Genealogies and Increasing Propagation of Chaos for Feynman-Kac and Genetic Models

P. Del Moral L. Miclo

LSP-CNRS and Université Toulouse III

abstract

A path-valued interacting particle systems model for the genealogical structure of genetic algorithms is presented. We connect the historical process and the distribution of the whole ancestral tree with a class of Feynman-Kac formulae on path space. We also prove increasing and uniform versions of propagation of chaos for appropriate particle block size and time horizon yielding what seems to be the first result of this type for this class of particle systems.

Keywords

Interacting particle systems, genetic algorithms, historical process, genealogy, relative entropy, propagation of chaos, non linear filtering, Feynman-Kac formula, empirical processes.

AMS Codes (91) : 60F17, 60K35, 60J05, 60G35, 93E11.

1 Introduction

Over the last two decades them have been important developments centering around the connections between genetic algorithms and Feynman-Kac formulae. This subject has natural links to biology, evolutionary computing, physics and advanced signal processing. The reader who wishes to know more details about these connections and specific applications is recommended to consult the survey paper [8] and references therein. In the previously referenced paper we essentially discussed the asymptotic behavior of the empirical measures associated to genetic-type particle systems as the number of particles tends to infinity. The strong versions of propagation of chaos presented here provide several measures of centrality and asymptotic independence for the distribution of a block of particles up to a given time horizon. These asymptotic results complement and strengthen those presented in [8].

Another side topic of the present work concerns the modeling and the convergence analysis of the historical process in population genetics. Aside from inherent and mathematical interest one of the practical reasons for studying the genealogical structure of a genetic algorithm stems from the fact that this set up is precisely what we need to solve numerically the so-called non linear filtering and smoothing problem.

This opening section is decomposed into three parts. we begin in section 1.1 with the Feynman-Kac formulae and provide a brief description of the corresponding genetic-type interacting particle system approximating model. In section 1.2 we describe in some details the main results of the paper. In section 1.3 we close with some comments on related works on the subject and some open problems.

Here are some standard notations to be used in all the paper. Let $\mathcal{P}(E)$ and $\mathcal{B}_b(E)$ denote respectively the set of probability measures and bounded measurable functions on a given measurable space $(E,\mathcal{E})$. As usual $\mathcal{B}_b(E)$ is regarded as a Banach space with the supremum norm

$$\forall f\in\mathcal{B}_b(E)\,,\qquad \|f\|=\sup_{x\in E}|f(x)|$$

The relative entropy $\mathrm{Ent}(\mu_1|\mu_2)$ and the total variation distance $\|\mu_1-\mu_2\|_{\mathrm{tv}}$ between probability measures μ_1, $\mu_2\in\mathcal{P}(E)$ are defined by

$$\mathrm{Ent}(\mu_1|\mu_2)=\int\ \log\frac{d\mu_1}{d\mu_2}\ d\mu_1$$

if $\mu_1<<\mu_2$ and ∞ otherwise, and

$$\|\mu_1-\mu_2\|_{\mathrm{tv}}=\sup\{|\mu_1(f)-\mu_2(f)|\ :\ f\in\mathcal{B}_b(E)\,,\ \ \|f\|\leq 1\}$$

For a distribution $\mu\in\mathcal{P}(E)$ and $\alpha\geq 1$ we also write $\|\,\boldsymbol{.}\,\|_{\alpha,\mu}$ the $\mathbb{L}_\alpha(\mu)$-norm

$$\forall f\in\mathbb{L}_\alpha(\mu)\,,\qquad \|f\|_{\alpha,\mu}=\left(\int\ |f|^\alpha\ d\mu\right)^{\frac{1}{\alpha}}$$

We also recall that any Markov transition $K(x_1,dx_2)$ from a measurable space $(E_1,\mathcal{E}_1)$ into another measurable space $(E_2,\mathcal{E}_2)$ generates two operators. One acting on bounded $\mathcal{E}_2$-measurable functions $f\in\mathcal{B}_b(E_2)$ and taking values in $\mathcal{B}_b(E_1)$

$$\forall(x_1,f)\in(E_1\times\mathcal{B}_b(E_2))\ ,\qquad (Kf)(x_1)=\int_{E_2}\ K(x_1,dx_2)\ f(x_2)$$

and the other one acting on measures $\mu_1\in\mathcal{P}(E_1)$ and taking values in $\mathcal{P}(E_2)$

$$\forall(\mu_1,A)\in(\mathcal{P}(E_1)\times\mathcal{E}_2)\ ,\qquad (\mu_1K)(A)=\int_{E_1}\ \mu_1(dx_1)\ K(x_1,A)$$

Finally, if $L(x_2,dx_3)$ is a Markov transition from $(E_2,\mathcal{E}_2)$ into another measurable space $(E_3,\mathcal{E}_3)$ then we denote $(KL)(x_1,dx_3)$ the composite operator

$$(KL)(x_1,dx_3)=\int_{E_2}\ K(x_1,dx_2)\ L(x_2,dx_3)$$

1.1 Feynman-Kac formulae and Genetic Algorithms

Throughout the sequel $\{(E_n, \mathcal{E}_n) \; ; \; n \in \mathbb{N}\}$ denotes a collection of measurable spaces. We further assume that $X = \{X_n \; ; \; n \in \mathbb{N}\}$ is a time inhomogeneous Markov chain such that at the n-th instant of time the state X_n takes values in E_n and we denote by $K_n(x_{n-1}, dx_n)$ the Markov transition of X at time $n \in \mathbb{N}$. Let there also be given a collection of bounded $\mathcal{E}_n$-measurable and strictly positive functions g_n on E_n, $n \in \mathbb{N}$.

We have shown in earlier papers that genetic algorithms arise naturally as particle approximating models of distributions given by

$$\forall n \in \mathbb{N}\,,\ \forall f \in \mathcal{B}_b(E_n)\,, \qquad \eta_n(f) = \frac{\gamma_n(f)}{\gamma_n(1)}$$

where $\gamma_n(f)$ is a Feynman-Kac formula given by

$$\gamma_n(f) = \mathbb{E}(f(X_n) \prod_{p=0}^{n-1} g_p(X_p)) \tag{1}$$

(see for instance [8] and references therein). To be more precise we recall that the distributions $\{\eta_n \; ; \; n \in \mathbb{N}\}$ are solution of a measure-valued dynamical system

$$\forall n \in \mathbb{N}\,, \qquad \eta_{n+1} = \Phi_{n+1}(\eta_n) \tag{2}$$

where the initial condition $\eta_0 \in \mathcal{P}(E_0)$ is the distribution of X_0. For each $n \in \mathbb{N}$ the one step mapping

$$\Phi_{n+1} : \mathcal{P}(E_n) \longrightarrow \mathcal{P}(E_{n+1})$$

associates to any $\eta \in \mathcal{P}(E_n)$ a probability measure $\Phi_{n+1}(\eta) \in \mathcal{P}(E_{n+1})$ given by

$$\forall f \in \mathcal{B}_b(E_{n+1})\,, \qquad \Phi_{n+1}(\eta)(f) = \frac{\eta(\; g_n \; (K_{n+1}f) \;)}{\eta(\; g_n \;)}, \tag{3}$$

We recall that the N-interacting particle systems approximating model associated to an abstract measure valued dynamical process (2) is the Markov chain $\{\xi_n \; ; \; n \in \mathbb{N}\}$ taking values at each time $n \in \mathbb{N}$ in the product state spaces $\{E_n^N \; ; \; n \in \mathbb{N}\}$ and defined by

$$\mathbb{P}(\xi_{n+1} \in dx_{n+1} \mid \xi_n = x_n) = \prod_{p=1}^{N} \Phi_{n+1}\left(m(x_n)\right)(dx_{n+1}^p) \tag{4}$$

where for each $n \geq 0$, $m(x_n) = \frac{1}{N}\sum_{i=1}^N \delta_{x_n^i}$ is the empirical measure associated to $x_n = (x_n^1, \ldots, x_n^N) \in E_n^N$ and $dx_n = dx_n^1 \times \cdots \times dx_n^N$ is an infinitesimal neighborhood of x_n. The initial system of particles $\xi_0 = (\xi_0^1, \ldots, \xi_0^N)$ consists in N independent random variables with common law η_0. We refer to [4] for a study of this abstract and general N-interacting particle systems model. In our framework and in view of (3) we have that

$$\Phi_{n+1}\left(m(x_n)\right) = \sum_{i=1}^{N} \frac{g_n(x_n^i)}{\sum_{j=1}^N g_n(x_n^j)} \; K_{n+1}(x_n^i, .)$$

Thus, we see that the transition $\xi_n \to \xi_{n+1}$ of the former Markov model splits up into two separate mechanisms

$$\xi_n = (\xi_n^1, \ldots, \xi_n^N) \in E_n^N \xrightarrow{\text{Selection}} \widehat{\xi}_n \quad = \quad (\widehat{\xi}_n^1, \ldots, \widehat{\xi}_n^N) \in E_n^N \xrightarrow{\text{Mutation}} \xi_{n+1} \in E_{n+1}^N \tag{5}$$

The selection transition consists in choosing randomly N particles $(\widehat{\xi}_n^1, \ldots, \widehat{\xi}_n^N)$ with common law

$$\sum_{i=1}^N \frac{g_n(\xi_n^i)}{\sum_{j=1}^N g_n(\xi_n^j)} \, \delta_{\xi_n^i}$$

After selection each particle $\widehat{\xi}_n^i$, $1 \leq i \leq N$, evolves according to the transition K_{n+1}. In other words ξ_{n+1}^i is a random variable with law $K_{n+1}(\widehat{\xi}_n^i, \,.\,)$.

Note that the larger $g_n(\xi_n^i)$ is the more likely particles $\widehat{\xi}_n^j$, $1 \leq j \leq N$, are inserted in site ξ_n^i. When a particle $\widehat{\xi}_n^j$ chooses a site ξ_n^i we can interpret ξ_n^i as being the "parent" of the individual $\widehat{\xi}_n^j$. In the same vein, recalling that ξ_n^i has been sampled according to distribution $K_n(\widehat{\xi}_{n-1}^i, \,.\,)$, the particle $\widehat{\xi}_{n-1}^i$ can also be regarded as the parent of ξ_n^i. In this way $\widehat{\xi}_{n-1}^i$ is an ancestor of $\widehat{\xi}_n^j$. Running back in time we can mentally trace in this way the whole genealogy of any particle.

Unfortunately the previous particle model does not carry any information about such ancestors. In this study we propose a particle Markovian model which allows us to trace at each time the complete genealogy of each individual.
Incidently through a suitable state space augmentation the historical process has exactly the same form as before. Therefore most of the asymptotic results known for previous genetic algorithm will also be valid for this genealogical model.
The key idea is to consider path-valued particles so that to trace the ancestral information back in time. Intuitively the resulting selection transition will consist in exchanging the whole ancestral information of the particles and the mutation only consists in extending each path with an elementary mutation transition.

1.2 Statement of the main results

This study is related to the one in [8] where the asymptotic behavior of the empirical measures

$$\eta_n^N = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_n^i} \in \mathcal{P}(E_n), \quad \text{and} \quad \eta_{[0,n]}^N = \frac{1}{N} \sum_{i=1}^N \delta_{(\xi_0^i, \ldots, \xi_n^i)} \in \mathcal{P}(E_0 \times \cdots \times E_n) \tag{6}$$

as N tends to infinity is considered. Although we have permitted here time dependent state spaces all results such as fluctuations and large deviations principles presented in there can be used without further work to study the convergence of the random empirical measures (6) to the deterministic ones

$$\eta_n \in \mathcal{P}(E_n), \quad \text{and} \quad \eta_{[0,n]} \stackrel{\text{def.}}{=} \eta_0 \otimes \cdots \otimes \eta_n \in \mathcal{P}(E_0 \times \cdots \times E_n)$$

In fact the time parameter $n \in \mathbb{N}$ can always be added to the state space as an additional variable. As a parenthesis if we consider the sequence $\mathcal{X}_n = (n, X_n)$ on the state space $E = \cup_n(\{n\} \times E_n)$ we do get a time homogeneous Markov chain $\mathcal{X}$ with transition

$$\mathcal{K}((p,x), d(q,y)) = \delta_{p+1}(dq)\ K_q(x,dy)$$

Nevertheless various useful mixing type conditions on the time inhomogeneous transitions K_q introduced in [8] and required in the sequel are not preserved by this state space augmentation. This is one of the reasons why we have chosen to consider an abstract time homogeneous Markov chain X. The other reason comes from the fact that the historical process model given next is best introduced in terms of a path-valued genetic model. Here again the time parameter can be added in the state space but we believe it is more transparent to describe the genealogy of individuals at each time $n \in \mathbb{N}$ in terms of ancestral paths from the origins up to time n.
We do not get into more details since a full discussion would be too great digression here but we would like to mention that the only assumption needed in the study of the convergence of the particle density profiles η_n^N is the following condition on the fitness function $\{g_n\ ;\ n \geq 0\}$

$$(\mathcal{G}) \qquad \forall n \in \mathbb{N}, \qquad \sup_{x,y\in E_n} |\log(g_n(x)/g_n(y))| < \infty.$$

In contrast, the investigation of uniform convergence results with respect to the time parameter for the N-approximating measures η_n^N and the asymptotic behavior of the empirical measures on path space $\eta_{[0,n]}^N$ rely on several strengthening of an additional mixing type condition on the mutation transitions $\{K_n\ ;\ n \geq 1\}$. To describe these different levels of assumptions it is useful to introduce a sequence of conditions indexed by a parameter $\alpha \in [1,\infty]$, namely

$$(\mathcal{K})_\alpha \qquad \forall n \in \mathbb{N}, \forall x \in E_n\ ,\ \ K_{n+1}(x,\ \bullet) \sim \eta_{n+1} \text{ and } \sup_{x\in E_n}(\tfrac{dK_{n+1}(x,\ \bullet)}{d\eta_{n+1}}) \in \mathbb{L}_\alpha(\eta_{n+1}).$$

With $\mathbb{P}_{[0,n]}^{(N)}$ denoting the distribution of the path particles

$$\forall 1 \leq i \leq N, \qquad \xi_{[0,n]}^i = (\xi_0^i, \ldots, \xi_n^i)$$

and $\mathbb{P}_{[0,n]}^{(N,q)}$, $1 \leq q \leq N$, its marginal on the first q-particles and $\mathbb{P}_n^{(N,q)}$ the marginal of the latter at time n our main result will basically be stated as follows.

