Branching and Interacting Particle Systems Approximations of Feynman-Kac Formulae with Applications to Non-Linear Filtering

P. Del Moral L. Miclo

LSP-CNRS and Université Toulouse III

abstract

This paper focuses on interacting particle systems methods for solving numerically a class of Feynman-Kac formulae arising in the study of certain parabolic differential equations, physics, biology, evolutionary computing, nonlinear filtering and elsewhere. We have tried to give an "exposé" of the mathematical theory that is useful for analyzing the convergence of such genetic-type and particle approximating models including law of large numbers, large deviations principles, fluctuations and empirical process theory as well as semigroup techniques and limit theorems for processes.
In addition, we investigate the delicate and probably the most important problem of the long time behavior of such interacting measure valued processes.
We will show how to link this problem with the asymptotic stability of the corresponding limiting process in order to derive useful uniform convergence results with respect to the time parameter.
Several variations including branching particle models with random population size will also be presented. In the last part of this work we apply these results to continuous time and discrete time filtering problems.

Keywords:

Interacting and branching particle systems, genetic algorithms, weighted sampling Moran processes, measure valued dynamical systems defined by Feynman-Kac formulae, asymptotic stability, chaos weak propagation, large deviations principles, central limit theorems, nonlinear filtering.

A.M.S. codes:

60G35, 60F10, 60H10, 60G57, 60K35, 60F05, 62L20, 92D25, 92D15, 93E10, 93E11.

Contents

1 Introduction

1.1 Background and Motivations

The aim of this set of notes is the design of a branching and interacting particle system (abbreviate **BIPS**) approach for the numerical solving of a class of Feynman-Kac formulae which arise in the study of certain parabolic differential equations, physics, biology, evolutionary computing, economic modelling and nonlinear filtering problems.

Our major motivation is from advanced signal processing and particularly from optimal nonlinear filtering problems. Recall that this consists in computing the conditional distribution of a partially observed Markov process.

In discrete time and in a rather general setting the classical nonlinear filtering problem can be summarized as to find distributions of the form

$$\forall\ f \in \mathcal{B}_{\mathrm{b}}(E),\ \forall\ n \geq 0, \qquad \eta_n(f) \ = \ \frac{\gamma_n(f)}{\gamma_n(1)} \tag{1}$$

where $\mathcal{B}_{\mathrm{b}}(E)$ is the space of real bounded measurable functions over a Polish state space E and $\gamma_n(f)$ is a Feynman-Kac formula given by

$$\gamma_n(f) = \mathbb{E}\left(f(X_n) \prod_{m=1}^{n} g_m(X_{m-1}) \right) \tag{2}$$

where $\{X_n\ ;\ n \geq 0\}$ is a given time inhomogeneous Markov chain taking values in E and $\{g_m\ ;\ m \geq 1\}$ is a given sequence of bounded positive functions.

In continuous time, the computation of the optimal pathwise filter can be summarized as to find the flow of distributions

$$\forall\ f \in \mathcal{B}_{\mathrm{b}}(E)\,, \forall\ t \in \mathbb{R}_+, \qquad \eta_t(f) \ = \ \frac{\gamma_t(f)}{\gamma_t(1)} \tag{3}$$

where $\gamma_t(f)$ is again defined through a Feynman-Kac formula of the following form

$$\gamma_t(f) = \mathbb{E}\left(f(X_t)\ \exp\left(\int_0^t U_s(X_s)\, ds \right) \right) \tag{4}$$

This time $\{X_t\ ;\ t \in \mathbb{R}_+\}$ denotes an E-valued càdlàg inhomogeneous Markov process and $\{U_t\ ;\ t \in \mathbb{R}_+\}$ is a measurable collection of locally bounded (in time) and measurable nonnegative functions.

Even if equations (1) and (3) look innocent they can rarely be solved analytically and their solving require extensive calculations. More precisely, with the notable exception of the so-called "linear-Gaussian" situation (Kalman-Bucy's filter [15]) or wider classes of models (Benes' filters [12]) optimal filters have no finitely recursive solution [20]. To obtain a computationally feasible solution some kind of approximation is needed.

Of course, there are many filtering algorithms that have been developed in mathematics and signal processing community. Until recently most works in this direction were based on fixed grid approximations, conventional linearization (Extended Kalman Filters) or determining the best linear filter (in least squares sense). These various numerical methods have never really cope with large scale systems or unstable processes. Comparisons and examples when the extended Kalman-Bucy filter fails can be found for instance in [15]. In addition all these deterministic schemes have to be handled very carefully mainly because they are usually based on specific criteria and rates of convergence are not always available.

The particle algorithms discussed in these lectures belong to the class of Monte Carlo methods and they do not use regularity informations on the coefficients of the models. Thus, large scale systems and nonlinear models with non sufficiently smooth coefficients represent classes of nonlinear filtering problems to which particle methods might be applied. These methods are in general robust and very efficient and many convergence results are available including uniform convergence results with respect to the time parameter. But, from a strict practical point of view, if there exists already a good specialized method for a specific filtering problem then the BIPS approach may not be the best tool for that application.

Let us briefly survey some distinct approaches and motivate our work.

In view of the functional representations (1) and (3) the temptation is also to apply classical Monte-Carlo simulations based on a sequence of independent copies of the process X. Unfortunately it is well known that the resulting particle scheme is not efficient mainly because the deviation of the particles may be too large and the growth of the exponential weights with respect to the time parameter is difficult to control (see for instance [28, 34, 57]).
In [34] we propose a way to regularize these weights and we give a natural ergodic assumption on the signal semigroup under which the resulting Monte-Carlo particle scheme converges in law to the optimal filter uniformly with respect to the time parameter.
In more general situations, complications occur mainly because this particle scheme is simply based on a sequence of independent copies of the signal. This is not surprising: roughly speaking the law of signal and the desired conditional distribution may differ considerably and they may be too few particles in the space regions with high probability mass.

Among the most exciting developments in nonlinear filtering theory are those centering around the recently established connections with branching and interacting particle systems. The evolution of this rapidly developing area of research may be seen quite directly through the following chains of papers [23, 21] [25, 24], [30, 33, 31] [35, 37, 36], [42], [41, 40], [45], [32, 47] as well as [11, 18, 46, 73, 64] and finally [59, 85, 84].

Instead of *hand crafting* algorithms, often based on the basis of had-hoc criteria, particle systems approaches provide powerful tools for solving a large class of nonlinear filtering problems. In contrast to the first Monte-Carlo scheme the branching and interacting particle approximating models involve the use of a system of particles which evolve in correlation with each other and give birth to a number of offsprings depending on the observation process. This guarantees an occupation of the probability space regions proportional to their probability mass thus providing a well behaved adaptative and stochastic grid approximating model. Furthermore these particle algorithms also belong to the class of resampling methods and they have been made valuable in practice by advances in computer technology [52]. Different adaptative locally refined but deterministic multi-grid methods can be found in [17]. In contrast to BIPS approaches the latter are limited to low dimensional state space examples.

It is hard to know where to start in describing contributions to BIPS approximations of Feynman-Kac formulae.

In discrete time and nonlinear filtering settings the more embryonic form of interacting particle scheme appeared in the independent works [32, 47], [64] and [73]. The first proof of convergence of these heuristics seems to be [30, 31]. The analysis of the convergence has been further developed in [33, 35, 37, 36, 42].

In continuous time settings the origins of interacting particle schemes is a more delicate problem. The first studies in continuous time settings seem to be [23] and [21]. These works were developed independently of the first set of referenced papers. The authors present a branching particle approximating model without any rates of convergence and the main difference with previous interacting particle schemes comes from the fact that the number of particle is not fixed but random. Moreover the authors made the crucial assumptions that we can exactly simulate random transitions according to the semigroup of the continuous time signal and stochastic integrals arising in Girsanov exponentials are exactly known. Therefore these particle algorithms do not applied directly to the continuous time case. On the other hand these branching particle models are based on a time discretization procedure. As a result the corresponding nonlinear filtering problem can be reduced to a suitably defined discrete time filtering problem. The corresponding discrete time version of such branching and interacting particle schemes as well as the first convergence rates are described in [25] and [24].

