17]. It differs from the branching particle algorithms introduced by Crisan and Lyons in [8], [9] and [10]. Intuitively speaking, our branching approach consists in introducing at each mutations a fixed number of auxiliary branching particles but at the end of the selection mechanism most of them are killed. Several numerical investigations have revealed that a clear benefit can be obtained by introducing auxiliary branching particles. In [17] we proved that the corresponding particle density profiles weakly converge to the desired conditional distribution as the size of the system is growing but we let open the question whether or not much loss of performance is incurred by one of these algorithms.
The main purpose of this section is to study the LDP associated to such approximations and to compare its rate function with the rate function H_n which governs the LDP associated to the interacting particle approach described in 3.3.
From such constructions we will show that its rate function is greater than the rate function H_n.

Let us describe our new process. Let $Y = y$ be a given sequence of observation records. The idea is to replace the Mutation/Prediction transition by a branching mechanism.

Namely, the system of particle is now described by the following Markov model

$$\begin{aligned} P_{[y]}(\zeta_0 \in dz) &= \prod_{p_1=1}^{N_1} \prod_{p_2=1}^{N_2} \nu(dz^{p_1,p_2}) \\ P_{[y]}(\widehat{\zeta}_n \in dx / \zeta_n = z) &= \prod_{p_1=1}^{N_1} \sum_{i_1,i_2=1}^{N_1,N_2} \frac{\overline{g}_n(y_n, z^{i_1,i_2})}{\sum_{j_1,j_2=1}^{N_1,N_2} \overline{g}_n(y_n, z^{j_1,j_2})} \, \delta_{z^{i_1,i_2}}(dx^{p_1}) \\ P_{[y]}(\zeta_{n+1} \in dz / \widehat{\zeta}_n = x) &= \prod_{p_1=1}^{N_1} \prod_{p_2=1}^{N_2} K_{n+1}(x^{p_1}, dz^{p_1,p_2}) \end{aligned}$$

where dx (resp. dz) is an infinitesimal neighborhood of $x = (x^1, \ldots, x^{N_1})$ and $z = (z^{i_1,i_2})_{1 \le i_1 \le N_1, 1 \le i_2 \le N_2}$. Note that

$$\phi_{n+1}\left(y_n, \eta^{N_1,N_2}(\zeta_n)\right) = \left(\sum_{i_1,i_2=1}^{N_1,N_2} \frac{\overline{g}_n(y_n, \zeta_n^{i_1,i_2})}{\sum_{j_1,j_2=1}^{N_1,N_2} \overline{g}_n(y_n, \zeta_n^{j_1,j_2})} \, \delta_{\zeta_n^{i_1,i_2}} \right) K_{n+1}$$

and

$$P_{[y]}(\widehat{\zeta}_n \in dx / \zeta_n = z) = \prod_{p_1=1}^{N_1} \sum_{i_1=1}^{N_1} \frac{\sum_{k_2=1}^{N_2} \overline{g}_n(y_n, z^{i_1,k_2})}{\sum_{j_1,j_2=1}^{N_1,N_2} \overline{g}_n(y_n, z^{j_1,j_2})} \sum_{i_2=1}^{N_2} \frac{\overline{g}_n(y_n, z^{i_1,i_2})}{\sum_{k_2=1}^{N_2} \overline{g}_n(y_n, z^{i_1,k_2})} \delta_{z^{i_1,i_2}}(dx^{p_1}) \tag{37}$$

The evolution in time of the particle systems is now described as follows :

1. At the time $n = 0$:
   The particle system $\zeta_0 = \left(\zeta_0^{i_1,i_2} \; ; \; 1 \le i_1 \le N_2, 1 \le i_2 \le N_2\right)$ consists of $N_1 N_2$ i.i.d. variables with the same distribution ν.
2. At the time $n \ge 1$:
   At the time n, the particle system $\zeta_n = \left(\zeta_n^{i_1,i_2} \; ; \; 1 \le i_1 \le N_1, 1 \le i_2 \le N_2\right)$ consists of $N_1 N_2$ particles.
   (a) When the observation $Y_n = y_n$ is received, each particle $\widehat{\zeta}_n^{i_1}$, $1 \le i_1 \le N_1$ chooses a sub-system of auxiliary particles $(\zeta_n^{j_1,1}, \ldots, \zeta_n^{j_1,N_2})$, $1 \le j_1 \le N_1$, at random with probability