Theorem 1.1 *For any $n \geq 0$ and $1 \leq q \leq N$ we have the following implications*

$$(\mathcal{G}) \implies \|\mathbb{P}_n^{(N,q)} - \eta_n^{\otimes q}\|_{\rm tv} \leq \frac{\exp(q\,c(n))}{N} \tag{7}$$

$$(\mathcal{G}) \text{ and } (\mathcal{K})_2 \implies \mathrm{Ent}(\mathbb{P}_{[0,n]}^{(N,q)} \mid \eta_{[0,n]}^{\otimes q}) \leq \frac{q\ c(n)}{N} \tag{8}$$

$$(\mathcal{G}) \text{ and } (\mathcal{K})_\alpha \implies \|\frac{d\mathbb{P}_n^{(N,q)}}{d\eta_n^{\otimes q}} - 1\|_{\alpha,\eta_n^{\otimes q}} \leq \frac{\exp(q\,c(n))}{N} \times \|\sup_x \frac{dK_n(x,\ .)}{d\eta_n}\|_{\alpha,\eta_n} \tag{9}$$

for some finite constant $c(n)$ whose value only depends on the time parameter n.
If $(\mathcal{G})$ and $(\mathcal{K})_1$ are both satisfied with

$$\sup_{n,x,y}\left|\log\frac{g_n(x)}{g_n(y)}\right| < \infty \quad \text{and} \quad \sup_{n,x,y}\left|\log\frac{dK_n(x,\ .)}{d\eta_n}(y)\right| < \infty \tag{10}$$

then

$$\operatorname{Ent}\left(\mathbb{P}^{(N,q)}_{[0,n]}\Big|\eta^{\otimes q}_{[0,n]}\right) \leq \frac{q\ n\ c}{N} \qquad \text{and} \qquad \sup_{n\geq 0}\|\frac{d\mathbb{P}^{(N,q)}_n}{d\eta_n^{\otimes q}} - 1\|_{\infty,\eta_n^{\otimes q}} \leq \frac{\exp{(q\,c)}}{N} \tag{11}$$

for some finite constant c which does not depend on the time parameter n.

Remark 1.2: The entropy and $\mathbb{L}_\alpha$-estimates are stronger than those in variation. In fact if we want to translate all these estimates in terms of the total variation distance we use the inequalities

$$\forall \mu_1 << \mu_2\,, \qquad 2\ \|\mu_1-\mu_2\|^2_{\mathrm{tv}} \leq \operatorname{Ent}(\mu_1|\mu_2)\ \leq\ \|\frac{d\mu_1}{d\mu_2}-1\|^2_{2,\mu_2}$$

(cf. [2], Theorem 4.1 for the first inequality and in this paper Lemma 3.10, p. 28 for the $\mathbb{L}_2$ bound). Consequently, under $(\mathcal{K})_2$ we also have the estimates for the total variation distance

$$\|\mathbb{P}^{(N,q)}_{[0,n]} - \eta^{\otimes q}_{[0,n]}\|_{\mathrm{tv}} \leq \sqrt{\frac{q}{N}}\ c'(n) \quad \text{and} \quad \|\mathbb{P}^{(N,q)}_{n} - \eta^{\otimes q}_{n}\|_{\mathrm{tv}} \leq \frac{\exp{(q\,c'(n))}}{N}$$

for some finite constant $c'(n)$ whose value only depends on the parameter n. In addition, if the uniform bounds (10) hold we have

$$\|\mathbb{P}^{(N,q)}_{[0,n]} - \eta^{\otimes q}_{[0,n]}\|_{\mathrm{tv}} \leq \sqrt{\frac{q\ n}{N}}\ c' \quad \text{and} \quad \sup_{n\geq 0}\|\mathbb{P}^{(N,q)}_{n} - \eta^{\otimes q}_{n}\|_{\mathrm{tv}} \leq \frac{\exp{(q\,c')}}{N}$$

for some finite constant c' which does not depend on the time parameter.
As a direct corollary of Theorem 1.1 we deduce strong versions of propagation of chaos for appropriate increasing block size $q(N)$ and time horizon $n(N)$.

Corollary 1.3 *Under $(\mathcal{G})$ and $(\mathcal{K})_2$ the following implication holds.*

$$\lim_{N\rightarrow\infty}\frac{q(N)}{N}=0 \implies \forall n\geq 0\,, \quad \lim_{N\rightarrow\infty}\operatorname{Ent}(\mathbb{P}^{(N,q(N))}_{[0,n]}\ |\ \eta^{\otimes q(N)}_{[0,n]})=0$$

In addition if (10) holds then

$$\lim_{N\rightarrow\infty}\frac{q(N)\ n(N)}{N}=0 \implies \lim_{N\rightarrow\infty}\operatorname{Ent}(\mathbb{P}^{(N,q(N))}_{[0,n(N)]}\ |\ \eta^{\otimes q(N)}_{[0,n(N)]})=0$$

and there exists some finite constant $q_0<\infty$ such that

$$\forall N\geq 1\,,\quad q(N)=q_0\log N \implies \lim_{N\rightarrow\infty}\sup_{n\geq 0}\|\frac{d\mathbb{P}^{(N,q(N))}_n}{d\eta_n^{\otimes q(N)}}-1\|_{\infty,\eta_n^{\otimes q(N)}}=0$$

1.3 Notes and Contents

This study has been influenced by papers of Ben Arous and Zeitouni [1], Donnelly and Kurtz [11, 12], Graham and Méléard [14] and Méléard [15].

To the best of our knowledge the idea to use genealogical paths to trace the ancestral information of particles first appeared in Ethier and Griffiths [13]. In this work the authors study the historical process with mutations but no selection as a measure valued diffusion process. More recently Donnelly and Kurtz study in [11, 12] genealogical processes with selections and recombinations in the context of Fleming-Viot processes. In this work the authors consider a specific infinite population model with a particular labeling of the individuals so that to identify the type of ancestors of populations. In contrast to our situation the limiting Fleming-Viot model is a random processes. Although our discrete generation particle models differ from the ones discussed in [11, 12] such labeling techniques can probably be used in our setting. In contrast, the strategy we have chosen here is to augment the state space so that to represent the historical process as a genetic type interacting particle system in path space. We believe this construction is more transparent than previously published genealogical models in genetic populations literature. It also gives novel and explicit connections between the distributions of genealogical trees of genetic populations, path-valued genetic algorithms and Feynman-Kac formulae.

In [14, 15] the authors present strong propagation of chaos for the total variation distance for the N-particle approximating model associated to a class of generalized Boltzmann equations. Their approach is essentially based on interacting graphs and precise coupling techniques. They show that the order of convergence for the total variation distance between the law of the q-first particles and the limiting distribution on a compact interval $[0, t]$ is $q^2\, c(t)/N$. Their result does not depend on the form of the mutation transition and it implies a increasing propagation of chaos for the total variation distance for block size $q(N) = o(\sqrt{N})$. The relationships between spatially homogeneous Boltzmann equations and continuous time Feynman-Kac formulae are described in some details in [8, 9]. In this connection the increasing propagation of chaos for the relative entropy for block size $q(N) = o(N)$ given in Corollary 1.3 can be viewed as an improvement of earlier results. But in view of the work of Graham and Méléard [14] and Méléard [15] we believe that the estimate in total variation distance (7) is not sharp and it can probably be extended to distributions in path space. In other words we make the plausible conjecture that the r.h.s. upper bound in (7) can be replaced by $q^2\, c(n)/N$.

As we shall see in the further development of section 3.2 the entropy estimates on path space presented in Theorem 1.1 are based on the fact that the distribution of particles can be regarded as a mean field Gibbs measure. In the recent and seminal paper [1] Ben Arous and Zeitouni proved increasing propagation of chaos for block size $q(N) = o(N)$ for a large class of Gibbs measures with polynomial interactions and bounded Gibbs potential function ([1] also contains a useful bibliography on this subject). As will be seen our potential function does not satisfy these conditions and it does not fit into the framework of classical literature on the subject. In this work we present an alternative approach. In contrast to [1] our strategy is not based on Laplace asymptotics and Banach space embedding. We also extend the previous result in three different ways. First we extend this result to a fairly general class of N-particle approximating models associated to an abstract sequence of functions $\{\Phi_n \,;\, n \geq 1\}$. Second we present an increasing propagation of chaos for a pair (block size/time horizon) $q(N) \times n(N) = o(N)$. Finally we propose a novel uniform increasing propagation of chaos, w.r.t. the time parameter for the n-th marginals $\mathbb{P}_n^{(N,q)}$.

This paper is divided in two parts devoted respectively to the modeling of the historical process in terms of genetic algorithms in path space and the study of strong versions of propagation of chaos.

In section 2 we propose a particle model for the historical process associated to a genetic algorithm. Section 2.1 introduces a path-valued interacting particle systems model. In section 2.2 we show that the latter can be regarded as the historical process associated to a genetic algorithm. The modeling impact of this framework in the study of non linear filtering and smoothing problems is performed in section 2.3. In section 2.3 we examine several examples of fitness functions and mutation transitions which fit into our framework.

Section 3 is concerned with the proof of Theorem 1.1. This section is divided in two parts. In section 3.1 we study the asymptotic behavior as N tends to infinity of the time marginals $\{P_n^{(N,q)}\ ;\ N\geq 1\}$. In section 3.2 we prove the entropy estimates on path space. The approach taken in here is different from the one we took in section 3.1 and it can be read independently of the latter. Furthermore we propose $\mathbb{L}_\alpha$-conditions on the N-particle approximating model underwhich an increasing propagation of chaos holds for a fairly general class of functions $\{\Phi_n\ ;\ n\geq 0\}$. Incidently we shall see that these criteria also appear to be useful in the study of weak convergence of the corresponding empirical processes.

2 Path-valued Interacting Particle Systems

2.1 Description of the Models

Let us suppose that X is a path-valued Markov chain of the following type

$$\forall n\in\mathbb{N}\,,\qquad X_n=(X'_0,\ldots,X'_n)\in E_n=E'_0\times\cdots\times E'_n \tag{12}$$

where $X'=\{X'_n\ ;\ n\geq 0\}$ is an auxiliary Markov chain taking values in an additional collection of measurable spaces $\{(E'_n,\mathcal{E}'_n)\ ;\ n\geq 0\}$. Since $X_0=X'_0$ we have that $E_0=E'_0$. We also notice that the Markov transitions $\{K_n\ ;\ n\geq 1\}$ of X and the Markov transitions $\{K'_n\ ;\ n\geq 1\}$ of X' are connected by the formula

$$\begin{aligned}&\forall n\in\mathbb{N}\,,\\ &K_{n+1}((x'_0,\ldots,x'_n),d(y'_0,\ldots,y'_{n+1}))=\delta_{(x'_0,\ldots,x'_n)}(d(y'_0,\ldots,y'_n))\ K'_{n+1}(y'_n,dy'_{n+1})\end{aligned} \tag{13}$$

We finally note that for this path-valued Markov chain the Feynman-Kac distributions (1) have the form

$$\forall n\in\mathbb{N}\,,\ \forall f\in\mathcal{B}_b(E_n)\,,\qquad \begin{aligned}\gamma_n(f)&=\mathbb{E}(\quad f(X_n)\qquad \textstyle\prod_{m=0}^{n-1}\ g_m(X_m))\\ &=\mathbb{E}(f(X'_0,\ldots,X'_n)\ \textstyle\prod_{m=0}^{n-1}\ g_m(X'_0,\ldots,X'_m))\end{aligned} \tag{14}$$

Using the framework of the preceding section we see that the N-genetic approximating model consists here in path-valued particles

$$\forall n\in\mathbb{N}\,,\forall 1\leq i\leq N\,,\qquad \left\{\begin{array}{lcl}\xi_n^i&=&(\xi'^i_{0,n},\ldots,\xi'^i_{n,n})\in E_n=E'_0\times\cdots\times E'_n\\ \widehat{\xi}_n^i&=&(\widehat{\xi}'^i_{0,n},\ldots,\widehat{\xi}'^i_{n,n})\in E_n=E'_0\times\cdots\times E'_n\end{array}\right. \tag{15}$$

The selection transition consists in choosing randomly N-path particles $(\widehat{\xi}'^i_{0,n},\ldots,\widehat{\xi}'^i_{n,n})$, $1\leq i\leq N$, with common law

$$\sum_{i=1}^N\ \frac{g_n(\xi'^i_{0,n},\ldots,\xi'^i_{n,n})}{\sum_{j=1}^N g_n(\xi'^j_{0,n},\ldots,\xi'^j_{n,n})}\ \delta_{(\xi'^i_{0,n},\ldots,\xi'^i_{n,n})}$$

In view of (13) and during the mutation transition each end point $\widehat{\xi}'^i_{n,n}$ evolves randomly according to the transition K'_{n+1}, that is

$$\begin{aligned} \xi^i_{n+1} &= ((\xi'^i_{0,n+1},\ldots,\xi'^i_{n,n+1}),\xi'^i_{n+1,n+1}) \\ &= ((\widehat{\xi}'^i_{0,n},\ldots\ldots,\widehat{\xi}'^i_{n,n}),\ \ \xi'^i_{n+1,n+1}) \in E_{n+1} = E_n \times E'_{n+1} \end{aligned}$$

where $\xi'^i_{n+1,n+1}$ is a random variable with law $K'_{n+1}(\widehat{\xi}'^i_{n,n},\,\boldsymbol{.})$.