The studies [41] and [40, 39] discuss several new interacting particle schemes for solving nonlinear filtering problems where the signal is continuous but the number of observations is finite. To get some feasible solution which can be used in practice several additional levels of approximations including time discretizations are also analyzed. In contrast to previous referenced papers these schemes can also be used for solving numerically filtering problems with correlated signal and observation noise sources.

As we shall see in the further development of section 1.2 the interacting or branching particle schemes based on an additional time discretization procedure are not really efficient for solving continuous time filtering problems. The authors presented in [45] a genuine continuous time genetic type particle scheme for solving the robust optimal filter. This scheme will be discussed in section 1.3 and section 3.

The connections between this IPS model, the classical Moran IPS and the Nanbu IPS (arising respectively in the literature of genetic algorithms and Boltzmann equations) are discussed in section 1.3.

The modelling and the analysis of such particle approximating models has matured over the past ten years in ways which make it much more complete and rather beautiful to learn and to use. One objective of these notes is to introduce the reader to branching and interacting particle interpretations of Feynman-Kac formulae of type (1)-(4).
We have also tried to give an "exposé" of the mathematical theory that it is useful in analyzing the convergence of such approximating models including law of large numbers, large deviations, fluctuations and empirical process theory, as well as semigroup techniques and functional limit theorems for stochastic processes.

Although only a selection of existing results is presented, many results appear here for the first time and several points have been improved. The proofs of existing results are only sketched but the methodologies are described carefully. Deeper informations are available in the list of referenced papers.

The material for this paper has also been chosen in order to give some feel of the variety of the theory but the development is guided by the classical interplay between theory and detailed consideration of application to specific nonlinear filtering models.

This set of notes is very far from being exhaustive and only surveys results that are closely related to BIPS-approximations of Feynman-Kac formulae and non linear filtering problems. Among the topics omitted are those centering around evolutionary computing and numerical function optimization problems. Among the huge literature on evolutionary computing and genetic algorithms we refer to [6, 7, 14], [19], [44], [60, 61, 62], [63] and [111].
We emphasize that the so-called simple genetic algorithm is a special case of the BIPS models presented in this work.
In this connection, the measure valued distribution flows (1)-(4) and the corresponding interacting particle approximating models can be regarded as the so-called infinite and finite population models. Therefore the methodologies presented in this work can be used for establishing the most diverse limit theorems on the long time behavior of these models as well as the asymptotics of the finite population model as the number of individuals tends to infinity.

An overview of the material presented in these notes was presented in a three one hour lectures for the Symposium/Workshop on Numerical Stochastics (April 1999) at the Fields Institute for Research in Mathematical Sciences (Toronto). At the same period they were presented at the University of Alberta Edmonton with the support of the Canadian Mathematics of Information Technology and Complex Systems project (MITACS).
We would like to thank Professor M. Kouritzin from the University of Alberta for stimulating discussions.

A part of this material presented in this set of notes results from collaborations of one of the authors with D. Crisan and T. Lyons [25, 24], with A. Guionnet [35, 37, 36], with J. Jacod [40] and Ph. Protter [41], and with M. Ledoux [42].

We also heartily thank Gérard Ben Arous, Carl Graham and Sylvie Méléard for encouraging and fruitful discussions and Michel Ledoux for inviting us to write these notes for *Le Séminaire de Probabilités*.

We gratefully acknowledge CNRS research fellowship No 97N23/0019 "Modélisation et simulation numérique", European community for the CEC Contracts No ERB-FMRX-CT96-0075 and INTAS-RFBR No 95-0091 and the Canadian Network MITACS-Prediction In Interacting Systems.

1.2 Motivating Examples

The interacting particle interpretation of the Feynman-Kac models (1)-(4) has had numerous applications in many nonlinear filtering problems; to name a few, radar signal processing ([46, 47]), global positioning system ([18]) and tracking problems ([73, 85, 84, 64]). Other numerical experiments are also given in [21] and [41].

The purpose of these notes is to expose not only the theory of interacting particle approximations of Feynman-Kac formulae but also to provide a firm basis for the understanding and solving nonlinear filtering problems. To guide the reader and motivate this study we present here two generic models and the discrete and continuous time Feynman-Kac formulations of the corresponding optimal filter.

The distinction between continuous and discrete time will lead to different kind of interacting particle approximating models. Intuitively speaking continuous time models correspond to processes of classical physics while discrete time models arise in a rather natural way as soon as computers are part of the process. More general and detailed models will be discussed in the further development of section 4.

In discrete time settings the state signal $X = \{X_n \ ; \ n \geq 0\}$ is an $\mathbb{R}^p$-valued Markov chain usually defined through a recursion of the form

$$X_n = F_n(X_{n-1}, W_n)$$

where $W = \{W_n \ ; \ n \geq 1\}$ is a noise sequence of independent and $\mathbb{R}^q$-valued random variables. For each $n \geq 1$ the function $F_n : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^p$ is measurable and the initial value X_0 is independent of W. The above recursion contains the laws of evolution for system states such as the laws of evolution of a target in tracking problems, an aircraft in radar processing or inertial navigation errors in GPS signal processing. The noise component W models the statistics of unknown control laws of an aircraft or a cruise control in an automobile or a non cooperative target as well as uncertainties in the choice of the stochastic mathematical model.

For future reference it is convenient to generalize the definition of the state signal X. More precisely an alternative way to define X consists in embedding the latter random dynamical system through its transition probabilities. This approach gives some insights into ways of thinking the evolution of the marginal laws of X and it also allows to consider signals taking values on infinite dimensional spaces. For these reasons we will systematically assume that the sequence $X = \{X_n \ ; \ n \geq 0\}$ is a Markov process taking values in a Polish state space E with transition probabilities $\{K_n \ ; \ n \geq 1\}$ and initial distribution η_0. As we shall see in the further development this formulation also allows a uniform treatment of filtering problems with continuous time signals and discrete time observations. The signal X is not known but partially observed. Usually we assume that the observation process $Y = \{Y_n \ ; \ n \geq 1\}$ is a sequence of $\mathbb{R}^r$-valued random variables given by

$$Y_n = h_n(X_{n-1}) + V_n$$

where $V = \{V_n \ ; \ n \geq 1\}$ are independent and $\mathbb{R}^r$ valued random variables whose marginal distributions possess a density $\varphi_n(v)$ with respect to Lebesgue measure on $\mathbb{R}^r$ and for each $n \geq 1$, $h_n : E \to \mathbb{R}^r$ is a measurable function.
Here again the design of the disturbance sequence V depends on the class of sensors at hand. For instance noise sources acting on sensors model thermic noise resulting from electronic devices or atmospheric propagation delays and/or received clock bias in GPS signal processing. For more details we refer the reader to the set of referenced articles.