$$\frac{\sum_{k=1}^{N_2} \overline{g}_n(y_n, \zeta_n^{j_1,k})}{\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \overline{g}_n(y_n, \zeta_n^{i,j})}$$

   and moves randomly to the site $\zeta_n^{j_1,j_2}$, $1 \le j_2 \le N_2$, in the chosen sub-system with probability

$$\frac{\overline{g}_n(y_n, \zeta_n^{j_1,j_2})}{\sum_{k=1}^{N_2} \overline{g}_n(y_n, \zeta_n^{j_1,k})}$$

   Therefore, the particle system $\widehat{\zeta}_n$ consists of N_1 particles.

(b) In the branching mechanism each particle $\widehat{\zeta}_n^{i_1}$, $1 \leq i_1 \leq N_1$, branches independently into a fixed number N_2 of auxiliary i.i.d.particles with common law $K_{n+1}(\widehat{\zeta}_n^{i_1}, \boldsymbol{.})$. Namely,

$$\widehat{\zeta}_n^{i_1} \in E \longrightarrow (\zeta_{n+1}^{i_1,1}, \ldots, \zeta_{n+1}^{i_1,N_2}) \in E^{N_2} \qquad \text{i.i.d.} \sim K_{n+1}(\widehat{\zeta}_n^{i_1}, \boldsymbol{.})$$

Therefore, at the time $(n+1)$ the particle system ζ_{n+1} consists of $N_1 N_2$ particles.

Remark 3.5: In view of the preceding description we see that the selection/updating mechanism (a) is decomposed into two separate transitions. In the first one each particle $\widehat{\zeta}_n^{i_1}$, $1 \leq i_1 \leq N_1$, chooses one of the N_1 sub-systems

$$\left\{ \left(\zeta_n^{j_1,k}\right)_{1 \leq k \leq N_2} \; : \; 1 \leq j_1 \leq N_1 \right\}$$

in accordance with the observation and the position of the N_2 auxiliary particles. When the sub-system $\left(\zeta_n^{j_1,k}\right)_{1 \leq k \leq N_2}$, $1 \leq j_1 \leq N_1$ is chosen the particle $\widehat{\zeta}_n^{i_1}$ moves to a given site in this sub-system according to the distribution

$$\sum_{k=1}^{N_2} \frac{\overline{g}_n(y_n, \zeta_n^{j_1,k})}{\sum_{l=1}^{N_2} \overline{g}_n(y_n, \zeta_n^{j_1,l})} \, \delta_{\zeta_n^{j_1,k}}$$

We begin by noting that the latter transition can be written

$$\frac{\overline{g}_n(y_n, x)}{\displaystyle\int \overline{g}_n(y_n, z) \, \frac{1}{N_2} \sum_{l=1}^{N_2} \delta_{\zeta_n^{j_1,l}}(dz)} \; \frac{1}{N_2} \sum_{k=1}^{N_2} \delta_{\zeta_n^{j_1,k}}(dx) \tag{38}$$

By definition of the sub-system $\left(\zeta_n^{j_1,k}\right)_{1 \leq k \leq N_2}$, with $n \geq 1$, the empirical measure

$$\frac{1}{N_2} \sum_{k=1}^{N_2} \delta_{\zeta_n^{j_1,k}}(dx)$$

is a particle approximation of the probability measure $K_n(\widehat{\zeta}_{n-1}^{j_1}, dx)$. Thus, intuitively speaking, (38) approximates

$$\frac{\overline{g}_n(y_n, x)}{\displaystyle\int \overline{g}_n(y_n, z) \, K_n(\widehat{\zeta}_{n-1}^{j_1}, dz)} \; K_n(\widehat{\zeta}_{n-1}^{j_1}, dx)$$

which is the conditional distribution of X_n with respect to $Y_n = y_n$ and $X_{n-1} = \widehat{\zeta}_{n-1}^{j_1}$. In other words the second transition in the mechanism (a) can be viewed as a mutation for each particle in accordance with the observation. In this situation we see that the particles track the unknown process by using mutations depending on the observation records.