2.2 The Historical Process

To see the strength of the preceding modeling it is instructive to note that each path-particle

$$\xi^i_n = (\xi'^i_{0,n},\ldots,\xi'^i_{n,n}) \quad \text{and} \quad \widehat{\xi}^i_n = (\widehat{\xi}'^i_{0,n},\ldots,\widehat{\xi}'^i_{n,n})$$

can be regarded as the genealogical branch of the end point particles $\xi'^i_{n,n}$ and $\widehat{\xi}'^i_{n,n}$. In this sense the former model can be regarded as a genealogical path-valued genetic algorithm. If we use a graphical representation we easily see that the set of all individuals and vertices defined formally by setting

$$\forall n\in\mathbb{N}\,,\ \ \forall 1\leq i\leq N\,,\qquad \xi'^i_{0,n}\longrightarrow \xi'^i_{0,n}\longrightarrow\ldots\longrightarrow \xi'^i_{n-1,n}\longrightarrow \xi'^i_{n,n}$$

represents the complete genealogy of the population $\{\xi'^i_{n,n}\ ;\ 1\leq i\leq N\}$ at time $n\in\mathbb{N}$. At closer inspection we also notice that selection acts on the whole ancestors but the mutation stage does not affect the ancestry levels but each genealogical path is only extended with an elementary move according to the transitions K'_n.

Next we examine the situation in which the selection pressure only depends on the end point particles. In this situation we will see that the resulting marginal model formed by the projections of the path particles in $E'_0\times\ldots\times E'_n$ on the sets E'_n is again a selection/mutation genetic algorithm. Furthermore the complete genealogy of this marginal model can be recovered in a natural and simple way from the path-valued model.

Suppose the fitness function g_n only depends on the n-th component, that is

$$\forall n\in\mathbb{N}\,,\qquad \begin{array}{rlcl} g_n: & E_n = E'_0\times\cdots\times E'_n & \longrightarrow &]0,\infty[\\ & x_n=(x'_0,\ldots,x'_n) & \longrightarrow & g_n(x_n)=g'_n(x'_n) \end{array}$$

for some bounded and strictly positive function g'_n on E'_n. In this situation we see that the marginal model

$$\xi'_n \stackrel{\text{def.}}{=} (\xi'^1_{n,n},\ldots,\xi'^N_{n,n})\in E'^N_n \xrightarrow{\text{Selection}} \widehat{\xi}'_n \stackrel{\text{def.}}{=} (\widehat{\xi}'^1_{n,n},\ldots,\widehat{\xi}'^N_{n,n})\in E'^N_n \xrightarrow{\text{Mutation}} \xi'_{n+1}\in E'^N_{n+1}$$

is again a selection/mutation genetic algorithm. It is clearly defined as in (5) by replacing g_n and K_n by g'_n and K'_n.

To clarify the presentation we slight abuse notations suppressing the double notational dependence on the time parameter and we simply write ξ'^i_n and $\widehat{\xi}'^i_n$ instead of $\xi'^i_{n,n}$ and $\widehat{\xi}'^i_{n,n}$

Using these simplified notations the selection/mutation transitions are defined as follows. During the selection stage the N-particles $\widehat{\xi}'_n = (\widehat{\xi}'^1_n, \ldots, \widehat{\xi}'^N_n)$ are chosen independently with the distribution

$$\sum_{i=1}^{N} \frac{g'_n(\xi'^i_n)}{\sum_{j=1}^{N} g'_n(\xi'^j_n)} \, \delta_{\xi'^i_n}$$

After selection, each particle $\widehat{\xi}'^i_n$ evolves randomly according to K'_{n+1}, so that for each $1 \leq i \leq N$, ξ'^i_{n+1} denotes a random variable with law $K'_{n+1}(\widehat{\xi}'^i_n, \centerdot)$.

As announced, the path valued selection/mutation genetic algorithm $\{\xi_n, \widehat{\xi}_n \ ; \ n \geq 0\}$ gives precisely the time evolution of the genealogical structure of the latter. More precisely the path-valued particles

$$\forall 1 \leq i \leq N \,, \qquad \widehat{\xi}^i_n = (\widehat{\xi}'^i_{0,n}, \ldots, \widehat{\xi}'^i_{n,n}) \in E_n = E'_0 \times \cdots \times E'_n$$

represent the line of ancestors $(\widehat{\xi}'^i_{0,n}, \ldots, \widehat{\xi}'^i_{n-1,n})$ of the individual $\widehat{\xi}'^i_n$. During the mutation transition the branch of ancestors does not change and the "parent" particle $\widehat{\xi}'^i_{n,n}$ evolves according to K'_{n+1}.

From previous observations it is also easily seen that the Markov chain

$$\{\xi'_n, \widehat{\xi}'_n \ ; \ n \in \mathbb{N}\} \tag{16}$$

is the N-genetic approximating model associated to a measure valued dynamical system

$$\forall n \in \mathbb{N} \,, \qquad \eta'_{n+1} = \Phi'_{n+1}(\eta'_n)$$

where $\eta'_0 = \eta_0 \in \mathcal{P}(E'_0) = \mathcal{P}(E_0)$ and the one step mappings

$$\forall n \in \mathbb{N} \,, \qquad \Phi'_{n+1} : \mathcal{P}(E'_n) \longrightarrow \mathcal{P}(E'_{n+1})$$

are defined as in (3) by replacing g_n, K_n and E_n by g'_n, K'_n and E'_n. It is also easily checked that the resulting distributions γ'_n and η'_n are the n-th marginal of γ_n and η_n, that is

$$\forall n \in \mathbb{N} \,, \forall f' \in \mathcal{B}_b(E'_n) \,, \qquad \eta'_n(f') = \frac{\gamma'_n(f')}{\gamma'_n(1)} \tag{17}$$

with

$$\gamma'_n(f') = \mathbb{E}(f'(X'_n) \prod_{m=0}^{n-1} g'_m(X'_m)) \tag{18}$$

Condition $(\mathcal{K})_1$ is clearly never met for the mutation transitions $\{K_n \ ; \ n \geq 1\}$ defined in (13). However if the elementary transitions $\{K'_n \ ; \ n \geq 1\}$ of the marginal model satisfy for some $\alpha \in [1, \infty]$ the mixing type condition

$$(\mathcal{K}')_\alpha \quad \forall n \in \mathbb{N} \,, \forall x \in E'_n \,, \ \ K'_{n+1}(x, \bullet) \sim \eta'_{n+1} \ \text{ and } \ \sup_{x \in E_n}\left(\frac{dK'_{n+1}(x, \bullet)}{d\eta'_{n+1}}\right) \in \mathbb{L}_\alpha(\eta'_{n+1}).$$

then techniques presented in [8] apply to study uniform convergence results and the asymptotic behavior of the empirical measures

$$\eta'^N_n = \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi'^i_n} \in \mathcal{P}(E'_n), \quad \text{and} \quad \eta'^N_{[0,n]} = \frac{1}{N} \sum_{i=1}^{N} \delta_{(\xi'^i_0, \ldots, \xi'^i_n)} \in \mathcal{P}(E'_0 \times \cdots \times E'_n)$$

as $N \to \infty$ to the deterministic measures

$$\eta'_n \in \mathcal{P}(E'_n), \quad \text{and} \quad \eta'_{[0,n]} \stackrel{\text{def.}}{=} \eta'_0 \otimes \cdots \otimes \eta'_n \in \mathcal{P}(E'_0 \times \cdots \times E'_n)$$

Moreover the conclusions of theorem 1.1 remains valid if we replace ξ_n, E_n, K_n and $(\mathcal{K})_\alpha$ by ξ'_n, E'_n, K'_n and $(\mathcal{K}')_\alpha$.

Another remark is that most of the convergence results presented in [8] such as fluctuations, Donsker and Glivenko-Cantelli Theorem but also exponential rates and large deviations only rely on assumption $(\mathcal{G})$. Recalling that the historical process is nothing else than a particular example of genetic model the latter results can be used without further work to study the asymptotic behavior of the genealogical-path empirical measures

$$\eta_n^N = \frac{1}{N}\sum_{i=1}^N \delta_{\xi_n^i} = \frac{1}{N}\sum_{i=1}^N \delta_{(\xi'^i_{0,n},\ldots,\xi'^i_{n,n})} \in \mathcal{P}(E_n) = \mathcal{P}(E'_0 \times \ldots \times E'_n)$$

Our last remark is that the estimate (7) in Theorem 1.1 is again independent on the form of the mutation transition. Thus it can be used to evaluate the distance in total variation between $\eta_n^{\otimes q}$ and the distribution of the first q-genealogical path-valued particles $\left\{(\xi'^i_{0,n},\ldots,\xi'^i_{n,n})\ ;\ 1 \le i \le q\right\}$.

2.3 Applications to Non Linear Filtering

The Feynman-Kac formulae (14) and (18) and their particle approximating models play a major role in the theory of non linear filtering. In mathematical terms the non linear filtering problem can be expressed as follows.

Let $(X,Y) = \{(X_n,Y_n)\ ;\ n \ge 0\}$ be a Markov chain taking values in some product spaces $\{(E_n \times F_n)\ ;\ n \ge 0\}$. Here $\{(F_n,\mathcal{F}_n)\ ;\ n \ge 0\}$ is an auxiliary sequence of measurable spaces. Further we assume that the initial distribution μ_0 the Markov transitions $\{G_n\ ;\ n \ge 1\}$ of (X,Y) have the form

$$\begin{aligned} \mu_0(d(x_0,y_0)) &= \overline{g}_n(x_0,y_0)\ \eta_0(dx_0)\ \gamma_0(dy_0) &(19)\\ G_n((x_{n-1},y_{n-1}),d(x_n,y_n)) &= \overline{g}_n(x_n,y_n)\ K_n(x_{n-1},dx_n)\ \gamma_n(dy_n) &(20)\end{aligned}$$

where, for each $n \in \mathbb{N}$, $\overline{g}_n : E_n \times F_n \to]0,\infty[$ is a strictly positive function and $\gamma_n \in \mathcal{P}(F_n)$.

The non linear filtering problem consists in computing the conditional distributions of the state signal X given the observations Y. To understand the motivations behind this problem we can think the signal X as being the Markovian model for the time evolution of a target in tracking problems or an aircraft in radar signal processing. The observation process models the noisy and partial information delivered by sensors as radars or sonars. Of course the exact values of the signal X and the values of the various disturbance sources are not known but it is reasonable to assume that we know their statistics. This corresponds to the situation in which η_0, $\overline{g}_n$ and K_n are explicitly known (the interested reader is referred to [6] for a discussion of some practical problems in which we also need to approximate these three parameters).

In engineering and advanced signal processing literature an alternative and more classical way to define the pair (signal/observation) Markov process (X, Y) is as follows. The signal $X = \{X_n \,;\, n \in \mathbb{N}\}$ is a Markov chain with transition probability kernels $\{K_n \,;\, n \geq 1\}$ and taking values at each time n in some measurable space $(E_n, \mathcal{E}_n)$ and the observation process is defined by

$$\forall n \in \mathbb{N}\,, \qquad Y_n = H_n(X_n, V_n) \tag{21}$$

The sequence $V = \{V_n \,;\, n \in \mathbb{N}\}$ is independent of X and it represents the noise sources. It consists of a collection of independent random variables taking values in some auxiliary measurable spaces $\{(S_n, \mathcal{S}_n) \,;\, n \in \mathbb{N}\}$. For each $n \in \mathbb{N}$, the random variable V_n is distributed according to a probability measure $\gamma_n \in \mathcal{P}(S_n)$. The collection of measurable functions $H_n : E_n \times S_n \to F_n$ is chosen so that

$$\forall n \in \mathbb{N}\,,\ \forall x_n \in E_n\,, \qquad \mathbb{P}\left(H_n(x, V_n) \in dy\right) = \overline{g}_n(x_n, y)\ \gamma_n(dy)$$

In other words the laws of the random variables $H_n(x, V_n)$ and V_n are absolutely continuous and $\overline{g}_n(x_n, \textbf{.})$ is the corresponding density.

If we fix the sequence of observations $Y = y$ then a version of the conditional distributions of the states of the signal given their noisy observations can be expressed in terms of Feynman-Kac formulae of the same type as the ones discussed above. More precisely, if we take

$$\forall n \in \mathbb{N}\,,\ \forall x \in E_n\,, \qquad g_n(x) = \overline{g}_n(x_n, y_n)$$

in (1) we have for any $f \in \mathcal{B}_b(E_n)$,

$$\begin{aligned}
\eta_n(f) &= \mathbb{E}(\ f(X_n) \mid Y_0 = y_0, \ldots, Y_{n-1} = y_{n-1}\ ) \\
\widehat{\eta}_n(f) &= \frac{\eta_n(f g_n)}{\eta_n(g_n)} = \mathbb{E}(\ f(X_n) \mid Y_0 = y_0, \ldots, Y_n = y_n\ )
\end{aligned}$$

It is also interesting to examine the situation where X is given by (12). Namely, suppose we have that

$$\forall n \in \mathbb{N}\,, \qquad X_n = (X'_0, \ldots, X'_n) \in E_n = E'_0 \times \cdots \times E'_n$$

where $X' = \{X'_n \,;\, n \geq 0\}$ is Markov chain taking values in some measurable spaces $\{(E'_n, \mathcal{E}'_n) \,;\, n \geq 0\}$ with initial distribution η_0 and transitions $\{K'_n \,;\, n \geq 1\}$. In this situation the observation sequence (21) takes the form

$$\forall n \in \mathbb{N}\,, \qquad Y_n = H_n((X'_0, \ldots, X'_n), V_n)$$

This means that the information delivered by sensors at each time n depends on the whole path $(X'_0, \ldots, X'_n)$ of the signal X' back from the origin and up to time n. Note that in this case the function $\overline{g}_n((x'_0, \ldots, x'_n), y_n)$ depends on the current observation $Y_n = y_n$ and on the whole path-coordinates $(x'_0, \ldots, x'_n)$. This type of sensor is in fact more general than those arising in practice. In classical filtering problems the observation sequence is rather defined by

$$\forall n \in \mathbb{N}\,, \qquad Y_n = H'_n(X'_n, V_n)$$

for some appropriate function $H'_n : E'_n \times S_n \to F_n$ and the resulting function $\overline{g}_n(x'_n, y_n)$ only depends the end point coordinate x'_n of the path $(x'_0, \ldots, x'_n)$ that is

$$\overline{g}_n((x'_0, \ldots, x'_n), y_n) = \overline{g}'_n(x'_n, y_n)$$

for some strictly positive function $\overline{g}'_n : E'_n \to]0, \infty[$. We emphasize that in this particular situation the pair process $(X', Y) = \{(X'_n, Y_n) \ ; \ n \geq 0\}$ has the same form as before. It is a Markov chain taking values in the measurable spaces $\{(E'_n \times F_n) \ ; \ n \geq 0\}$. The initial distribution and the Markov transitions of (X', Y) are defined as in (19) and (20) by replacing $(\overline{g}_n, K_n)$ by $(\overline{g}'_n, K'_n)$. From these observations one concludes that

$$\begin{array}{rcl} \eta'_n(f') & = & \mathbb{E}(\ f'(X'_n) \mid Y_0 = y_0, \ldots, Y_{n-1} = y_{n-1}\ ) \\ \eta_n(f) & = & \mathbb{E}(\ f(X'_0, \ldots, X'_n) \mid Y_0 = y_0, \ldots, Y_{n-1} = y_{n-1}\ ) \end{array} \qquad (22)$$

for any $f' \in \mathcal{B}_b(E'_n)$ and $f \in \mathcal{B}_b(E'_0 \times \ldots \times E'_n)$. The historical process (15) associated to the N-particle approximating model (16) of the limiting distributions $\{\eta'_n \ ; \ n \in \mathbb{N}\}$ is therefore an particle approximating model of distributions $\{\eta_n \ ; \ n \in \mathbb{N}\}$.