Given the stochastic nature of the pair signal/observation process and given the observation values $Y_n = y_n$, for each $n \geq 1$, the nonlinear filtering problem consists in computing recursively in time the one step predictor conditional probabilities η_n and the filter conditional distributions $\widehat{\eta}_n$ given for any bounded Borel test function f by

$$\begin{aligned}
\eta_n(f) &= \mathbb{E}(f(X_n)|Y_1 = y_1, \ldots, Y_n = y_n) \\
\widehat{\eta}_n(f) &= \mathbb{E}(f(X_n)|Y_1 = y_1, \ldots, Y_n = y_n, Y_{n+1} = y_{n+1})
\end{aligned}$$

As usually the n-step filter $\widehat{\eta}_n$ is written in terms of η_n as

$$\widehat{\eta}_n(f) = \frac{\displaystyle\int_E f(x) \ \varphi_{n+1}(y_{n+1} - h_{n+1}(x)) \ \eta_n(dx)}{\displaystyle\int_E \varphi_{n+1}(y_{n+1} - h_{n+1}(x)) \ \eta_n(dx)}$$

and the n-step predictor may be defined in terms of the Feynman-Kac type formula

$$\eta_n(f) = \frac{\gamma_n(f)}{\gamma_n(1)}$$

with

$$\gamma_n(f) = \mathbb{E}\left(f(X_n) \ \prod_{m=1}^{n} \varphi_m(y_m - h_m(X_{m-1})) \right)$$

We will return to this model with more detailed examples in section 5.

The second filtering model presented hereafter is a continuous time (signal/observation) Markov process $\{(S_t, Y_t)\ ;\ t \in \mathbb{R}_+\}$ taking values in $\mathbb{R}^p \times \mathbb{R}^q$. It is solution of the Itô's stochastic differential equation

$$\left\{\begin{array}{lcl} dS_t &=& A(t,S_t)\,dt + B(t,S_t)\,dW_t + \displaystyle\int_{\mathbb{R}^m} C(t,S_{t-},u)\,(\mu(dt,du) - \nu(dt,du)) \\ dY_t &=& h(S_t)\,dt + \sigma\,dV_t \end{array}\right.$$

(V,W) is an $(q+r)$ dimensional standard Wiener process and μ is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}^m$ with intensity measure $\nu(dt,du) = dt \otimes F(du)$ and F is a positive σ-finite measure on $\mathbb{R}^m$. The mappings $A : \mathbb{R}_+ \times \mathbb{R}^p \to \mathbb{R}^p$, $B : \mathbb{R}_+ \times \mathbb{R}^p \to \mathbb{R}^p \otimes \mathbb{R}^r$, $C : \mathbb{R}_+ \times \mathbb{R}^p \times \mathbb{R}^m \to \mathbb{R}^p$, and $h : \mathbb{R}^p \to \mathbb{R}^q$ are Borel functions, S_0 is a random variables independent of (V,W,μ) and $Y_0 = 0$.
Here again the first equation represents the evolution laws of the physical signal process at hand. For instance the Poisson random measure μ may represent jumps variations of a moving and non cooperative target (see for instance [47]).

Next we examine three situations. In the first one we assume that observations are given only at regularly spaced times $t_0, t_1, \ldots, t_n, \ldots \in \mathbb{R}_+$ ($t_0 = 0, Y_0 = 0$) and we are interested in the conditional distributions given for any bounded Borel function $f : \mathbb{R}^p \to \mathbb{R}$ by

$$\tilde{\pi}_n(f) \stackrel{\text{def.}}{=} \mathbb{E}(\,f(S_{t_n})\,|Y_{t_1} = y_1, \ldots, Y_{t_n} = y_n)$$

where $y_1, \ldots, y_n \in \mathbb{R}^q$ is a given sequence of observations.
If we denote $E = D([0,t_1[,\mathbb{R}^p)$ the Polish space of càdlàg paths from $[0,t_1[$ into $\mathbb{R}^p$ then the discrete time sequence

$$\forall\, n \geq 0, \qquad X_n \ \stackrel{\text{def.}}{=}\ S_{[t_n,t_{n+1}[} \circ \theta_{t_n}$$

(where $(\theta_t)_{t\geq 0}$ stands for the usual family of time shifts), is an E-valued Markov chain. On the other hand if $H : E \to \mathbb{R}^q$ is the mapping defined by

$$\forall x \in E\,, \qquad H(x) = \int_0^{t_1} h(x_s)\,ds$$

then using the above notations for any $n \geq 1$:

$$Y_{t_n} - Y_{t_{n-1}} = H(X_{n-1}) + \sigma(V_{t_n} - V_{t_{n-1}})$$

This description is interesting in that it shows that the latter filtering problem is equivalent to the previous discrete time model. In this connection it is also worth noting that for any bounded Borel test function $f : E \to \mathbb{R}$

$$\eta_n(f) \stackrel{\text{def.}}{=} \mathbb{E}\big(\,f(S_{[t_n,t_{n+1}[})\,|Y_{t_1} = y_1, \ldots, Y_{t_n} = y_n\big) = \frac{\gamma_n(f)}{\gamma_n(1)}$$

with

$$\gamma_n(f) = \mathbb{E}\left(f(X_n)\ \prod_{m=1}^n \varphi_m(y_m - y_{m-1} - H(X_{m-1}))\right)$$

where, for any $m \geq 1$, φ_m is the density of the Gaussian variable $\sigma(V_{t_m} - V_{t_{m-1}})$. This observation also explain why it is necessary to undertake a study of Feynman-Kac formula of type (1) with Polish valued Markov processes X.

A more traditional question in continuous time nonlinear filtering settings is to compute the conditional distributions $\{\pi_t \; ; \; t \in \mathbb{R}_+\}$ given for any bounded Borel functions $f : \mathbb{R}^p \to \mathbb{R}$ by setting

$$\pi_t(f) = \mathbb{E}\left(f(S_t) \,\middle|\, \mathcal{Y}_{[0,t]}\right) \tag{5}$$

where $\mathcal{Y}_{[0,t]}$ is the filtration generated by the observations up to time t. Roughly speaking this continuous time problem is close to the previous discrete time model when the time step $\Delta t = t_n - t_{n-1}$ is sufficiently small and when observations are delivered continuously in time. For instance in radar processing the measurements derived from the return signal of a moving target can be regarded as a pulse train of rectangular or Gaussian pulses with period 10^{-4} seconds.
In real applications one usually consider the discrete time signal model

$$\{S_{t_n} \; ; \; n \geq 0\}$$

as $\mathbb{R}^p$-valued Markov chain and one replaces the continuous time observation process by the discrete time sequence

$$\Delta Y_{t_n} \stackrel{\text{def.}}{=} h\left(S_{t_{n-1}}\right) \, \Delta t + \sigma(V_{t_n} - V_{t_{n-1}}) \tag{6}$$

This first level of approximation is commonly used in practice and the error caused by the discretization of the time interval is well understood (see for instance [76] and references therein as well as section 4.3 in these notes).
One consequence of the previous discretizarion is that the filtering problem is now reduced to find a way of computing the conditional distributions given for any bounded Borel function $f : \mathbb{R}^p \to \mathbb{R}$ by

$$\pi_n^\Delta(f) \stackrel{\text{def.}}{=} \mathbb{E}(\, f(S_{t_n}) \,|\Delta Y_{t_1} = y_1 - y_0, \ldots, \Delta Y_{t_n} = y_n - y_{n-1}\,) = \frac{\gamma_n^\Delta(f)}{\gamma_n^\Delta(1)}$$

with $y_0 = 0$ and

$$\gamma_n^\Delta(f) = \mathbb{E}\left(f(S_{t_n}) \prod_{m=1}^n \varphi_m(y_m - y_{m-1} - h(S_{t_{m-1}})\Delta t)\right)$$

One drawback of this formulation is that the corresponding IPS approximating model is not really efficient. As we shall see in the further development of section 1.3 and section 4.3 the evolution of this scheme is decomposed into two genetic type selection/mutation transitions. During each selection stage the system of particles takes advantage of the current observation data in order to produce an adaptative grid.
The underlying n^{th} step selection transition is related to the fitness functions g_n defined by

$$g_n(x) = \varphi_n(y_n - y_{n-1} - h(S_{t_{n-1}})\Delta t)$$

but the physical noise in sensors as well as the choice of the short time step Δt critically corrupt the information delivered between two dates (recall that the current observation at time n has the form (6)). One consequence is that the resulting particle

scheme combined with the previous time discretization converges slowly to the desired conditional distributions (5) (see for instance section 4.3 and [24] as well as [23, 21] for a branching particle alternative scheme).
One way to improve these rates is to use a genuine continuous time and interacting particle approximating model. In this alternative approach the information used at each selection date is not the "increments" of the observation process but the current observation value at that time.