Our next objective is to study large deviations upper bound for the laws $P_n^{N_1,N_2}$ of the particle density profiles

$$\eta^{N_1,N_2}(\zeta_n) = \frac{1}{N_1 N_2} \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \delta_{\zeta_n^{i_1,i_2}}$$

The arguments are similar to those used to prove theorem 2.7. We will also work with the condition (ET) page 12. If this condition takes place then, using the same line of arguments as those used in the proof of proposition 2.5, one has the following result

Proposition 3.6 *Assume that condition (H0) holds. Then for any $L > 0$, $N_2 \geq 1$ and $n \geq 0$ we can find a compact set $K_L \subset \mathbf{M}_1(E)$ so that*

$$P\left(\eta^{N_1,N_2}(\zeta_n) \in K_L^c\right) \leq 4\, e^{-N_1\, L}. \tag{39}$$

We state now the main result of this section

Theorem 3.7 *Assume that (H0) and (H1) hold. Then, for any $n \geq 0$ and $N_2 \geq 1$, the family $\{P_n^{N_1,N_2} \ : \ N_1 \geq 1\}$ obeys a LDP with rate function $H_n^{N_2}$ given by*

$$\begin{cases} H_n^{N_2}(\mu) = \sup_{V\in\mathcal{C}_b(E)} \left(\mu(V) + \inf_{\eta\in\mathbf{M}_1(E)} \left(H_{n-1}^{N_2}(\eta) - \log \psi_{n-1}(y_{n-1},\eta)((K_n e^{\frac{V}{N_2}})^{N_2})\right)\right) \\ H_0^{N_2}(\mu) = N_2 I(\mu|\nu). \end{cases}$$

Moreover, $H_n \leq H_n^{N_2} \leq N_2\, H_n$ for any $n \geq 0$, where H_n is the rate function introduced in proposition 3.2.

Proof:
By definition of ζ_0, for any $N_2 \geq 1$ $\{P_0^{N_1,N_2} \ : \ N_1 \geq 1\}$ obeys a LDP with rate function

$$H_1^{N_2}(\mu) = \sup_{V\in\mathcal{C}_b(E)} \left(\mu(V) - N_2 \log \nu\left(e^{\frac{V}{N_2}}\right)\right) = N_2 I(\mu|\nu)$$

Let us examine how to obtain a LDP at time n from a LDP at time $(n-1)$.
Assume that $\{P_{n-1}^{N_1,N_2} \ : \ N_1 \geq 1\}$ obeys a LDP with rate function $H_{n-1}^{N_2}(\mu)$ for some $n \geq 1$.
By definition of the particle system ζ_n, one gets for every $V \in \mathcal{C}_b(E)$

$$\frac{1}{N_1} \log E_y\left(\exp\left(N_1 \eta^{N_1,N_2}(\zeta_n)(V)\right)\right) \quad \leq \quad \frac{1}{N_1} \log E_y\left(\exp\left(N_1 F_n^{N_2}(\eta^{N_1,N_2}(\zeta_{n-1}))\right)\right)$$

with

$$F_n^{N_2}(\eta) = \log \int \left(\int e^{\frac{V(z)}{N_2}}\, K_n(x,dz)\right)^{N_2}\, \psi_{n-1}(y_{n-1},\eta)(dx)$$

Under our assumptions we have $F_n^{N_2} \in \mathcal{C}_b(\mathbf{M}_1(E))$. Thus, Varadhan's lemma and the induction hypothesis at rank $(n-1)$ imply that

$$\lim_{N_1 \to \infty} \frac{1}{N_1} \log E_y \left(\exp \left(N_1 \eta^{N_1,N_2}(\zeta_n)(V) \right) \right) \leq \Lambda_n^{N_2}(V)$$

with

$$\Lambda_n^{N_2}(V) \stackrel{\text{def}}{=} - \inf_{\eta \in \mathbf{M}_1(E)} \left(H_{n-1}^{N_2}(\eta) - F_n^{N_2}(\eta) \right)$$