2.4 Examples

To illustrate and motivate our study we end this introductory section with some comments on assumptions $(\mathcal{G})$ and $(\mathcal{K})_\alpha$. Assumption $(\mathcal{G})$ clearly holds as soon as the fitness functions satisfy

$$\forall n \geq 0\,, \forall x, x' \in E_n\,, \qquad \frac{1}{a_n} \leq \frac{g_n(x)}{g_n(x')} \leq a_n \qquad (23)$$

for some collection of numbers $a_n \in [1, \infty[$, $n \geq 0$. In non linear filtering settings the fitness function g_n depends on the current observation $Y_n = y_n$ at time n, that is

$$\forall n \in \mathbb{N}\,, \forall x \in E_n\,, \qquad g_n(x) = \overline{g}_n(x, y_n)$$

In terms of the densities $\overline{g}_n$ condition (23) reads

$$\forall n \geq 0\,, \forall x, x' \in E_n\,, \forall y \in F_n\,, \qquad \frac{1}{\overline{a}_n(y)} \leq \frac{\overline{g}_n(x, y)}{\overline{g}_n(x', y)} \leq \overline{a}_n(y) \qquad (24)$$

for some functions $\overline{a}_n : F_n \to [1, \infty[$ and with these notations $a_n = \overline{a}_n(y_n)$.

Example 2.1 *As a typical example of nonlinear filtering problem assume the observations take values in $\mathbb{R}^d$ and densities $\overline{g}_n$ are given by*

$$\forall n \in \mathbb{N}\,,\ \ \overline{g}_n(x, y) = \frac{1}{((2\pi)^d |R_n|)^{1/2}}\ \exp\left(-\frac{1}{2}\,(y - h_n(x))'\ R_n^{-1}\ (y_n - h_n(x))\right)$$

For any $n \in \mathbb{N}$ $h_n : E_n \to \mathbb{R}^d$ is a bounded measurable functions and R_n is a $d \times d$ symmetric positive matrix. This correspond to the situation where the observation sequence is given by

$$\forall n \in \mathbb{N}, \qquad Y_n = h_n(X_n) + V_n$$

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 can check that (24) holds with

$$\log \overline{a}_n(y) = \mathrm{osc}_\alpha(h_n)\ |||R_n^{-1}|||_\beta\ (\|y\|_\beta + \|h_n\|_\beta)$$

as soon as for some $\alpha, \beta \geq 1$ with $\frac{1}{\alpha} + \frac{1}{\beta} = 1$ we have

$$\begin{aligned} |||R_n^{-1}|||_\beta &= \sup\left\{||R_n^{-1}z||_\beta\ ;\ z \in \mathbb{R}^d\ ,\ \ \|z\|_\beta = 1\right\} < \infty \\ \mathrm{osc}_\alpha(h_n) &= \sup_{x,x' \in E_n} \|h_n(x) - h_n(x')\|_\alpha < \infty \quad \text{and} \quad \|h_n\|_\beta = \sup_{x \in E_n} \|h_n(x)\|_\beta < \infty \end{aligned}$$

To connect assumptions $(\mathcal{K})_\alpha$ with some more easily verifiable properties on the one step mappings K_n let us suppose that for each $n \geq 1$ there exists a reference probability measure $\lambda_n \in \mathcal{P}(E_n)$ such that $K_n(x, .) \sim \lambda_n$ for any $x \in E_{n-1}$ and

$$\forall y \in E_n\,, \qquad \sup_{x \in E_{n-1}} |\log l_n(x,y)| \leq b_n(y) \qquad \text{where} \quad l_n(x,y) = \frac{dK_n(x,.)}{d\lambda_n}(y)$$

for some positive function $b_n : E_n \to [0,\infty[$. A simple calculation shows that for any choice of the reference probability measure $\lambda_n \in \mathcal{P}(E_n)$ and for any $\alpha \geq 1$ we have

$$\int \sup_x \left(\frac{dK_n(x,.)}{d\eta_n}\right)^\alpha d\eta_n = \int \sup_x \left(\frac{dK_n(x,.)}{d\lambda_n}\right)^\alpha \left(\frac{d\eta_n}{d\lambda_n}\right)^{1-\alpha} d\eta_n \tag{25}$$

Since

$$\frac{d\eta_n}{d\lambda_n}(y) = \frac{\eta_{n-1}(g_{n-1}\ l_n(.,y))}{\eta_{n-1}(g_{n-1})}$$

we find that

$$\left|\log \frac{d\eta_n}{d\lambda_n}\right| \leq b_n$$

From (25) one concludes that

$$\int \sup_x \left(\frac{dK_n(x,.)}{d\eta_n}\right)^\alpha d\eta_n \leq \int e^{(2\alpha-1)\,b_n}\, d\lambda_n$$

It is now easily checked that

$$\begin{aligned} &\forall n \geq 1\,, \quad \int e^{(2\alpha-1)b_n(y)}\ \lambda_n(dy) \implies (\mathcal{K})_\alpha \\ &\sup_{n\geq 0} a_n < \infty \ \text{ and } \ \sup_{n \geq 1} \|b_n\| < \infty \implies (10) \end{aligned}$$

Example 2.2 *Suppose that $E = \mathbb{R}^d$ and K_n is given by*

$$K_n(x,dy) = \frac{1}{((2\pi)^d|Q_n|)^{1/2}}\ \exp\left(-\frac{1}{2}(y - B_n(x))'\ Q_n^{-1}\ (y - B_n(x))\right)\, dy$$

where Q_n is a $d \times d$ symmetric nonnegative matrix and $B_n : \mathbb{R}^d \to \mathbb{R}^d$ is a bounded function. Using previous observations it is not difficult to check that $(\mathcal{K})_\alpha$ is satisfied for any $\alpha \geq 1$ with

$$\lambda_n(dy) = \frac{1}{((2\pi)^d|Q_n|)^{1/2}}\ \exp\left(-\frac{1}{2}y'\ Q_n^{-1}\ y\right)\, dy.$$

and

$$\begin{aligned} \left|\log \frac{dK_n(x,.)}{d\lambda_n}(y)\right| &= \left|b_n(x)'Q_n^{-1}y\right| + \frac{1}{2}\ \left|b_n(x)'\ Q_n^{-1}\ b_n(x)\right| \\ &\leq\ b_n(y) = \|B_n\|\ |||Q_n^{-1}|||_1\ \|y\|_1 + \frac{d}{2}\ \|B_n\|^2\ |||Q_n^{-1}|||_1 \end{aligned}$$

where

$$\|B_n\| = \sup_{1\leq i\leq d} \sup_x |B_n^i(x)|\,, \quad |||Q_n^{-1}|||_1 = \sup_{1\leq j\leq d} \sum_{i=1}^d |(Q_n^{-1})_{i,j}|\,, \quad \|y\|_1 = \sum_{i=1}^d |y^i|$$

For $d=1$ and K_n given by

$$K_n(x,dy) = \frac{c(n)}{2}\; e^{-c(n)\,|y-B_n(x)|}\; dy$$

for some $c(n)>0$ conditions $(\mathcal{K})_\alpha$ also hold for $\alpha=\infty$. Indeed, if we choose

$$\lambda_n(dy) = \frac{c(n)}{2}\; e^{-c(n)\,|y|}\; dy$$

we clearly have

$$\left|\log \frac{dK_n(x,\centerdot)}{d\lambda_n}(y)\right| \leq b_n(y) \quad \text{with} \quad \forall y\in\mathbb{R}\,, \;\; b_n(y) = c(n)\;\|B_n\|$$

We also notice that in this situation $\sup_{n\geq 1}(c(n)\;\|B_n\|)<\infty \Rightarrow \sup_{n\geq 1}(\|b_n\|)$.

3 Strong Propagation of Chaos

The object of this section is to prove Theorem 1.1. As announced in the introductive section 1.3 our analysis involves two different techniques. Therefore we have chosen to decompose the proof of the theorem in two independent sections.

In section 3.1 we study the asymptotic behavior as N tends to infinity of the time marginals $\{P_n^{(N,q)}\;;\;N\geq 1\}$. The approach taken here is partly based on ideas presented in [8]. More precisely our strategy will be to extend the analysis of the asymptotic behavior of the empirical measures η_n^N as N tends to infinity to their q-tensor product measures $(\eta_n^N)^{\otimes q}$.

Section 3.2 discusses entropy estimates on path space. This investigation relies on the fact that, under $(\mathcal{K})_1$, the distribution of the particles can be viewed as a mean field Gibbs measure on path space with partition function equals to 1. As we shall see in the further development of section 3.2 this property simplifies considerably the analysis since it allows us to relate without further work the desired relative entropy with more tractable $\mathbb{L}_2$-norms.

3.1 Time Marginal Estimates

Throughout this section $\mathcal{B}_b^+(E)$ denotes the set of bounded and strictly positive functions g on a measurable space $(E,\mathcal{E})$ such that $\sup_{x,y}|g(x)/g(y)|<\infty$.

Let N and q be two natural numbers such that $N,q\geq 1$. For any N-vector $x=(x^1,\ldots,x^N)\in E^N$, let $m(x)^{\odot q}$ be the finite measure on E^q defined by

$$m(x)^{\odot q} = \frac{1}{N^q}\sum_{(i_1,\ldots,i_q)\in I(q)} \delta_{(x^{i_1},\ldots,x^{i_q})}$$

where $I(q)$ is the set of q-indices $(i_1,\ldots,i_q)\in\{1,\ldots,N\}^q$ such that $i_j\neq i_{j'}$ for any $1\leq j\neq j'\leq N$.

The main object of the section is to prove the following result.

Theorem 3.1 *Under assumption* $(\mathcal{G})$ *we have*

$$\forall n\in\mathbb{N}\,,\forall(f,g)\in(\mathcal{B}_b(E_n^q)\times\mathcal{B}_b^+(E_n^q))\,,\forall\sqrt{N/2}\geq q\geq 1\,,$$

$$\mathbb{E}\left(\left|\Psi_g(m(\xi_n)^{\odot q})(f)-\Psi_g(\eta_n^{\otimes q})(f)\right|^2\right)^{\frac{1}{2}} \leq \frac{e^{qc(n)}}{\sqrt{N}}\,\|f\|\,\sup_{x,y}|g(x)/g(y)|^2 \tag{26}$$

$$\left|\mathbb{E}\left(\Psi_g(m(\xi_n)^{\odot q})(f)\right)-\Psi_g(\eta_n^{\otimes q})(f)\right| \leq \frac{e^{qc(n)}}{N}\,\|f\|\,\sup_{x,y}|g(x)/g(y)|^3 \tag{27}$$

where

$$\forall(\mu,f)\in(\mathcal{P}(E_n^q)\times\mathcal{B}_b(E_n^q))\,,\quad \Psi_g(\mu)(f)=\mu(gf)/\mu(g)$$

In addition, if $(\mathcal{K})_\alpha$ *holds for some* $\alpha\in[1,\infty]$ *then we have the* $\mathbb{L}_\alpha(\eta_n^{\otimes q})$*-estimate*

$$\|\frac{d\mathbb{P}_n^{(N,q)}}{d\eta_n^{\otimes q}}-1\|_{\alpha,\eta_n^{\otimes q}}\leq\frac{\exp{(qc(n))}}{N}\times\|\sup_x\frac{dK_n(x,\,\boldsymbol{\cdot}\,)}{d\eta_n}\|_{\alpha,\eta_n}^q \tag{28}$$

The constant $c(n)$ *only depends on the parameter* n*. If* $(\mathcal{G})$ *and* $(\mathcal{K})_1$ *are both satisfied with (10) then the estimates (26), (27) and (28) are true for some constant* $c(n)=c$ *which does not depend on the the time parameter.*

Remarks 3.2: a) Observing that

$$\mathbb{E}(f(\xi_n^1,\ldots,\xi_n^q))-\eta_n^{\otimes q}(f)=\left(1-\frac{I(q)}{N^q}\right)\,\mathbb{E}(f(\xi_n^1,\ldots,\xi_n^q))+\mathbb{E}(m(\xi_n)^{\odot q}(f))-\eta_n^{\otimes q}(f)$$

and

$$1-\frac{|I(q)|}{N^q}\leq 1-\frac{N(N-1)\ldots(N-(q-1))}{N^q}\leq 1-(1-\frac{q-1}{N})^{q-1}\leq\frac{(q-1)^2}{N} \tag{29}$$

we can easily deduce (7) from (27).