The key idea presented in [45] and further developed in this work is to study the robust and pathwise filter defined for any $y \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}^q)$ (and not only on a set of probability measure 1) and for any bounded Borel function $f : \mathbb{R}^p \to \mathbb{R}$ by a formula of the form

$$\pi_{y,t}(f) = \frac{\displaystyle\int_{\mathbb{R}^p} f(x)\ e^{h^\star(x)y_t}\ \eta_{y,t}(dx)}{\displaystyle\int_E e^{h^\star(x)y_t}\ \eta_{y,t}(dx)} \qquad \text{with} \quad \eta_{y,t}(f) \stackrel{\text{def.}}{=} \frac{\gamma_{y,t}(f)}{\gamma_{y,t}(1)}$$

and $\gamma_{y,t}$ is again defined through a Feynman-Kac type formula

$$\gamma_{y,t}(f) = \mathbb{E}\left(f(X_t^y)\ \exp\int_0^t\ V_s(X_s^y, y_s)\ ds\right)$$

For any $s \in \mathbb{R}_+$, $V_s : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}_+$ is a Borel measurable function which depends on the coefficients of the filtering problem at hand and $\{X_t^y\ ;\ t \in \mathbb{R}_+\}$ is a Markov process which depends on the observations. To describe precisely these mathematical models we need to introduce several additional notations. We will return to this model with detailed examples of signal processes which can be handled in our framework in section 4.2.

The convergence results for the resulting interacting particle approximating model will improve the one presented in section 4.3 and in the articles [23, 21, 24]. From a practitioner's view point the main difference between these two approaches lies in the fact that in the former the selection and interaction mechanism only depends on the current observation y_t and on the fitness function $U_t(x) \stackrel{\text{def.}}{=} V_t(x, y_t)$. Moreover under natural stability assumptions the resulting IPS scheme converges uniformly in time to the desired optimal filter.

1.3 Description of the Models

To provide a red line in this work and to point out the connections with classical mean-field interacting particle system theory, the models are focused around two approximating interacting particle systems (abbreviate **IPS**). The research literature abounds with variation of these two IPS models. The interested reader will also find a detailed description of several variants including branching schemes with random population size, periodic selection schemes and conditional mutations.

Most of the terminology we have used is drawn from mean field IPS and measure valued processes theory. We shall see that the flows of distributions (1) and (3) are solutions of deterministic measure valued and nonlinear evolution equations. These equations will be used to define the transition probability kernel of the corresponding IPS approximating models. In this sense the deterministic evolution equations will be regarded as the limiting measure valued process associated with a sequence of IPS schemes.

The resulting IPS models can also be regarded as genetic type algorithms as those arising in numerical function optimization, biology and population analysis (cf. [62, 67, 87, 63]).
The range of research areas in which genetic algorithms arise is also quite board, to name a few: machine learning [61], control systems [60], electromagnetics [69, 110], economics and finance [8, 80, 97], aircraft landing [1, 2], topological optimum design [72] and identification of mechanical inclusions [70, 71].

In continuous time settings the corresponding IPS can be viewed as a weighted sampling Moran particle system model. Moran particle models arise in population genetics. They usually model the evolution in time of the genetic structure of a large but finite population (see for instance [29] and references therein). In the classical theory of measure valued processes the limiting process is random and it is called the Fleming-Viot process. In this setting the limiting process is commonly used to predict the collective behavior of the system with a finite but large number of particles.
In contrast to the above situation the desired distributions (1) and (3) are not random (except, in filtering settings, through the observation process). It is therefore necessary to find a new strategy to define an IPS scheme that will approximate (1) and (3).

In time homogeneous settings and in the case where $E = \mathbb{R}^d$, $d \geq 1$, the evolution equation (11) and the corresponding IPS can also be regarded, in some sense, as a simple generalized and spatially homogeneous Boltzmann equation and as the corresponding Nanbu type IPS (see for instance [65, 83] and references therein). At this juncture many results presented in this work can be applied, but this is outside the scope of these notes, to study the fluctuations or to prove uniform convergence results for a class of Nanbu IPS.

Let us finally mention that the Feynman-Kac models considered in this study can be regarded as the distributions of a random Markov particle X killed at a given rate and conditioned by non-extinction (see for instance [96]). In this connection the asymptotic stability properties of the limiting processes associated with the Feynman-Kac models can be used to give conditions under which a killed particle conditioned by non-extinction forgets exponentially fast its initial condition (see for instance [43]).

In view of the above discussion the objects on which the limiting processes (8), (11) and the corresponding IPS schemes are sought may vary considerably. It is therefore necessary to undertake a study of the convergence in an abstract and time inhomogeneous setting.
For instance, in nonlinear filtering applications the time-inhomogeneous assumption is essential since the objects depend on the observation process.

The genetic type algorithms arising in numerical function analysis or nonlinear control problems usually also depends on a specific cooling schedule parameter (see for instance [19, 44])

1.3.1 The Limiting Measure Valued Models

To describe precisely the limiting measure valued models, let us introduce some notations. Let (E, r) be a Polish space, ie E is a separable topological space whose topology is given by a metric r which is supposed to be complete.
Let $\mathbf{B}(E)$ be the σ-field of Borel subsets of E. We denote by $\mathbf{M}(E)$ the space of all finite and signed Borel measures on E. Let $\mathbf{M}_1(E) \subset \mathbf{M}(E)$ be the subset of all probability measures. As usual, both spaces will be furnished with the weak topology. We recall that weak topology is generated by the Banach space $\mathcal{C}_{\mathrm{b}}(E)$ of all bounded continuous functions, endowed with the supremum norm, defined by

$$\forall\ f \in \mathcal{C}_{\mathrm{b}}(E), \qquad \|f\| \;=\; \sup_{x \in E} |f(x)|$$

(since $\mathbf{M}(E)$ can be regarded naturally as a part of the dual of $\mathcal{C}_{\mathrm{b}}(E)$). More generally, the norm $\|\cdot\|$ is defined in the same way on $\mathcal{B}_{\mathrm{b}}(E)$, the set of all bounded measurable functions, which is also a Banach space.

We denote by μK the measure given by $\mu K(A) = \int_E \mu(dx)\, K(x, A)$ where K is any integral operator on $\mathcal{B}_{\mathrm{b}}(E)$, $\mu \in \mathbf{M}(E)$ and $A \in \mathbf{B}(E)$. We also write

$$\forall\ f \in \mathcal{B}_{\mathrm{b}}(E)\,, \qquad \mu K(f) = \int \mu(dx)\, K(x, dz)\, f(z). \tag{7}$$

If K_1 and K_1 are two integral operators on $\mathcal{B}_{\mathrm{b}}(E)$ we denote by $K_1 K_2$ the composite operator on $\mathcal{B}_{\mathrm{b}}(E)$ defined for any $f \in \mathcal{B}_{\mathrm{b}}(E)$ by

$$K_1 K_2 f(x) = \int_E K_1(x, dy)\, K_2(y, dz)\, f(z)$$

We also denote by $\{K_n\ ;\ n \geq 1\}$ the transition probability kernels (respectively $\{L_t\ ;\ t \geq 0\}$ the family of pregenerators) of the discrete time (resp. the continuous time) Markov process X.