Thus, according to Dembo-Zeitouni [18], for any $N_2 \geq 1$ the family of probability measures $\{P_n^{N_1,N_2} \,:\, N_1 \geq 1\}$ obeys a LDP with rate function $H_n^{N_2}$ with value at μ

$$\inf_{\eta \in \mathbf{M}_1(E)} \left(H_{n-1}^{N_2}(\eta) + \sup_{V \in \mathcal{C}_b(E)} \left(\mu(V) - \log \int \left(\int e^{\frac{V(z)}{N_2}} \, K_n(x,dz) \right)^{N_2} \psi_{n-1}(y_{n-1},\eta)(dx) \right) \right)$$

A clear induction gives the last assertion. This completes the proof of the theorem. ∎

Acknowledgments :

We gratefully acknowledge the European Community for the research fellowships CEC Contract No ERB-FMRX-CT96-0075 and INTAS-RFBR 95-0091. This work was also supported by EPSRC grant GR/ L354661.
We heartly thank T. Lyons, D. Crisan and B. Zegarlinskii for their warm hospitality during our stay in the Imperial College of London where we started this work. We also wish to thank G. Ben Arous for encouraging and fruitful discussions.

References

[1] AZENCOTT R. (1980) Grandres déviations et applications, in "Ecoles d'Eté de Probabilités de Saint Flour VIII-1978", Edited by P.L. Hennequin. Lecture Notes in Mathematics 774, Springer, Berlin, pp. 1-176.

[2] BEN AROUS, G., BRUNAUD, M. (1990). Methode de Laplace : Etude variationnelle des fluctuations de diffusions de type "champ moyen". *Stochastics* **31-32**, 79-144

[3] BENES B.E. (1981). Exact Finite-Dimensional Filters for Certain Diffusions with Nonlinear Drift. *Stochastics,* **Vol.5***, pp 65-92.*

[4] BILLINGSLEY P. (1968). Convergence of probability measures. *Wiley, New York*

[5] CARVALHO H. (1995), Filtrage optimal non linéaire du signal GPS NAVSTAR en racalage de centrales de navigation, *Thèse de L'Ecole Nationale Supérieure de l'Aéronautique et de l'Espace, Sept..*

[6] CARVALHO H., DEL MORAL P., MONIN A., SALUT G. (1995). Optimal nonlinear filtering in GPS/INS Integration. *IEEE Trans. on Aerospace and Electronic Systems.*

[7] CHALEYAT-MAUREL M., MICHEL D. (1983), Des résultats de non existence de filtres de dimension finie. *C.R.Acad.Sci. Paris, Serie I,* **t. 296***.*

[8] CRISAN D., LYONS T.J. (1997). Nonlinear Filtering And Measure Valued Processes. *Proba. Theory and Related Fields, 109, 217-244.*

[9] CRISAN D., LYONS T.J. (1996). Convergence of a Branching Particle Method to the Solution of the Zakai Equation. *Imperial College London, preprint.*

[10] CRISAN D., J. GAINES, LYONS T.J. (1998). A Particle Approximation of the Solution of the Kushner-Stratonovitch Equation. To appear in SIAM Journal of Applied Math. 1998.

[11] DAWSON D.A. .Stochastic Evolution Equation and Related Measure Processes. *Journal of Multivariate Analysis,* **5***, 1-52.*

[12] DAWSON, D.A., GARTNER, J. (1987) Large déviations from the Mc Kean-Vlasov limit for weakly interacting diffusions. *Stochastics*, **20**, 247-308

[13] DEL MORAL P., RIGAL G., NOYER J.C., SALUT G. (1993), Traitement nonlinéaire du signal par reseau particulaire : Application Radar. *quatorzième colloque GRETSI, Juan les Pins, 13-16 Septembre.*

[14] DEL MORAL P., NOYER J.C., SALUT G. (1995). Résolution particulaire et traitement non-linéaire du signal : Application Radar/Sonar. *Traitement du signal, Septembre.*

[15] DEL MORAL P. (1996). Non linear filtering using random particles. *Theor. Prob. Appl.,* **Vol. 40***, (4).*

[16] DEL MORAL P. (1996). Non-linear Filtering: Interacting particle resolution. *Markov Processes and Related Fields,* **Vol.2***, 555-580.*

[17] DEL MORAL P. (1996). Measure Valued Processes and Interacting Particle Systems. Application to Non Linear Filtering Problems.*Publications du Laboratoire de Statistiques et Probabilités, Université Paul Sabatier,* **No 15-96**.. To appear in Annals of Applied Proba. (1998).