b) We can reduce the proof of (26) and (27) to the case where $g=1$. Indeed, suppose we have proved (26) and (27) for $g=1$. Then using the decomposition

$$\Psi_g(m(\xi_n)^{\odot q})(f)-\Psi_g(\eta_n^{\otimes q})(f)=\frac{\eta_n^{\otimes q}(g)}{m(\xi_n)^{\odot q}(g)}\times m(\xi_n)^{\odot q}\left(\frac{g}{\eta_n^{\otimes q}(g)}\left(f-\Psi_g(\eta_n^{\otimes q})(f)\right)\right)$$

and noticing that

$$\eta_n^{\otimes q}\left(\frac{g}{\eta_n^{\otimes q}(g)}\left(f-\Psi_g(\eta_n^{\otimes q})(f)\right)\right)=0\quad\text{and}\quad\left|\frac{\eta_n^{\otimes q}(g)}{m(\xi_n)^{\odot q}(g)}\right|\leq 2\sup_{x,y}\left|\frac{g(x)}{g(y)}\right|$$

(since $\frac{N^q}{|I(q)|}\leq 2$ as soon as $\sqrt{N/2}\geq q$) one deduce (26) for any $g\in\mathcal{B}_b(E_n^q)$. To treat (27) we use the alternative decomposition

$$\Psi_g(m(\xi_n)^{\odot q})(f)-\Psi_g(\eta_n^{\otimes q})(f)=A_1+A_2+A_3$$

with

$$\begin{aligned}
A_1 &= m(\xi_n)^{\odot q}\left(\frac{gf}{\eta_n^{\otimes q}(g)}\right) - \eta_n^{\otimes q}\left(\frac{gf}{\eta_n^{\otimes q}(g)}\right), \quad A_2 = \Psi_g(\eta_n^{\otimes q})(f)\ \left(1 - \frac{m(\xi_n)^{\odot q}(g)}{\eta_n^{\otimes q}(g)}\right) \\
A_3 &= \left(\Psi_g(m(\xi_n)^{\odot q})(f) - \Psi_g(\eta_n^{\otimes q})(f)\right)\ \left(1 - \frac{m(\xi_n)^{\odot q}(g)}{\eta_n^{\otimes q}(g)}\right)
\end{aligned}$$

Assuming that (27) has been proved for $g = 1$ and (26) is true for any $g \in \mathcal{B}_b^+(E_n^q)$ one can check that

$$|\mathbb{E}(A_1)| + |\mathbb{E}(A_2)| \leq 2\ \frac{e^{qc(n)}}{N}\ \|f\|\ \sup_{x,y} |g(x)/g(y)|$$

and

$$\begin{aligned}
|\mathbb{E}(A_3)| &\leq \mathbb{E}\left(\left(\Psi_g(m(\xi_n)^{\odot q})(f) - \Psi_g(\eta_n^{\otimes q})(f)\right)^2\right)^{1/2} \mathbb{E}\left(\left(1 - \frac{m(\xi_n)^{\odot q}(g)}{\eta_n^{\otimes q}(g)}\right)^2\right)^{1/2} \\
&\leq \frac{e^{qc(n)}}{N}\ \|f\|\ \sup_{x,y} |g(x)/g(y)|^3
\end{aligned}$$

This clearly implies (27). From previous observations it suffices to check (26) and (27) for $g = 1$.

The proof of Theorem 3.1 is ultimately based on the observation that the sequence of q-tensor product measures $\{\eta_n^{\otimes q}\ ;\ n \in \mathbb{N}\}$ is solution of a measure valued dynamical system of the same form as in (2), namely

$$\forall n \in \mathbb{N}, \qquad \eta_{n+1}^{\otimes q} = \Phi_{n+1}^{(q)}(\eta_n^{\otimes q})$$

where $\Phi_{n+1}^{(q)} : \mathcal{P}(E_n^q) \longrightarrow \mathcal{P}(E_{n+1}^q)$ are given by

$$\forall n \in \mathbb{N}, \forall \eta \in \mathcal{P}(E_n^q), \forall f \in \mathcal{B}_b(E_{n+1}^q),$$

$$\Phi_{n+1}^{(q)}(\eta)(f) = \frac{\eta(\ g_n^{(q)}\ (K_{n+1}^{(q)} f)\ )}{\eta(\ g_n^{(q)}\ )},$$

with for any $(x_1, \ldots, x_q) \in E_n^q$ and $(y_1, \ldots, y_q) \in E_{n+1}^q$

$$g_n^{(q)}(x_1, \ldots, x_q) = g_n^{\otimes q}(x_1, \ldots, x_q) = g_n(x_1) \ldots g_n(x_q) \tag{30}$$

and

$$K_{n+1}^{(q)}((x_1, \ldots, x_q), d(y_1, \ldots, y_q)) = K_{n+1}(x_1, dy_1) \ldots K_{n+1}(x_q, dy_q) \tag{31}$$

This observation allows us to apply the arguments of [8] and conclude that the non linear semigroup formed by the composite mappings

$$\forall 0 \leq p \leq n, \qquad \Phi_{n|p}^{(q)} = \Phi_n^{(q)} \circ \Phi_{n-1}^{(q)} \circ \cdots \circ \Phi_{p+1}^{(q)}$$

(with convention $\Phi_{n|n}^{(q)} = Id$) have again the same form as the one step mappings $\{\Phi_n^{(q)}\ ;\ n \geq 1\}$.

Lemma 3.3 ([8]) *For any $0 \leq p \leq n$, $\eta \in \mathcal{P}(E_p^q)$ and $f \in \mathcal{B}_b(E_n^q)$ we have that*

$$\Phi_{n|p}^{(q)}(\eta)(f) = \frac{\eta(\ g_{n|p}^{(q)}\ (K_{n|p}^{(q)} f)\ )}{\eta(\ g_{n|p}^{(q)}\ )}, \tag{32}$$

The fitness functions $g_{n|p}^{(q)} : E_p^q \to]0, \infty[$ and the Markov transitions $K_{n|p}^{(q)}$ (from E_p^q into E_n^q) satisfy the backward formulae

$$\forall 1 \leq p \leq n, \quad g_{n|p-1}^{(q)} = g_{p-1}^{(q)}\ K_p^{(q)}(g_{n|p}^{(q)}), \quad \text{and} \quad K_{n|p-1}^{(q)} = S_{n|p}^{(q)} K_{n|p}^{(q)} \tag{33}$$

with for any $f \in \mathcal{B}_b(E_p^q)$

$$S_{n|p}^{(q)}(f) = \frac{K_p^{(q)}(\ g_{n|p}^{(q)}\ f\ )}{K_p^{(q)}(\ g_{n|p}^{(q)}\ )}$$

and conventions $g_{n|n}^{(q)} = 1$ and $K_{n|n}^{(q)} = Id$.

As in [8] we will now use the decomposition of errors

$$\forall f \in \mathcal{B}_b(E_n^q)\,,$$

$$m(\xi_n)^{\odot q}(f) - \eta_n^{\otimes q}(f) = \sum_{p=0}^{n} \left[\Phi_{n|p}^{(q)}(m(\xi_p)^{\odot q})(f) - \Phi_{n|p}^{(q)}(\Phi_p^{(q)}(m(\xi_{p-1})^{\odot q}))(f)\right] \tag{34}$$

with convention for $p = 0$, $\Phi_0^{(q)}(m(\xi_{-1})^{\odot q}) = \eta_0^{\otimes q}$. Since

$$\Phi_p^{(q)}(m(\xi_{p-1})^{\otimes q}) = \Phi_p(m(\xi_{p-1}))^{\otimes q}$$

Formula (34) can also be written

$$m(\xi_n)^{\odot q}(f) - \eta_n^{\otimes q}(f) = J_1(f) + J_2(f)$$

with

$$J_1(f) = \sum_{p=0}^{n} \left[\Phi_{n|p}^{(q)}(m(\xi_p)^{\odot q})(f) - \Phi_{n|p}^{(q)}(\Phi_p(m(\xi_{p-1}))^{\otimes q})(f)\right]$$

and

$$J_2(f) = \sum_{p=0}^{n-1} \left[\Phi_{n|p}^{(q)}(m(\xi_p)^{\otimes q})(f) - \Phi_{n|p}^{(q)}(m(\xi_p)^{\odot q}))(f)\right]$$

Next lemma is instrumental for the proof of Theorem 3.1. Its proof follows the proof of the theorem.

Lemma 3.4 *Let $(E, \mathcal{E})$ be a measurable space and let $q, N \geq 1$. For any N-vector $x = (x^1, \ldots, x^N) \in E^N$, $\sqrt{N/2} \geq q \geq 1$ and $(f, g) \in (\mathcal{B}_b(E^q) \times \mathcal{B}_b^+(E^q))$ we have*

$$\|m(x)^{\odot q} - m(x)^{\otimes q}\|_{\text{tv}} \ \leq\ \frac{(q-1)^2}{N} \tag{35}$$

$$\left|\Psi_g(m(x)^{\odot q})(f) - \Psi_g(m(x)^{\otimes q})(f)\right| \ \leq\ \frac{2\ (q-1)^2}{N}\ \sup_{x,y}\left|\frac{g(x)}{g(y)}\right|^2\ \sup_{x,y}|f(x) - f(y)| \tag{36}$$

$$\text{where} \qquad \forall (\mu, f) \in (\mathcal{P}(E^q) \times \mathcal{B}_b(E^q)), \quad \Psi_g(\mu)(f) = \mu(gf)/\mu(g)$$

If $X = (X^1, \ldots, X^N)$ *are i.i.d. random variables with common law* $\eta \in \mathcal{P}(E)$ *then for any* $\sqrt{N/2} \geq q \geq 1$

$$\mathbb{E}\left(\left(\Psi_g(m(X)^{\odot q})(f) - \Psi_g(\eta^{\otimes q})(f)\right)^2\right)^{\frac{1}{2}} \leq \frac{q\,C}{\sqrt{N}} \sup_{x,y} \left|\frac{g(x)}{g(y)}\right|^2 \sup_{x,y} |f(x) - f(y)| \tag{37}$$

$$\left|\mathbb{E}\left(\Psi_g(m(X)^{\odot q})(f)\right) - \Psi_g(\eta^{\otimes q})(f)\right| \leq \frac{q^2\,C}{N} \sup_{x,y} \left|\frac{g(x)}{g(y)}\right|^3 \sup_{x,y} |f(x) - f(y)| \tag{38}$$

for some universal and finite constant C*.*

Proof of Theorem 3.1: In view of remark 3.2 page 16 it is enough to prove (26) and (27) for $g = 1$. We start proving (27). By (38) and (32) we easily that for any $\sqrt{N/2} \geq q \geq 1$

$$\begin{aligned}
|\mathbb{E}(J_1(f))| &\leq \sum_{p=0}^{n} \mathbb{E}\left(\left|\mathbb{E}\left(\Phi^{(q)}_{n|p}(m(\xi_p)^{\odot q})(f)\,|\xi_{p-1}\right) - \Phi^{(q)}_{n|p}(\Phi_p(m(\xi_{p-1}))^{\otimes q})(f)\right|\right) \\
&\leq \frac{q^2\ C}{N} \sum_{p=0}^{n} \sup_{x,y} \left|\frac{g^{(q)}_{n|p}(x)}{g^{(q)}_{n|p}(y)}\right|^3 \sup_{x,y} \left|K^{(q)}_{n|p}(f)(x) - K^{(q)}_{n|p}(f)(y)\right| \\
&\leq \frac{q^2\ C}{N} \|f\| \sum_{p=0}^{n} \sup_{x,y} \left|\frac{g^{(q)}_{n|p}(x)}{g^{(q)}_{n|p}(y)}\right|^3 \sup_{x,y} \left\|K^{(q)}_{n|p}(x,.) - K^{(q)}_{n|p}(y,.)\right\|_{\rm tv}
\end{aligned}$$

for some universal constant $C < \infty$. In much the same way (36) yields

$$|J_2(f)| \leq \frac{2\ q^2}{N} \|f\| \sum_{p=0}^{n-1} \sup_{x,y} \left|\frac{g^{(q)}_{n|p}(x)}{g^{(q)}_{n|p}(y)}\right|^2 \sup_{x,y} \left\|K^{(q)}_{n|p}(x,.) - K^{(q)}_{n|p}(y,.)\right\|_{\rm tv}$$

By the product definitions (30) and (31) of $g^{(q)}_n$ and $K^{(q)}_n$ a clear backward inductive proof gives that

$$g^{(q)}_{n|p}(x_1, \ldots, x_q) = g^{(1)}_{n|p}(x_1) \ldots g^{(1)}_{n|p}(x_q)$$

and

$$K^{(q)}_{n|p}((x_1, \ldots, x_q), d(y_1, \ldots, y_q)) = K^{(1)}_{n|p}(x_1, dy_1) \ldots K^{(1)}_{n|p}(x_q, dy_q)$$

It is now clear that for any $\sqrt{N/2} \geq q$

$$\left|\mathbb{E}(m(\xi_n)^{\odot q}(f)) - \eta_n^{\otimes q}(f)\right|$$

is bounded by

$$\frac{q^3\ C'}{N} \|f\| \sum_{p=0}^{n} \sup_{x,y} \left|\frac{g^{(1)}_{n|p}(x)}{g^{(1)}_{n|p}(y)}\right|^{3q} \sup_{x,y} \left\|K^{(1)}_{n|p}(x,.) - K^{(1)}_{n|p}(y,.)\right\|_{\rm tv}$$

for some universal and finite constant C' and the proof of (27) is now straightforward.