To get formally the nature of such schemes we first note that the distribution flow $\{\eta_n\ ;\ n \geq 0\}$ defined by (1) page 3 is a solution of the following measure valued dynamical system

$$\forall\ n \geq 1, \qquad \eta_n \;=\; \Phi_n(\eta_{n-1}) \tag{8}$$

For all $n \geq 1$, $\Phi_n : \mathbf{M}_1(E) \to \mathbf{M}_1(E)$ is the mapping defined by

$$\begin{aligned} \Phi_n(\eta) &= \Psi_n(\eta) K_n \\ \forall\ f \in \mathcal{B}_{\mathrm{b}}(E), \qquad \Psi_n(\eta)(f) &= \frac{\eta(g_n f)}{\eta(g_n)} \end{aligned} \tag{9}$$

We note that the recursion (8) involves two separate transitions:

$$\eta_{n-1} \xrightarrow{\text{Updating}} \widehat{\eta}_{n-1} \stackrel{\text{def}}{=} \psi_n(\eta_{n-1}) \xrightarrow{\text{Prediction}} \eta_n = \widehat{\eta}_{n-1} K_n \tag{10}$$

The first one is nonlinear and it will be called the updating step and the second one is linear and it will be called the prediction transition with reference to filtering theory.

In the continuous time situation, the distributions flow $\{\eta_t \; ; \; t \geq 0\}$ defined by (3) page 3 satisfies for any regular test function f the following nonlinear evolution equation

$$\frac{d}{dt}\,\eta_t(f) = \eta_t(\mathcal{L}_{t,\eta_t}(f)) \tag{11}$$

where $\mathcal{L}_{t,\eta}$, for $t \geq 0$ and $\eta \in \mathbf{M}_1(E)$ fixed, is a pregenerator on E, defined on a suitable domain by

$$\mathcal{L}_{t,\eta}(f)(x) = L_t f(x) + \int (f(z) - f(x))\, U_t(z)\, \eta(dz) \tag{12}$$

1.3.2 Interacting Particle Systems Models

In this section we introduce the IPS approximating models of the previous evolution equations (8) and (11). To give an initial introduction and to illustrate the idea in a simple form we emphasize that the particle approximation approach described here is not restricted to the particular form of mappings $\{\Phi_n \; ; n \geq 1\}$ or to the nature of the pregenerators $\{\mathcal{L}_{t,\eta} \; ; \; t \geq 0, \;\; \eta \in \mathbf{M}_1(E)\}$.

In discrete time, starting from a collection of mappings

$$\Phi_n : \mathbf{M}_1(E) \to \mathbf{M}_1(E), \qquad n \geq 1,$$

we consider an N-IPS, $\xi_n = (\xi_n^1, \ldots, \xi_n^N)$, $n \geq 0$, which is a Markov chain on the product state space E^N with transition probability kernels satisfying

$$P\left(\xi_n \in dx \,|\, \xi_{n-1} = z\right) = \prod_{p=1}^{N} \Phi_n(m(z))(dx^p) \tag{13}$$

where $dx \stackrel{\text{def}}{=} dx^1 \times \cdots \times dx^N$ is an infinitesimal neighborhood of the point $x = (x^1, \ldots, x^N) \in E^N$, $z = (z^1, \ldots, z^N) \in E^N$, δ_a stands for the Dirac measure at $a \in E$ and

$$\forall\; z = (z^1, \ldots, z^N) \in E^N, \qquad m(z) \;\; = \;\; \frac{1}{N} \sum_{i=1}^{N} \delta_{z^i} \;\in\; \mathbf{M}_1(E) \tag{14}$$

The initial system $\xi_0 = (\xi_0^1, \ldots, \xi_0^N)$ consists in N independent particles with common law η_0.
Intuitively speaking it is quite transparent from the above definition that if Φ_n is sufficiently regular and if $m(\xi_{n-1})$ is close to the desired distribution η_{n-1} then one expects that $\Phi_n(m(\xi_{n-1}))$ is a nice approximating measure for η_n. Therefore at the next step the particle system $\xi_n = (\xi_n^1, \ldots, \xi_n^N)$ looks like a sequence of independent random variables with common law η_n. This general and abstract IPS model first appeared in [33, 31] and its analysis was further developed in [35].

In much the same way starting from a family of pregenerators

$$\{\mathcal{L}_{t,\eta}\ ;\ t \geq 0,\ \eta \in \mathbf{M}_1(E)\}$$

we consider an interacting N-particle system $(\xi_t)_{t\geq 0} = ((\xi_t^1, \ldots, \xi_t^N))_{t\geq 0}$, which is a time-inhomogeneous Markov process on the product space E^N, $N \geq 1$, whose pregenerator acts on functions ϕ belonging to a good domain by

$$\forall\ (x_1, \ldots, x_N) \in E^N, \quad \mathcal{L}_t^{(N)}(\phi)(x_1, \ldots, x_N) \ = \ \sum_{i=1}^N \mathcal{L}_{t,m(x)}^{(i)}(\phi)(x_1, \ldots, x_N) \tag{15}$$

The notation $\mathcal{L}_{t,\eta}^{(i)}$ has been used instead of $\mathcal{L}_{t,\eta}$ when it acts on the i-th variable of $\phi(x_1, \ldots, x_N)$. This abstract and general formulation is well known in mean field IPS literature (the interested reader is for instance referred to [83] and [98] and references therein).
In section 3.2 we will give a more detailed description of (15) when the pregenerators $\mathcal{L}_{t,\eta}$ are defined by (12).

In discrete time settings, a more explicit description of (13) in terms of a two stage genetic type process can already been done. More precisely, using the fact that the mappings Φ_n under study are given by (9) and

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

we see that the resulting motion of the particles is decomposed into two separate mechanisms

$$\xi_{n-1} \xrightarrow{\text{Selection/Updating}} \widehat{\xi}_{n-1} \xrightarrow{\text{Mutation/Prediction}} \xi_n$$

These mechanisms can be modelled as follows:

Selection/Updating:

$$P(\widehat{\xi}_{n-1} \in dx \,|\, \xi_{n-1} = z) = \prod_{p=1}^N \sum_{i=1}^N \frac{g_n(z^i)}{\sum_{j=1}^N g_n(z^j)}\,\delta_{z^i}(dx^p).$$

Mutation/Prediction:

$$P(\xi_n \in dz \,|\, \widehat{\xi}_{n-1} = x) = \prod_{p=1}^N K_n(x^p, dz^p). \tag{16}$$

Thus, we see that the particles move according to the following rules. In the selection transition, one updates the positions in accordance with the fitness functions $\{g_n;\ n \geq 1\}$ and the current configuration. More precisely, at each time $n \geq 1$, each particle examines the system $\xi_{n-1} = \big(\xi_{n-1}^1, \ldots, \xi_{n-1}^N\big)$ and chooses randomly a site

ξ_{n-1}^i, $1 \le i \le N$, with a probability which depends on the entire configuration ξ_{n-1} and given by

$$\frac{g_n\left(\xi_{n-1}^i\right)}{\sum_{j=1}^N g_n\left(\xi_{n-1}^j\right)}.$$

This mechanism is called the selection/updating transition as the particles are selected for reproduction, the most fit individuals being more likely to be selected. In other words, this transition allows particles to give birth to some particles at the expense of light particles which die.
The second mechanism is called mutation/prediction since at this step each particle evolves randomly according to a given transition probability kernel.

The preceding scheme is clearly a system of interacting particles undergoing adaptation in a time non-homogeneous environment represented by the fitness functions $\{g_n;\, n \ge 1\}$. Roughly speaking the natural idea is to approximate the two step transitions (10) of the system (8) by a two step Markov chain taking values in the set of finitely discrete probability measures with atoms of size some integer multiple of $1/N$. Namely, we have that

$$\eta_{n-1}^N \stackrel{\text{def.}}{=} \frac{1}{N}\sum_{i=1}^N \delta_{\xi_{n-1}^i} \xrightarrow{\text{Selection}} \widehat{\eta}_{n-1}^N \stackrel{\text{def.}}{=} \frac{1}{N}\sum_{i=1}^N \delta_{\widehat{\xi}_{n-1}^i} \xrightarrow{\text{Mutation}} \eta_n^N = \frac{1}{N}\sum_{i=1}^N \delta_{\xi_n^i}.$$

As it will be obvious in subsequent sections the continuous time IPS model is again a genetic type model involving selection and mutation transitions.