[18] DEMBO, A., ZEITOUNI, O (1993) Large deviations techniques and Applications. Jones and Bartlett.

[19] DEUSCHEL, J-D., STROOCK, D.W.(1989) Larges deviations. *Academic Press.*

[20] GORDON N.J., SALMON D.J., SMITH A.F.M. (1993). Novel Approach to non-linear/non-Gaussian Bayesian state estimation.,*IEE*

[21] HITSUDA, M., TANAKA, H.(1981) Central limit theorem for a simple diffusion model for interacting particles. *Hiroshima Mathematical Journal,***11**, 415-423

[22] HOLLAND J.H. (1975) Adaptation in Natural and Artificial Systems.*University of Michigan Press, Ann Arbor*

[23] JAZWINSKI A.H. (1970). Stochastic Processes and Filtering Theory. *Academic Press.*

[24] KALLIANPUR G., STRIEBEL C. (1967). Stochastic differential equations occuring in the estimation of continuous parameter stochastic processes. *Tech. Rep. No 103, Department of statistics, Univ. of Minnesota.*

[25] KUNITA H. (1971). Asymptotic Behavior of Nonlinear Filtering Errors of Markov Processes. *Journal of Multivariate Analysis,* **Vol. 1***, No 4, p. 365-393*

[26] LIANG D., McMILLAN J. (1989). Development of a marine integrated navigation system. *In Kalman Filter Integration of Modern Guidance and Navigation Systems, AGARD-LS-166, OTAN.*

[27] OCONE D.L. (1980). Topics in Nonlinear Filtering Theory. *Ph.D. Thesis, MIT, Cambridge.*

[28] PARDOUX E. (1991). Filtrage Non Linéaire et Equations aux Dérivés Partielles Stochastiques Associées. *Ecole d'été de Probabilités de Saint-Flour XIX-1989, Lecture Notes in Mathematics, 1464, Springer-Verlag*

[29] PARTHASARATHY K.R. (1968).
Probability measures on Metric Spaces.*Academic Press Inc., New York.*

[30] PINSKER M.S. (1964). Information and Information Stability of Random Variables and Processes. *Holden Day, San Francisco.*

[31] STETTNER L. (1988). On Invariant Measures of Filtering Processes, *in K. Helmes, N. Kohlman (Eds.), Stochastic Differential Systems*, Proc. 4th Bad Honnef Conf. 1988, Lecture Notes in Control and Information Sciences, 279-292.

[32] SHIRYAEV A.N. (1966). On stochastic equations in the theory of conditional Markov processes. *Theor. Prob. Appl.* **(11)** *(179-184)*

[33] STRATONOVICH R.L. (1960). Conditional Markov processes. *Theor. Prob. Appl.* **(5)**, *(156-178).*

[34] STROOCK D.W. (1984). An Introduction to the Theory of Large Deviations. *Universitext, Springer-Verlag New-York Inc.*

[35] SZNITMAN A.S. (1984). Non linear reflecting diffusion process and the propagation of chaos and fluctuations associated. *J. Functional Anal.* **56** *(3) 311-336.*

[36] SZNITMAN A.S. (1989). Topics in Propagation of Chaos. *Ecole d'été de Probabilités de Saint-Flour XIX-1989, Lecture Notes in Mathematics,* **No 1464.**

[37] TANAKA, H. (1984) Limit theorems for certain diffusion processes. Proceedings of the Taniguchi Symp, Katata 1982, 469-488, Tokyo, Kinokuniya.

[38] VARADHAN S.R.S. (1966), Asymptotic probabilities and differential equations, Comm. Pure Applied Mathematics, 19, 261-286.