The proof of (26) is a repeat of arguments used in the proof of (27). It is therefore only sketched. We use again decomposition (34) to check that

$$\mathbb{E}\left(|m(\xi_n)^{\odot q}(f) - \eta_n^{\otimes q}(f)|^2\right)^{\frac{1}{2}} \leq J_1'(f) + J_2'(f)$$

with

$$\begin{aligned} J_1'(f) &= \sum_{p=0}^{n} \mathbb{E}\left(\left[\Phi^{(q)}_{n|p}(m(\xi_p)^{\odot q})(f) - \Phi^{(q)}_{n|p}(\Phi_p(m(\xi_{p-1}))^{\otimes q})(f)\right]^2\right)^{\frac{1}{2}} \\ J_2'(f) &= \sum_{p=0}^{n-1} \mathbb{E}\left[|\Phi^{(q)}_{n|p}(m(\xi_p)^{\otimes q})(f) - \Phi^{(q)}_{n|p}(m(\xi_p)^{\odot q}))(f)|\right] \end{aligned}$$

Then we use (36) and (37) to check that the sum $J_1'(f) + J_2'(f)$ is bounded by

$$\frac{Cq^2}{\sqrt{N}} \;\|f\| \sum_{p=0}^{n} \sup_{x,y} \left|\frac{g^{(1)}_{n|p}(x)}{g^{(1)}_{n|p}(y)}\right|^{2q} \quad \sup_{x,y} \left\|K^{(1)}_{n|p}(x,\,.\,) - K^{(1)}_{n|p}(y,\,.\,)\right\|_{\mathrm{tv}}$$

for some universal and finite constant C'.

In the case when $(\mathcal{G})$ and $(\mathcal{K})_1$ hold with

$$M(g) = \sup_{n,x,y} \left|\log \frac{g_n(x)}{g_n(y)}\right| < \infty \quad \text{and} \quad \frac{M(K)}{2} = \sup_{n,x,y} \left|\log \frac{dK_n(x,\,.\,)}{d\eta_n}(y)\right| < \infty$$

we can use the backward formula (33) to check that

$$\forall (x,y) \in E_p^2\,, \qquad \left|\frac{g^{(1)}_{n|p}(x)}{g^{(1)}_{n|p}(y)}\right| = \left|\frac{g_p(x)}{g_p(y)}\right| \left|\frac{K_{p+1}(g^{(1)}_{n|p+1}(x))}{K_{p+1}(g^{(1)}_{n|p+1}(y))}\right| \leq \exp M(g,K)$$

with

$$M(g,K) = M(g) + M(K)$$

Moreover using the same line of arguments as [8] we also have

$$\sup_{x,y} \left\|K^{(1)}_{n|p}(x,\,.\,) - K^{(1)}_{n|p}(y,\,.\,)\right\|_{\mathrm{tv}} \leq \prod_{k=1}^{n-p} \left(1 - \alpha(S^{(1)}_{n|p+k})\right)$$

where $\alpha(S)$ is the Dobrushin ergodic coefficient of a Markov transition S on a measurable state space E defined as

$$\alpha(S) = \inf \sum_{i=1}^{m} \min\left(S(x, A_i), S(z, A_i)\right)$$

where the infimum is taken over all x, z and all resolutions of the state space into pairs of non-intersecting subsets $\{A_i\;;\; 1 \leq i \leq m\}$ and $m \geq 1$.

Since for any $0 \leq p \leq n$, $x \in E_{p-1}$ and $A \in \mathcal{E}_p$

$$S^{(1)}_{n|p}(x, A) = \frac{K_p(\;g^{(1)}_{n|p}\;1_A\;)(x)}{K_p(\;g^{(1)}_{n|p}\;)(x)} \geq e^{-M(K)} \;\frac{\eta_p(\;g^{(1)}_{n|p}\;1_A\;)}{\eta_p(\;g^{(1)}_{n|p}\;)}$$

we have $\alpha(S^{(1)}_{n|p}) \geq e^{-M(K)}$. This implies that

$$\sup_{x,y} \left\| K^{(1)}_{n|p}(x, .) - K^{(1)}_{n|p}(y, .) \right\|_{\text{tv}} \leq \ \left(1 - e^{-M(K)}\right)^{n-p}$$

If we combine these two estimates we conclude that

$$\mathbb{E}\left(|m(\xi_n)^{\odot q}(f) - \eta_n^{\otimes q}(f)|^2\right)^{\frac{1}{2}} \leq \frac{q^2\ C'}{\sqrt{N}}\ \|f\|\ \frac{\exp\left(2q\ M(g, K)\right)}{1 - \exp -M(K)}$$

and

$$\left|\mathbb{E}(m(\xi_n)^{\odot q}(f)) - \eta_n^{\otimes q}(f)\right| \leq \frac{q^3\ C'}{N}\ \|f\|\ \frac{\exp\left(3q\ M(g, K)\right)}{1 - \exp -M(K)}$$

According to a) remark 3.2 this ends the proof of (26) and (27).

We now turn to the proof of (28). Under $(\mathcal{K})_1$ we use the representation

$\forall (y_1, \ldots, y_q) \in E_n^q$,

$$\begin{aligned}
\frac{d\mathbb{P}_n^{(N,q)}}{d\eta_n^{\otimes q}}(y_1, \ldots, y_q) &= \mathbb{E}\left(\frac{d\Phi_n(m(\xi_{n-1}))^{\otimes q}}{d\eta_n^{\otimes q}}(y_1, \ldots, y_q)\right) \\
&= \mathbb{E}\left(\frac{d\Phi_n^{(q)}(m(\xi_{n-1})^{\otimes q})}{d\Phi_n^{(q)}(\eta_{n-1}^{\otimes q})}(y_1, \ldots, y_q)\right)
\end{aligned}$$

Again we observe that

$$\frac{d\Phi_n^{(q)}(m(\xi_{n-1})^{\otimes q})}{d\Phi_n^{(q)}(\eta_{n-1}^{\otimes q})}(y_1, \ldots, y_q) = \frac{m(\xi_{n-1})^{\otimes q}(g_{n-1}^{(q)}\, k_n^{(q)}(., (y_1, \ldots, y_q)))}{m(\xi_{n-1})^{\otimes q}(g_{n-1}^{(q)})}$$

with

$$\begin{aligned}
k_n^{(q)}((x_1, \ldots, x_q), (y_1, \ldots, y_q)) &= \frac{dK_n^{(q)}((x_1, \ldots, x_q), .)}{d\eta_n^{\otimes q}}(y_1, \ldots, y_q) \\
&= \frac{dK_n(x_1, .)}{d\eta_n}(y_1) \ldots \frac{dK_n(x_q, .)}{d\eta_n}(y_q) \\
&= k_n(x_1, y_1) \ldots k_n(x_q, y_q)
\end{aligned}$$

Using (27) and (36) one gets after some tedious but easy calculations

$$\left|\frac{d\mathbb{P}_n^{(N,q)}}{d\eta_n^{\otimes q}}(y_1, \ldots, y_q) - 1\right| \leq \frac{e^{q\, c'(n)}}{N}\ \sup_{x,y}\left|\frac{g_{n-1}(x)}{g_{n-1}(y)}\right|^{3q}\ \prod_{p=1}^{q}\ \sup_{x_p} k_n(x_p, y_p)$$

for some constant $c'(n)$ which has the boundness properties stated in the theorem. This clearly implies that

$$\forall \alpha \in [1, \infty]\,, \qquad \left\|\frac{d\mathbb{P}_n^{(N,q)}}{d\eta_n^{\otimes q}} - 1\right\|_{\alpha, \eta_n^{\otimes q}} \leq \frac{e^{q\, c'(n)}}{N}\ \sup_{x,y}\left|\frac{g_{n-1}(x)}{g_{n-1}(y)}\right|^{3q}\ \left\|\sup_{x} k_n(x, .)\right\|_{\alpha, \eta_n}^{q}$$

This ends the proofs of the theorem. ∎

The proof of Lemma 3.4 is now completed.

Proof of Lemma 3.4: Since

$$|m(x)^{\otimes q}(f) - m(x)^{\odot q}(f)| \leq \left(1 - \frac{|I(q)|}{N^q}\right) \|f\|$$

the proof of (35) is a clear consequence of (29). To prove (36) we rewrite

$$\Psi_g(m(x)^{\odot q})(f) - \Psi_g(m(x)^{\otimes q})(f)$$

as

$$\left(\frac{m(x)^{\otimes q}(g)}{m(x)^{\odot q}(g)}\right) \quad m(x)^{\odot q}\left[\frac{g}{m(x)^{\otimes q}(g)} \ \left(f - \Psi_g(m(x)^{\otimes q})(f)\right)\right]$$

Using the fact that

$$\frac{N^q}{|I(q)|} \leq (1 - \frac{q^2}{N})^{-1} \leq 2$$

for any $\sqrt{N/2} \geq q$ one gets

$$\left|\frac{m(x)^{\otimes q}(g)}{m(x)^{\odot q}(g)}\right| \leq 2 \sup_{x,y}\left|\frac{g(x)}{g(y)}\right|$$

Since

$$f(u) - \Psi_g(m(x)^{\otimes q})(f) = \int (f(u) - f(v)) \ \frac{g(v)}{m(x)^{\otimes q}(g)} \ m(x)^{\otimes q}(dv)$$

one easily concludes that

$$\left|\Psi_g(m(x)^{\odot q})(f) - \Psi_g(m(x)^{\otimes q})(f)\right|$$

is bounded by

$$2 \sup_{x,y}\left|\frac{g(x)}{g(y)}\right|^2 \ \sup_{x,y}|f(x) - f(y)| \ \|m(x)^{\odot q} - m(x)^{\otimes q}\|_{\rm tv}$$

and the proof of (36) is now a consequence of (35). The proof of (38) is a little more involved. We start by noting that

$$\eta^{\otimes q}(f) - \mathbb{E}(m(X)^{\odot q}(f)) = \left(1 - \frac{|I(q)|}{N^q}\right) \ \eta^{\otimes q}(f)$$

Using (35) the above decomposition yields

$$\left|\mathbb{E}(m(X)^{\otimes q}(f)) - \eta^{\otimes q}(f)\right| \leq 2 \ \frac{(q-1)^2}{N} \ \|f\|$$

Replacing q and f by $(q+q)$ and $(f \otimes f)$ we arrive at

$$\left|\mathbb{E}(\left[m(X)^{\otimes q}(f)\right]^2) - \left[\eta^{\otimes q}(f)\right]^2\right| \leq 8 \ \frac{q^2}{N} \ \|f\|^2$$

Combining the preceding two inequalities yields

$$\mathbb{E}([m(X)^{\otimes q}(f)-\eta^{\otimes q}(f)]^2)$$

$$=\mathbb{E}([m(X)^{\otimes q}(f)]^2)-[\eta^{\otimes q}(f)]^2+2\eta^{\otimes q}(f)\ \left(\eta^{\otimes q}(f)-\mathbb{E}(m(X)^{\otimes q}(f))\right)\leq \tfrac{12q^2}{N}\ \|f\|^2$$

and therefore

$$\begin{aligned}
\mathbb{E}(\left[m(X)^{\otimes q}(f)-\eta^{\otimes q}(f)\right]^2)^{\frac{1}{2}} &\leq \frac{4q}{\sqrt{N}}\ \|f\| \\
\mathbb{E}(\left[m(X)^{\odot q}(f)-\eta^{\otimes q}(f)\right]^2)^{\frac{1}{2}} &\leq [\frac{4q}{\sqrt{N}}+\frac{q^2}{N}]\ \|f\|\leq \frac{8q}{\sqrt{N}}\ \|f\|
\end{aligned}\tag{39}$$

as soon as $\sqrt{N}\geq\sqrt{2}\ q\ (\geq\ q/4)$. Now we use again the decomposition

$$\Psi_g(m(X)^{\odot q})(f)-\Psi_g(\eta^{\otimes q})(f)$$

$$=\left(\frac{\eta^{\otimes q}(g)}{m(X)^{\odot q}(g)}\right)\quad m(X)^{\odot q}\left[\ \frac{g}{\eta^{\otimes q}(g)}\ \left(f-\Psi_g(\eta^{\otimes q})(f)\right)\right]$$

As earlier we observe that

$$\left|\frac{\eta^{\otimes q}(g)}{m(X)^{\odot q}(g)}\right|\leq 2\ \sup_{x,y}\left|\frac{g(x)}{g(y)}\right|$$

and

$$f(u)-\Psi_g(\eta^{\otimes q})(f)=\int\ (f(u)-f(v))\ \frac{g(v)}{\eta^{\otimes q}(g)}\ \eta^{\otimes q}(dv)$$

We conclude that for any $\sqrt{N/2}\geq q$

$$\mathbb{E}\left(\left|\Psi_g(m(X)^{\odot q})(f)-\Psi_g(\eta^{\otimes q})(f)\right|^2\right)^{\frac{1}{2}}$$

is bounded by

$$2\ \sup_{x,y}\left|\frac{g(x)}{g(y)}\right|\ \mathbb{E}\left(\left|m(X)^{\odot q}\left[\ \frac{g}{\eta^{\otimes q}(g)}\ \left(f-\Psi_g(\eta^{\otimes q})(f)\right)\right]\right|^2\right)^{\frac{1}{2}}$$

Since

$$\eta^{\otimes q}\left[\ \frac{g}{\eta^{\otimes q}(g)}\ \left(f-\Psi_g(\eta^{\otimes q})(f)\right)\right]=0$$

it does follows from (39) that

$$\mathbb{E}\left(\left|\Psi_g(m(X)^{\odot q})(f)-\Psi_g(\eta^{\otimes q})(f)\right|^2\right)^{\frac{1}{2}}\leq 16\ \frac{q}{\sqrt{N}}\ \sup_{x,y}\left|\frac{g(x)}{g(y)}\right|^2\ \sup_{x,y}|f(x)-f(y)|$$

We end the proof of (38) by noting that

$$\mathbb{E}\left(\Psi_g(m(X)^{\odot q})(f)\right)-\Psi_g(\eta^{\otimes q})(f)$$

$$=\mathbb{E}\left(\left[\Psi_g(m(X)^{\odot q})(f)-\Psi_g(\eta^{\otimes q})(f)\right]\left[1-\tfrac{m(X)^{\odot q}(g)}{\eta^{\otimes q}(g)}\right]\right)$$

and using Cauchy-Schwartz's inequality.