Another interest feature of the IPS defined by (13) and (22) is that they can be used to approximate the Feynman-Kac formulae (2) and (4). One of the best ways for introducing the corresponding particle approximating models is through the following observation. In discrete time and on the basis of the definition of the distributions $\{\eta_n, \gamma_n \,;\, n \ge 0\}$ we have that

$$\forall n \ge 1\,, \qquad \eta_{n-1}(g_n) = \frac{\gamma_{n-1}(g_n)}{\gamma_{n-1}(1)} = \frac{\gamma_n(1)}{\gamma_{n-1}(1)}$$

Therefore for any $f \in \mathcal{B}_{\mathrm{b}}(E)$ and $n \ge 0$

$$\gamma_n(f) = \gamma_n(1)\,\eta_n(f) \qquad \text{with} \qquad \gamma_n(1) = \prod_{m=1}^n \eta_{m-1}(g_m)$$

(with the convention $\prod_\emptyset = 1$). Taking in consideration the above formula the natural particle approximating scheme for the "unnormalized" distributions $\{\gamma_n \,;\, n \ge 0\}$ is simply given by

$$\gamma_n^N(f) \stackrel{\text{def.}}{=} \gamma_n^N(1)\,\eta_n^N(f) \qquad \text{with} \qquad \gamma_n^N(1) \stackrel{\text{def.}}{=} \prod_{m=1}^n \eta_{m-1}^N(g_m) \tag{17}$$

Similarly, in continuous time settings we have that

$$\eta_t(U_t) = \frac{d}{dt}\log E\left[\exp\int_0^t U_s(X_s)\,ds\right]$$

in the Radon-Nykodim sense and a.s. for $t \geq 0$. Thus, one gets

$$\gamma_t(f) = \gamma_t(1)\, \eta_t(f) \qquad \text{with} \qquad \gamma_t(1) = \exp \int_0^t \eta_s(U_s)\, ds$$

Therefore the natural particle approximating model for the flow of "un-normalized" distributions is given by

$$\gamma_t^N(f) = \gamma_t^N(1)\, \eta_t^N(f) \qquad \text{with} \qquad \gamma_t^N(1) = \exp \int_0^t \eta_s^N(U_s)\, ds \tag{18}$$

where η_t^N is again the empirical measure of the system ξ_t.

1.4 Outline and Contents

These notes are essentially divided into three main parts devoted respectively to discrete time and continuous time models and applications to nonlinear filtering problems.

The first part (section 2) concerns discrete time models. Most of the material presented in this section is taken from the papers [25], [35, 37, 36] and [42].

Section 2.1 focuses on the standard properties of the limiting process needed in the further development of section 2.2. In the preliminary subsection 2.1.1 we give a brief exposition of basic terminologies and properties of the limiting process. In subsection 2.1.2 we discuss the delicate and important problem of the asymptotic stability of the nonlinear semigroup associated with the limiting process. These properties are treated under several type of hypothesis on the functions $\{g_n \ ; \ n \geq 1\}$ and on the transition semigroup associated with the Markov process X. We will also connect in section 2.2.3 these stability properties with the study of the long time behavior of the IPS approximating models.
In nonlinear filtering settings these properties are also essential in order to check whether or not the so-called nonlinear filtering equation forgets erroneous initial conditions. Applications to discrete time filtering problems are discussed in section 5.4.

Section 2.2 is concerned with the asymptotic behavior of the IPS models as the number of particles tends to infinity. This section is essentially divided into four parts. Each of them is devoted to a specific technique or notion to analyze the convergence of the IPS models. This section covers $\mathbb{L}^p$-mean error bounds and uniform convergence results with respect to the time parameter as well as functional central limit theorems, a Glivenko-Cantelli Theorem and large deviations principles.
The quick sketch of the contents of this section will be developed more fully in the preliminary and introductive section 2.2.1.

In section 3 we propose to extend some of the above results to continuous time models. This analysis involves different techniques and it is far from being complete. Among these techniques two are of great importance: semigroup techniques and limit theorems for processes.

In the preliminary section 3.1 we discuss the main hypothesis on the Feynman-Kac formula needed in the further development of section 3.3. In the first subsections 3.1.1 and 3.1.2 we present several regularity assumptions on the fitness functions $\{U_t\ ;\ t \geq 0\}$ and on the Markov process $\{X_t\ ;\ t \geq 0\}$. Section 3.1.3 is concerned with extending the techniques presented in section 2.1.2 to derive stability properties of the Feynman-Kac limiting process. These results will be important later to prove uniform convergence results with respect to the time parameter.

In section 3.2 we give a rigorous mathematical model for the continuous time IPS approximating scheme.

The asymptotic behavior of these IPS numerical schemes as the number of particles tends to infinity is discussed in section 3.3. In the first subsection 3.3.1 we propose $\mathbb{L}^p$-mean errors as well as uniform results. In the last subsection 3.3.2 we prove a functional central limit theorem. Many important problems such as that of the fluctuations and large deviations for the empirical measures on path space remain to be answered.

In view of their kinship to discrete time models the results presented in section 3 can be seen as a good departure to develop a fairly general methodology to study the same asymptotic properties as we did for the discrete time IPS models.

To guide the reader we give in all the development of section 2 and section 3 several comments on the assumptions needed in each specific situation. In section 5 we give detailed and precise examples for each particular assumption.
The applications of the previous results to nonlinear filtering problems are given in section 4. We will discuss continuous time as well as time discretizations and discrete time filtering problems.

The study of the asymptotic stability properties of the limiting system and the investigations of $\mathbb{L}^p$-mean errors, central limit theorems and large deviation principles will of course require quite specific tools. To each of these approaches and techniques corresponds an appropriate set of conditions on the transition probability kernels $\{K_n\ ;\ n \geq 1\}$ and on the so-called fitness functions $\{g_n\ ;\ n \geq 1\}$.

We have tried to present these conditions and to give results at a relevant level of generality so that each section can be read independently of each other. The reader only interested in IPS and BIPS approximating models is recommended to consult section 1.3.2 and section 2.3 as well as section 3 for continuous time models.

The $\mathbb{L}^p$-mean errors presented in section 2.2.2 as well as the central limit theorems for processes exposed in section 3.3.2 are only related to the dynamics structure of the limiting system studied in section 2.1.1. Section 2.2.4 and section 2.2.5 complement previous results of section 2.2.2 and section 2.2.3 by providing asymptotic but precise estimates of $\mathbb{L}^p$-mean errors and exponential rates.

The specific conditions needed in each section are easy to interpret. Furthermore they can be in surprisingly many circumstances be connected one each other. For all these reasons we have no examples in these sections. The interplay and connections between the employed conditions will be described in full details in section 5. We will also connect them to classical examples arising in nonlinear filtering literature. We hope that this choice of presentation will serve our reader well.

2 The Discrete Time Case

2.1 Structural Properties of the Limiting Process.

The investigations of law of large numbers, fluctuations and large deviation principles require quite specific mathematical tools. We shall see in the forthcoming development that these properties are also strongly related to the dynamics structure of the limiting measure valued process (8).
In this section we introduce some basic terminology and a few results on the dynamics structure and the stability properties of (8). In our study a dynamical system is said to be asymptotically stable when its long time behavior does not depends on its initial condition.

In section 2.2 we will prove that the asymptotic stability of the limiting system (8) is a sufficient condition for the uniform convergence of the density profiles of the IPS.