■

3.2 Entropy Estimates on Path Space

The study of the asymptotic behavior of distributions $\{\mathbb{P}^{(N,q)}_{[0,n]}\ ;\ n \geq 0\}$ as $N \to \infty$ becomes more transparent if we introduce a more abstract formulation.

In the further development we assume that $\{\xi_n\ ;\ n \in \mathbb{N}\}$ is the N-interacting particle systems approximating model associated to a given sequence of functions $\{\Phi_n\ ;\ n \in \mathbb{N}\}$. We denote as usual $\{\eta_n\ ;\ n \in \mathbb{N}\}$ the solution of the corresponding limiting system.

We will use in our development the following condition.

$(\mathcal{P})$ **For any $n \in \mathbb{N}$ and $\eta \in \mathcal{P}(E_n)$ we have $\Phi_{n+1}(\eta) \sim \eta_{n+1}$.**

Note that when $\{\Phi_n\ ;\ n \in \mathbb{N}\}$ are given by (3) condition $(\mathcal{P})$ holds if, and only if, for any $n \in \mathbb{N}$ and $x \in E_n$ we have $(\Phi_{n+1}(\delta_x) =)K_{n+1}(x, \boldsymbol{.}) \sim \eta_{n+1}$. This condition is not satisfied when K_n is the transition (13) of the path-valued Markov chain (12) but it is in many cases fulfilled when K_n represent the signal transition in classical non linear filtering problems (see examples given in section 2.4).

Assumption $(\mathcal{P})$ first appeared in [5]. In this work the authors prove large deviations principles for the laws of the empirical measures on path space. A crucial practical advantage of $(\mathcal{P})$ is that the distributions of the particles in path space can be regarded as a mean field Gibbs measure with unit partition function. The approach taken here to obtain useful entropy estimates and related increasing propagation of chaos is based on the similar ideas.

The following assumption on the N-interacting particle systems approximating model summarizes the only $\mathbb{L}_\alpha$-estimates on the particle density profiles $\{m(\xi_n)\ ;\ n \in \mathbb{N}\}$ needed in the sequel.

$(\mathcal{I})$ **There exists some $\alpha \geq 1$ and a collection of functions $\theta_{\alpha,n} \in \mathbb{L}_\alpha(\eta_n)$, $n \geq 1$ such that**

$$\forall n \geq 1\,, \forall y \in E_n\,, \qquad \mathbb{E}\left(\left|\frac{d\Phi_n(m(\xi_{n-1}))}{d\Phi_n(\eta_{n-1})}(y) - 1\right|^\alpha\right)^{\frac{1}{\alpha}} \leq \frac{\theta_{\alpha,n}(y)}{\sqrt{N}}$$

Conditions $(\mathcal{P})$ and $(\mathcal{I})$ also appear to be useful to study the weak convergence of the empirical processes associated to the N-particle approximating model. Although this subject is tangential to this paper we have chosen to present this result. The results for the genetic algorithm (5) will then be deduced directly from these results.

To state the main result of this section we need to introduce some additional notations. If $\mathcal{F}$ is a collection of bounded measurable functions f, $\|f\| \leq 1$, on a measurable space $(E, \mathcal{E})$ and $\mu \in \mathcal{P}(E)$ and $\epsilon > 0$ we denote $N(\epsilon, \mathcal{F}, \mathbb{L}_2(\mu))$ the minimal number of $\mathbb{L}_2(\mu)$-balls of radius ϵ needed to cover $\mathcal{F}$. We also denote by $\mathcal{N}(\varepsilon, \mathcal{F})$ and by $I(\mathcal{F})$ the uniform covering numbers and entropy integral given by

$$\mathcal{N}(\epsilon, \mathcal{F}) = \sup\{N(\epsilon, \mathcal{F}, \mathbb{L}_2(\mu));\ \mu \in \mathcal{P}(E)\} \quad \text{and} \quad I(\mathcal{F}) = \int_0^1 \sqrt{\log \mathcal{N}(\varepsilon, \mathcal{F})}\, d\epsilon$$

For any $n \geq 0$ and for any collection $\mathcal{F}_n$ of bounded measurable functions $f : E_n \to \mathbb{R}$, $\|f\| \leq 1$, we denote by $\{m(\xi_n)(f) \ ; \ f \in \mathcal{F}_n\}$ the $\mathcal{F}_n$-empirical process associated to $\mathcal{F}_n$ and the particle density profile $m(\xi_n)$. The Zolotarev $\mathcal{F}_n$-semi-distance between $m(\xi_n)$ and its limiting distribution η_n is defined by

$$\|m(\xi_n) - \eta_n\|_{\mathcal{F}_n} = \sup\{|m(\xi_n)(f) - \eta_n(f)| \ ; \ f \in \mathcal{F}_n\}$$

Theorem 3.5 *For each $n \in \mathbb{N}$, let $\mathcal{F}_n$ be a collection of bounded measurable functions $f : E_n \to \mathbb{R}$ such that $\|f\| \leq 1$. If the functions $\{\Phi_n \ ; \ n \in \mathbb{N}\}$ satisfy conditions $(\mathcal{P})$ and $(\mathcal{I})$ for some number $\alpha \geq 1$ then there exists some universal constant $A_\alpha < \infty$ such that*

$$\forall n \in \mathbb{N}\,, \forall N \geq 1\,, \quad \mathbb{E}\left(\|m(\xi_n) - \eta_n\|^\alpha_{\mathcal{F}_n}\right)^{\frac{1}{\alpha}} \leq \frac{A_\alpha}{\sqrt{N}} \ \left(I(\mathcal{F}_n) + \|\theta_{\alpha,n}\|_{\alpha,\eta_n}\right) \tag{40}$$

In addition, if $(\mathcal{I})$ holds for $\alpha = 2$ then we have that

$$\forall n \in \mathbb{N}\,, \forall 1 \leq q \leq N\,, \quad \mathrm{Ent}(\mathbb{P}^{(N,q)}_{[0,n]} \mid \eta^{\otimes q}_{[0,n]}) \leq \frac{2\ q}{N} \sum_{p=1}^{n} \ \|\theta_{2,n}\|^2_{2,\eta_n} \tag{41}$$

The proof of (40) is very simple therefore we give it first. Using the decomposition

$$\begin{aligned}
&m(\xi_n)(f) - \eta_n(f) \\
&= m(\xi_n)(f) - \Phi_n(m(\xi_{n-1}))(f) + \Phi_n(m(\xi_{n-1}))(f) - \Phi_n(\eta_{n-1})(f) \\
&= m(\xi_n)(f) - \Phi_n(m(\xi_{n-1}))(f) + \int \ f(y)\ [\frac{d\Phi_n(m(\xi_{n-1}))}{d\Phi_n(\eta_{n-1})}(y) - 1]\ \eta_n(dy)
\end{aligned}$$

we clearly have

$$\|m(\xi_n) - \eta_n\|_{\mathcal{F}_n} \leq \|m(\xi_n) - \Phi_n(m(\xi_{n-1}))\|_{\mathcal{F}_n} + \int \ |\frac{d\Phi_n(m(\xi_{n-1}))}{d\Phi_n(\eta_{n-1})}(y) - 1|\ \eta_n(dy)$$

Under our assumptions and using Marcinkiewicz-Zygmund's inequality to empirical processes (see [10] and Lemma 2.10 in [8]) we conclude that

$$\begin{aligned}
\mathbb{E}\left(\|m(\xi_n) - \eta_n\|^\alpha_{\mathcal{F}_n}\right)^{\frac{1}{\alpha}} &\leq \frac{B_\alpha}{\sqrt{N}}\ I(\mathcal{F}_n) + [\int \ \mathbb{E}(|\frac{d\Phi_n(m(\xi_{n-1}))}{d\Phi_n(\eta_{n-1})}(y) - 1|^\alpha)\ \eta_n(dy)]^{\frac{1}{\alpha}} \\
&\leq \frac{B_\alpha}{\sqrt{N}}\ I(\mathcal{F}_n) + \frac{1}{\sqrt{N}}\ [\int \ \theta^\alpha_{\alpha,n}\ d\eta_n]^{\frac{1}{\alpha}}
\end{aligned}$$

for some universal constant B_α which only depends on the parameter α. This clearly ends the proof of (40).

The weak convergence of empirical processes for the genetic approximating model (5) has been started in [7] and it is been further developed in [8]. In [8] the authors proved $\mathbb{L}_\alpha$-mean error estimates of the same type using different techniques without any assumptions on the mutation transitions. Here we present an alternative and more simple proof based on the mixing and $\mathbb{L}_\alpha$-conditions $(\mathcal{F})_\alpha$. Next analysis emphasizes the new points in which this approach differs from the one in [7, 8] with an eye toward precise uniform bounds w.r.t. the time parameter.

A basic result giving $\mathbb{L}_\alpha$-mean error bounds in terms of the fitness functions $\{g^{(1)}_{n|p}\ ;\ 0\leq p\leq n\}$ and the Markov transitions $\{K^{(1)}_{n|p}\ ;\ 0\leq p\leq n\}$ introduced in Lemma 3.3 is as follows.

Lemma 3.6 ([8],p.36) *If the functions $\{\Phi_n\ ;\ n\geq 1\}$ are given by (3) and condition $(\mathcal{G})$ holds then we have*

$$\forall n\geq 0\,,\forall f\in\mathcal{B}_b(E_n)\,,\forall\alpha\geq 1\,,\forall N\geq 1,$$

$$\mathbb{E}\left(|\eta^N_n(f)-\eta_n(f)|^\alpha\right)^{\frac{1}{\alpha}}\leq\frac{B_{\alpha,n}}{\sqrt{N}}\ \|f\|$$

where

$$B_{\alpha,n}=B_{\alpha,0}\ \sum_{p=0}^{n}\ \sup_{x,x'}\left|\frac{g^{(1)}_{n|p}(x)}{g^{(1)}_{n|p}(x')}\right|^2\ \sup_{x,x'}\left\|K^{(1)}_{n|p}(x,\,.\,)-K^{(1)}_{n|p}(x',\,.\,)\right\|_{\rm tv}$$

and $B_{\alpha,0}$ is a universal constant which only depends on the parameter α.

Remark 3.7: The assumptions on the preceding lemma are remarkably weak since they does not depend on the form of the mutation transitions. In the case where $(\mathcal{G})$ and $(\mathcal{K})_1$ are both satisfied with

$$M(g)=\sup_{n,x,y}\left|\log\frac{g_n(x)}{g_n(y)}\right|<\infty\quad\text{and}\quad\frac{M(K)}{2}=\sup_{n,x,y}\left|\log\frac{dK_n(x,\,.\,)}{d\eta_n}(y)\right|<\infty\qquad(42)$$

we proceed as in the proof of Theorem 3.1 to check that for any $\alpha\geq 1$ the sequence $\{B_{\alpha,n}\ ;\ n\in\mathbb{N}\}$ is uniformly bounded. More precisely in this situation we have that

$$\forall\alpha\geq 1\,,\qquad\sup_{n\geq 0}B_{\alpha,n}\leq B_{\alpha,0}\ \times\ \frac{\exp\left(2M(g,K)\right)}{1-\exp-M(K)}$$

Using the decomposition

$$\frac{d\Phi_{n+1}(m(\xi_n))}{d\Phi_{n+1}(\eta_n)}(y)-1=\frac{\eta_n(g_n)}{m(\xi_n)(g_n)}\ \times\ m(\xi_n)\left(\frac{g_n}{\eta_n(g_n)}\ \ (k_{n+1}(\,.\,,y)-1)\right)$$

with

$$\forall(x,y)\in(E_n\times E_{n+1})\,,\quad k_{n+1}(x,y)=\frac{dK_{n+1}(x,\,.\,)}{d\eta_{n+1}}(y)$$

it is easily checked that

$$|\frac{d\Phi_{n+1}(m(\xi_n))}{d\Phi_{n+1}(\eta_n)}(y)-1|\leq\sup_{x,x'}\left|\frac{g_n(x)}{g_n(x')}\right|\quad|m(\xi_n)(f^y_n)-\eta_n(f^y_n)|$$

with

$$\forall(x,y)\in(E_n\times E_{n+1})\,,\quad f^y_n(x)=\frac{g_n(x)}{\eta_n(g_n)}\ \ (k_{n+1}(x,y)-1)$$

Using Lemma 3.6 condition $(\mathcal{I})$ clearly holds for some $\alpha\geq 1$ and

$$\theta_{\alpha,n}(y)=B_{\alpha,n}\ \sup_{x,x'}\left|\frac{g_n(x)}{g_n(x')}\right|^2\ \ \left(1+\sup_x k_{n+1}(x,y)\right)\qquad(43)$$

as soon as $(K)_\alpha$ is met. Arguing as before if $(\mathcal{G})$ and $(\mathcal{K})_1$ are both satisfied with (42) one also concludes that

$$\sup_{n\geq 0} \|\theta_{\alpha,n}\|_{\alpha,\eta_n} \leq B_{\alpha,0} \ \times \ \frac{\exp\left(4M(g,K)\right)}{1-\exp -M(K)}$$

As a direct consequence of (40) and Lemma 3.6 we have.

Corollary 3.8 *For each $n \in \mathbb{N}$, let $\mathcal{F}_n$ be a collection of bounded measurable functions $f : E_n \to \mathbb{R}$ such that $\|f\| \leq 1$. Assume that the functions $\{\Phi_n \ ; \ n \geq 1\}$ are given by (3) and conditions $(\mathcal{G})$ and $(\mathcal{K})_\alpha$ are both satisfied for some $\alpha \geq 1$. Then there exists a universal constant c_α such that*

$$\forall n \in \mathbb{N}\,, \forall N \geq 1\,, \quad \mathbb{E}\left(\|m(\xi_n)-\eta_n\|^{\alpha}_{\mathcal{F}_n}\right)^{\frac{1}{\alpha}} \leq \frac{c_\alpha}{\sqrt{N}} \ \left(I(\mathcal{F}_n)+c'_n\right)$$

where c'_n is a finite constant which only depends on the time parameter parameter n and such that $\sup_n c'_n < \infty$ as soon as the uniform bounds (42) are satisfied.