Asymptotic stability properties are also particularly important in filtering settings mainly because the initial law of the signal is usually unknown. In this situation it is therefore essential to check whether the optimal filter is asymptotic stable otherwise all approximating numerical schemes will almost necessarily fails to describe the real optimal filter with the exact initial data.

The genetic type scheme (13) is also used as a random search procedure in numerical function optimization. As for most of stochastic search algorithms, a related question which is of primary interest is to check that their long time behavior does not depend on its initial value.

The study of the long time behavior of the nonlinear filtering equation has been started in [77, 78, 102, 103]. These papers are mainly concerned with the existence of invariant probability measures for the nonlinear filtering equation and do not discuss asymptotic stability properties. A first attempt in this direction was done in [88]. The authors use the above results to prove that the optimal filter "forgets" any erroneous initial condition if the unknown initial law of the signal is absolutely continuous with respect to this new starting point. In the so-called linear-Gaussian situation the optimal filter is also known to be exponentially asymptotically stable under some suitable detectability and observability assumptions (see for instance [88]).
In [5, 48] the authors employ Lyapunov exponent and Hilbert projective techniques to prove exponential asymptotic stability for finite state nonlinear filtering problems. More recently an exponentially asymptotic stability property has been obtained in [16]

for real transient signals, linear observations and bounded observation noise. Here again the proof entirely relies on using Hilbert's projective metrics and showing that the updating transition is a strict contraction with respect to this metric. As pointed out by the authors in [16] the drawback of this metric is its reliance on the assumption of bounded observation noise and does not apply when the optimal filter solution has not bounded support.
Another approach based on the use of Hilbert projective metric to control the logarithmic derivatives of Zakai's kernel is described in [26].
In [4] the authors extend their earlier Lyapunov exponent techniques to Polish-valued signals. Here again the technique consists in evaluating the rate of contraction in the projective Hilbert metric under a mixing type condition on the Markov process X. In discrete time settings this assumption is that the transition probability kernel K_n is homogeneous with respect to time (that is $K_n = K$) and it satisfies the following condition.

$(\mathcal{K})_\varepsilon$ ***There exist a reference probability measure*** $\lambda \in \mathbf{M}_1(E)$ ***and a positive number*** $\epsilon \in (0,1]$ ***so that*** $K(x, \bullet) \sim \lambda$ ***for any*** $x \in E$ ***and***

$$\epsilon \leq \frac{dK(x, \bullet)}{d\lambda} \leq \frac{1}{\epsilon}$$

Here we have chosen to present the novel approach introduced in [36, 38] and further developed in [43]. It is based on the powerful tools developed by R.L. Dobrushin to study central limit theorems for nonstationnary Markov chains [53]. We believe that this approach is more transparent than the previous ones and it also allows to relax considerably the assumption $(\mathcal{K})_\varepsilon$. The continuous time version of this approach will be given in section 3.1.3.

2.1.1 Dynamics Structure

Let us introduce some additional notations. Let $\{Q_{p,n}\ ;\ 0 \leq p \leq n\}$ be the time-inhomogeneous semigroup defined by the relations

$$Q_{p,n} = Q_{p+1} \ldots Q_{n-1} Q_n \quad \text{where} \quad Q_n(f)(x) \stackrel{\text{def.}}{=} g_n(x)\ (K_n f)(x)$$

for any $f \in \mathcal{B}_\mathrm{b}(E)$, $0 \leq p \leq n$ and with the convention $Q_{n,n} = \mathrm{Id}$. It is transparent from the definition of "un-normalized" distributions $\{\gamma_n\ ;\ n \geq 0\}$ that

$$\forall\ 0 \leq p \leq n, \qquad \gamma_n \quad = \quad \gamma_p Q_{p,n}$$

Similarly, we denote by $\{\Phi_{p,n}\ ;\ 0 \leq p \leq n\}$ the composite mappings

$$\Phi_{p,n} = \Phi_n \circ \Phi_{n-1} \circ \ldots \circ \Phi_{p+1}$$

with the convention $\Phi_{n,n} = \mathrm{Id}$. A clear backward induction on the parameter p shows that the composite mappings $\{\Phi_{p,n}\ ;\ 0 \leq p \leq n\}$ have the same form as the one step mappings $\{\Phi_n\ ;\ n \geq 0\}$. This is the content of next lemma.

Lemma 2.1 ([30]) *For any $0 \leq p \leq n$ and $f \in \mathcal{B}_{\mathrm{b}}(E)$ we have that*

$$\Phi_{p,n}(\mu)(f) = \frac{\mu(g_{p,n}\ K_{p,n}(f))}{\mu(g_{p,n})} \tag{19}$$

where

$$K_{p,n}(f) = \frac{Q_{p,n}(f)}{Q_{p,n}(1)} \quad \text{and} \quad g_{p,n} = Q_{p,n}(1)$$

The fitness functions $\{g_{p,n}\ ;\ 0 \leq p \leq n\}$ and the transition probability kernels $\{K_{p,n}\ ;\ 0 \leq p \leq n\}$ satisfy the backward formulae

$$K_{p-1,n}(f) = \frac{K_p(g_{p,n}\ K_{p,n}(f))}{K_p(g_{p,n})}, \quad g_{p-1,n} = g_p\ .\ K_p(g_{p,n}) \tag{20}$$

for any $f \in \mathcal{B}_{\mathrm{b}}(E)$, $1 \leq p \leq n$ and conventions $g_{n,n} = 1$, $K_{n,n} = \mathrm{Id}$.

Lemma 2.2 *For any $0 \leq p \leq n$, $\mu, \nu \in \mathbf{M}_1(E)$ and $f \in \mathcal{B}_{\mathrm{b}}(E)$*

$$\begin{aligned}
&\Phi_{p,n}(\mu)(f) - \Phi_{p,n}(\nu)(f) \\
&= \frac{\nu(g_{p,n})}{\mu(g_{p,n})}\ \mu\Big[\frac{g_{p,n}}{\nu(g_{p,n})} K_{p,n}\left(f - \Phi_{p,n}(\nu)(f)\right)\Big] \\
&= \frac{1}{\nu(g_{p,n})}\Big[(\mu(f_{p,n}) - \nu(f_{p,n})) + \Phi_{p,n}(\mu)(f)\ (\nu(g_{p,n}) - \mu(g_{p,n}))\Big]
\end{aligned} \tag{21}$$

where

$$f_{p,n} \stackrel{\text{def}}{=} g_{p,n}\ .\ K_{p,n}(f).$$

Unless otherwise stated we will always assume that the fitness functions $\{g_n\ ;\ n \geq 1\}$ are continuous and satisfy the following condition.

$(\mathcal{G})$ ***For any*** $n \geq 1$***, there exists*** $a_n \in [1, \infty)$ ***with***

$$\forall\ x \in E, \forall\ n \geq 1, \qquad \frac{1}{a_n}\ \leq\ g_n(x)\ \leq\ a_n \tag{22}$$

Under this assumption and using the above notations we notice that

- For any $x \in E$ and $0 \leq p \leq n$

$$\frac{1}{a_{p,n}} \leq g_{p,n}(x) \leq a_{p,n} \quad \text{where} \quad a_{p,n} = \prod_{q=p+1}^{n} a_q$$

- For any $f \in \mathcal{B}_{\mathrm{b}}(E)$ and $0 \leq p \leq n$

$$\|Q_{p,n}(f)\| \leq a_{p,n}\ \|f\| \qquad a_{0,n}^{-1} \leq \gamma_n(1) \leq a_{0,n}$$

In order that our paper is both broad and has some coherence we have chosen to concentrate on the analysis of the particle density profiles $\{\eta_n^N \; ; \; n \geq 0\}$ and the corresponding limiting system $\{\eta_n \; ; \; n \geq 0\}$.