As we said in the introduction the proof of the entropy estimates (41) on path space relies on the fact that $\mathbb{P}^{(N)}_{[0,n]}$ is a mean field Gibbs measure. To describe the potential function we notice that under $(\mathcal{P})$ we have for any $F \in \mathcal{B}_b(E_0^N \times \ldots \times E_n^N)$

$$\begin{aligned}
&\mathbb{E}(F(\xi_0,\ldots,\xi_n)) \\
&= \int \ F(x_0,\ldots,x_n) \ \Phi_n(m(x_{n-1}))^{\otimes N}(dx_n) \ \ldots \ \Phi_1(m(x_0))^{\otimes N}(dx_1) \ \eta_0^{\otimes N}(dx_0) \\
&= \int \ F(x_0,\ldots,x_n) \ \exp\left(H_n^{(N)}(x_0,\ldots,x_n)\right) \eta_n^{\otimes N}(dx_n) \ \ldots \ \eta_0^{\otimes N}(dx_0)
\end{aligned}$$

with

$$H_n^{(N)}(x_0,\ldots,x_n) = N\sum_{p=1}^{n} \ \int \ \log\frac{d\Phi_p(m(x_{p-1}))}{d\Phi_p(\eta_{p-1})} \ dm(x_p)$$

In other words $\mathbb{P}^{(N)}_{[0,n]}$ is absolutely continuous with respect to the tensor product measure

$$\eta^{\otimes N}_{[0,n]} = \eta_0^{\otimes N} \otimes \ldots \otimes \eta_n^{\otimes N} \in \mathcal{P}(E_0^N \times \ldots \times E_n^N)$$

and

$$\frac{d\mathbb{P}^{(N)}_{[0,n]}}{d\eta^{\otimes N}_{[0,n]}} = \exp H_n^{(N)} \qquad \eta^{\otimes N}_{[0,n]} - \text{a.e.}$$

In addition if we consider the function

$$\begin{array}{rcl}
\Phi_{[0,n]} : \mathcal{P}(E_0 \times \ldots \times E_n) & \longrightarrow & \mathcal{P}(E_0 \times \ldots \times E_n) \\
\mu & \mapsto & \Phi_{[0,n]}(\mu) = \eta_0 \otimes \Phi_1(\mu_0) \otimes \ldots \Phi_n(\mu_{n-1})
\end{array}$$

where μ_k, $0 \leq k \leq n$, stands for the k-th marginal of μ then it is easy to check that $H_n^{(N)}$ can be rewritten as

$$H_n^{(N)}(x_0,\ldots,x_n) \ \ = \ \ N\ \overline{H}_n\left(\frac{1}{N}\sum_{i=1}^{N} \ \delta_{(x_0^i,\ldots,x_n^i)}\right)$$

with the potential function $\overline{H}_n$

$$\forall \mu \in \mathcal{P}(E_0 \times \ldots \times E_n)\,, \qquad \overline{H}_n(\mu) = \int \log \frac{d\Phi_{[0,n]}(\mu)}{d\Phi_{[0,n]}(\eta_{[0,n]})}\; d\mu$$

Before bringing the proof of (41) two elementary lemmas are stated and proved. As noticed in [1] a natural tool for the analysis of a strong version of the propagation of chaos for mean field interacting particle systems is the following inequality due to Csiszar [3] (see also the proof of Theorem 2 in [1]).

Lemma 3.9 ([3] (2.10) p. 772) *Let $(E, \mathcal{E})$ be a measurable space and let $\mu^{(N)}$ be an exchangeable measure on the product space E^N such that $\mu^{(N)} << \eta^{\otimes N}$ for some $\eta \in \mathcal{P}(E)$. If $\mu^{(N,q)}$, $1 \leq q \leq N$, are the marginals of $\mu^{(N)}$ on the first q-coordinates then we have*

$$\mathrm{Ent}(\mu^{(N,q)} \mid \eta^{\otimes q}) \leq \frac{q}{N} \left(1 + \frac{\{N/q\}}{[N/q]}\right)\; \mathrm{Ent}(\mu^{(N)} \mid \eta^{\otimes N}) \tag{44}$$

where $[a]$ is the integer part of $a \in \mathbb{R}$ and $\{a\} = a - [a]$.

Proof: The proof of (44) is quite simple. From the variational definition of the relative entropy

$$\mathrm{Ent}(\mu|\eta) = \sup_{f \in \mathcal{C}_b(E)} \{\mu(f) - \log \eta(\exp(f))\}$$

we have already $\mathrm{Ent}(\mu^{(N)} \mid \eta^{\otimes N}) \geq (\mu^{(N)}(f^{(q)}) - \log \eta^{\otimes N}(\exp f^{(q)}))$ with

$$f^{(q)}(x_1, \ldots, x_N) = \sum_{p=1}^{[N/q]} \varphi(x_{(p-1)q+1}, \ldots, x_{(p-1)q+q})\,, \qquad \varphi \in \mathcal{C}_b(E^q)$$

Since $\mu^{(N)}(f^{(q)}) = [N/q]\; \mu^{(N,q)}(\varphi)$ and $\eta^{\otimes N}(\exp f^{(q)}) = (\eta^{\otimes q}(\varphi))^{[N/q]}$ taking the supremum over $\varphi \in \mathcal{C}_b(E^q)$ one concludes that

$$\mathrm{Ent}(\mu^{(N)} \mid \eta^{\otimes N}) \geq\; [N/q]\; \mathrm{Ent}(\mu^{(N,q)} \mid \eta^{\otimes q})$$

We end the proof by noting that

$$\forall a \in [1, \infty]\,, \quad \frac{1}{[a]} = \frac{1}{a}\; \frac{[a] + \{a\}}{[a]} = \frac{1}{a}\; \left(1 + \frac{\{a\}}{[a]}\right) \left(\leq \frac{2}{a}\right)$$

■

Lemma 3.10 *If μ is absolutely continuous with respect to η and $\frac{d\mu}{d\eta} \in \mathbb{L}_2(\eta)$ then we have*

$$\mathrm{Ent}(\mu|\eta) \leq \left\| \frac{d\mu}{d\eta} - 1 \right\|_{2,\eta}^2$$

Proof: Using the standard inequality, $\log u \leq u - 1$ which is valid for any $u \geq 0$ we clearly have

$$\mathrm{Ent}(\mu|\eta) = \int \log \frac{d\mu}{d\eta}\; d\mu \leq \int \left(\frac{d\mu}{d\eta} - 1\right)\; d\mu = \int \left(\frac{d\mu}{d\eta} - 1\right)\; \frac{d\mu}{d\eta}\; d\eta$$

From which on concludes that

$$\mathrm{Ent}(\mu|\eta) \leq \int \left(\frac{d\mu}{d\eta} - 1\right)^2 d\eta = \left\|\frac{d\mu}{d\eta} - 1\right\|_{2,\eta}^2$$

■

Lemma 3.9 highlights the relations between the relative entropy and the mean value of the potential function $\overline{H}_n^{(N)}$. More precisely according to Lemma 3.9 we have that

$$\mathrm{Ent}\left(\mathbb{P}_{[0,n]}^{(N,q)} \middle| \eta_{[0,n]}^{\otimes q}\right) \leq 2\ \frac{q}{N}\ \mathrm{Ent}\left(\mathbb{P}_{[0,n]}^{(N)} \middle| \eta_{[0,n]}^{\otimes N}\right) = 2\ \frac{q}{N}\ \mathbb{E}(H_n^{(N)}(\xi_0, \ldots, \xi_n)) \tag{45}$$

It can be seen from definitions of $H_n^{(N)}$ and ξ that

$$\begin{aligned}
\mathbb{E}(H_n^{(N)}(\xi_0, \ldots, \xi_n)) &= N \sum_{p=1}^{n} \mathbb{E}\left(\mathbb{E}\left(\ m(\xi_p)[\ \log\left(d\Phi_p(m(\xi_{p-1}))/d\eta_p\right)\ ]\ |\ \xi_{p-1}\ \right)\ \right) \\
&= N \sum_{p=1}^{n} \mathbb{E}(\ \Phi_p(m(\xi_{p-1}))[\ \log\left(d\Phi_p(m(\xi_{p-1}))/d\eta_p\right)\ ]\ )
\end{aligned}$$

Therefore one gets

$$\mathrm{Ent}\left(\mathbb{P}_{[0,n]}^{(N)} \middle| \eta_{[0,n]}^{\otimes N}\right) = N \sum_{p=0}^{n-1} \mathbb{E}\left(\ \mathrm{Ent}\left(\Phi_{p+1}(m(\xi_p))\,|\Phi_{p+1}(\eta_p)\right)\ \right)$$

Using Lemma 3.10 the end of the proof of Theorem 3.5 is now straightforward. Indeed, under the assumptions of the theorem, Lemma 3.10 implies that

$$\begin{aligned}
\mathbb{E}\left(\mathrm{Ent}\left(\Phi_{p+1}(m(\xi_p))\,|\Phi_{p+1}(\eta_p)\right)\right) &\leq \mathbb{E}\left(\|1 - d\Phi_{p+1}(m(\xi_p))/d\eta_{p+1}\|_{2,\eta_{p+1}}^2\right) \\
&= \int \mathbb{E}\left(|1 - \frac{d\Phi_{p+1}(m(\xi_p))}{d\Phi_{p+1}(\eta_p)}(y)|^2\right)\ \eta_{p+1}(dy) \\
&\leq \frac{1}{N} \int \theta_{p+1}^2\ d\eta_{p+1} = \frac{1}{N}\ \|\theta_{p+1}\|_{2,\eta_{p+1}}^2
\end{aligned}$$

From this and taking into account (45) on concludes that

$$\mathrm{Ent}\left(\mathbb{P}_{[0,n]}^{(N,q)} \middle| \eta_{[0,n]}^{\otimes q}\right) \leq \frac{2\ q}{N}\ \sum_{p=1}^{n} \|\theta_p\|_{2,\eta_p}^2$$

and the proof of the theorem is completed. The earlier discussion given in page 26 leads to the following immediate corollary of (41), Theorem 3.5.

Corollary 3.11 *If $\{\Phi_n\ ;\ n \geq 1\}$ are given by (3) and conditions $(\mathcal{G})$ and $(\mathcal{K})_2$ are both satisfied then we have that*

$$\forall n \geq 0\,,\ \forall 1 \leq q \leq N\,, \qquad \mathrm{Ent}(\mathbb{P}_{[0,n]}^{(N,q)}\ |\ \eta_{[0,n]}^{\otimes q}) \leq \frac{q\ c(n)}{N} \tag{46}$$

In addition if the uniform bounds in (10) are satisfied then (41) holds with some linear function of the time parameter, that is $c(n) = c\,n$ for some $c < \infty$.

References

[1] O. I Ben Arous, G.; Zeitouni. Increasing propagation of chaos for mean field models. *Ann. Inst. H. Poincar Probab. Statist.*, 35(1):85–102., 1999.

[2] I. Csiszar. Information type measures of difference of probability distributions and indirect observations. *Studia. Sci. Math. Hungar*, 2:299–318, 1967.

[3] I. Csiszar. Sanov property, generalized i-projection and a conditional limit theorem. *The Annals of Probability*, 12(3):768–793, 1984.

[4] P. Del Moral. Measure valued processes and interacting particle systems. Application to non linear filtering problems. *The Annals of Applied Probability*, 8(2):438–495, 1998.

[5] P. Del Moral and A. Guionnet. Large deviations for interacting particle systems. Applications to non linear filtering problems. *Stochastic Processes and their Applications*, 78:69–95, 1998.

[6] P. Del Moral and J. Jacod. Interacting particle filtering with discrete observations. Publications du Laboratoire de Statistiques et Probabilités, no 11-99, 1999.

[7] P. Del Moral and M. Ledoux. Convergence of empirical processes for interacting particle systems with applications to nonlinear filtering. *Journal of Theoretical Probability*, 13(1):225–257, 2000.

[8] P. Del Moral and L. Miclo. Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non linear filtering. In J. Azéma, M. Emery, M. Ledoux, and M. Yor, editors, *Séminaire de Probabilités XXXIV*, Lecture Notes in Mathematics 1729, pages 1–145. Springer-Verlag, Berlin, 2000.

[9] P. Del Moral and L. Miclo. A Moran particle system approximation of Feynman-Kac formulae. to appear in Stochastic Processes and their Applications, 2000.

[10] A.N. Van der Vaart and J.A. Wellner. *Weak Convergence and Empirical Processes with Applications to Statistics.* Springer Series in Statistics. Springer, 1996.

[11] P. Donnelly and T.G. Kurtz. Genealogical processes for fleming-viot models with selection and recombination. To appear in The Annals of Applied Probability.

[12] P. Donnelly and T.G. Kurtz. Particle representations for measure-valued population models. *Ann. Probab.*, 27:166–205, 1999.

[13] S.N. Ethier and R.C. Griffiths. The infinitely-many-sites-model as a measure valued diffusion. *Ann. Prob.*, 15:515–545, 1987.

[14] C. Graham and S. Méléard. Stochastic particle approximations for generalized Boltzmann models and convergence estimates. *The Annals of Probability*, 25(1):115–132, 1997.

[15] S. Méléard. Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. In D. Talay and L. Tubaro, editors, *Probabilistic Models for Nonlinear Partial Differential Equations, Montecatini Terme, 1995*, Lecture Notes in Mathematics 1627. Springer-Verlag, 1996.

Pierre Del Moral, Laurent Miclo
Laboratoire de Statistiques et Probabilités
CNRS UMR C5583
Université Paul Sabatier
118, route de Narbonne
31062 Toulouse cedex, France
E-mail: delmoral@cict.fr
E-mail: miclo@cict.fr