Most of the results presented in section 2.1.2 and section 2.2 can be used to obtain analog results for $\{\widehat{\eta}_n^N \; ; \; n \geq 0\}$ and the flow $\{\widehat{\eta}_n \; ; \; n \geq 0\}$. The interested reader is recommended to use the decompositions

$$\Phi_n(\mu)f - \Phi_n(\nu)f = \frac{1}{\nu(g_n)} \left([\mu(g_n K_n f) - \nu(g_n K_n f)] + (\Phi_n(\mu)f) \, [\nu(g_n) - \mu(g_n)] \right)$$

and

$$\widehat{\eta}_{n-1}^N f - \widehat{\eta}_{n-1} f = \left[\widehat{\eta}_{n-1}^N f - \Psi_n(\eta_{n-1}^N) f\right] + \left[\Psi_n(\eta_{n-1}^N) f - \Psi_n(\eta_{n-1}) f\right]$$

and to recall that $\widehat{\eta}_{n-1}^N$ is the empirical measure associated with N conditionally independent random variables with common law $\Psi_n(\eta_{n-1}^N)$.
There also exists a different strategy to approximate the flow of distributions $\{\widehat{\eta}_n \; ; \; n \geq 0\}$. This alternative approach is simply based on the observation that the dynamics structure of the latter is again defined by a recursion of the form (8). More precisely, in view of (10) we have that

$$\forall \, n \geq 1, \qquad \widehat{\eta}_n \;\; = \;\; \widehat{\Phi}_n(\widehat{\eta}_{n-1}) \tag{23}$$

where $\widehat{\Phi}_n(\eta) = \widehat{\Psi}_n(\eta)\widehat{K}_n$ and for any $f \in \mathcal{B}_{\mathrm{b}}(E)$,

$$\widehat{\Psi}_n(\eta)(f) \stackrel{\text{def.}}{=} \frac{\eta(\widehat{g}_n f)}{\eta(\widehat{g}_n)}, \qquad \widehat{K}_n f \stackrel{\text{def.}}{=} \frac{K_n(g_{n+1} f)}{K_n(g_{n+1})}, \qquad \widehat{g}_n \stackrel{\text{def.}}{=} K_n(g_{n+1}).$$

Noticing that the new fitness functions $\{\widehat{g}_n \; ; \; n \geq 0\}$ again satisfy $(\mathcal{G})$ we see that the study of (23) and its corresponding IPS approximating model is reduced to the former by replacing the fitness functions $\{g_n \; ; \; n \geq 1\}$ and the transitions $\{K_n \; ; \; n \geq 0\}$ by $\{\widehat{g}_n \; ; \; n \geq 0\}$ and $\{\widehat{K}_n \; ; \; n \geq 1\}$ and the initial data η_0 by the distribution $\widehat{\eta}_0$.

Nevertheless the above description shows that the formulation of the new fitness functions involves integrations over the whole state space E. In practice these integrals are not known exactly and another level of approximation is therefore needed. We also remark that we use the well known "rejection/acceptation" sampling method to produce transitions of the IPS approximating scheme. But, when the constants $\{a_n \; ; \; n \geq 1\}$ are too large it is well known that the former sampling technique is "time-consuming" and not really efficient.

We have chosen to restrict our attention to the IPS approximating model $\{\eta_n^N \; ; \; n \geq 0\}$ for several reasons:

- First of all it is defined as a two stage transition which can serve as a Markov model for classical genetic algorithms arising in biology and nonlinear estimation problems.

- Secondly, some results such as the fluctuations on path space for the IPS $\{\widehat{\xi}_n \; ; \; n \geq 0\}$ are not yet available.

- Another important reason is that the limiting system $\{\eta_n \; ; \; n \geq 0\}$ is exactly the discrete time approximating model of the Kushner-Stratonovitch equation (see for instance [24] and references therein as well as section 4.3 page 122).

- Finally the evolution equation of the distributions $\{\widehat{\eta}_n \; ; \; n \geq 0\}$ is a special case of (8) so that the results on (8) will also include those on (23).

2.1.2 Asymptotic Stability

In this section we discuss the asymptotic stability properties of (8). The main difficulty lies in the fact that (8) is a two stage process and it may have completely different kinds of long time behavior.

For instance, if the fitness functions are constant functions then (8) is simply based on prediction transitions. In this situation the recursion (8) describes the time evolution of the distributions of the Markov process $\{X_n \; ; \; n \geq 0\}$. In this very special case the theory of Markov processes and stochastic stability can be applied.

On the other hand, if the transition probability kernels $\{K_n \; ; \; n \geq 1\}$ are trivial, that is $K_n = \mathrm{Id}$, $n \geq 1$, then (8) is only based on updating transitions. In this case its long time behavior is strongly related on its initial value. For instance, if $g_n = \exp(-U)$, for some $U : E \to \mathbb{R}_+$, then for any bounded continuous function $f : E \to \mathbb{R}_+$ with compact support

$$\eta_n(f) = \frac{\eta_0(f\, e^{-nU})}{\eta_0(e^{-nU})} \xrightarrow[n \to \infty]{} \frac{\eta_0(f\, 1_{U^\star})}{\eta_0(U^\star)}$$

where $U^\star \stackrel{\text{def.}}{=} \{x \in E \; ; \; U(x) = \mathrm{essinf}_{\eta_0} U\}$ (at least if $\eta_0(U^*) > 0$).

If we write $\mathbf{M}_0(E) \subset \mathbf{M}(E)$ the subset of measures μ such that $\mu(E) = 0$ then any transition function $T(x, dz)$ on E maps $\mathbf{M}_0(E)$ into $\mathbf{M}_0(E)$. We recall that its norm is given by

$$\beta(T) \stackrel{\text{def.}}{=} \sup_{\mu \in \mathbf{M}_0(E)} \frac{\|\mu T\|_{tv}}{\|\mu\|_{tv}} = \sup_{\mu,\nu \in \mathbf{M}_1(E)} \frac{\|\mu T - \nu T\|_{tv}}{\|\mu - \nu\|_{tv}}$$

The quantity $\beta(T)$ is a measure of contraction of the total variation distance of probability measures induced by T. It can also be defined as

$$\beta(T) = \sup_{x,y \in E} \|\delta_x T - \delta_y T\|_{tv} = 1 - \alpha(T) \tag{24}$$

The quantity $\alpha(T)$ is called the Dobrushin ergodic coefficient of T defined by the formula

$$\alpha(T) = \inf \sum_{i=1}^{m} \min\left(T(x, A_i), T(z, A_i)\right) \tag{25}$$

where the infimum is taken over all $x, z \in E$ and all resolutions of E into pairs of non-intersecting subsets $\{A_i \; ; \; 1 \leq i \leq m\}$ and $m \geq 1$ (see for instance [53]). Our analysis is based on the following observation: In view of (20) the Markov transition probability kernels $\{K_{n,p}\ ;\ 0 \leq p \leq n\}$ are composite operators of time-inhomogeneous but linear Markov operators. More precisely it can be checked directly from (20) that

$$K_{p-1,n} = S_p^{(n)} K_{p,n} = S_p^{(n)} S_{p+1}^{(n)} \ldots S_{n-1}^{(n)} S_n^{(n)}, \qquad S_p^{(n)} f = \frac{K_p(g_{p,n}\ f)}{K_p(g_{p,n})} \tag{26}$$

The usefulness of Dobrushin ergodic coefficient in the study of the asymptotic stability of the nonlinear dynamical system (8) can already been seen from the next theorem.

Theorem 2.3 *Assume that the fitness functions $\{g_n\ ;\ n \geq 1\}$ is a sequence of bounded and positive functions on E. For any $n \geq p$ we have that*

$$\|\Phi_{p,n}(\mu) - \Phi_{p,n}(\nu)\|_{\mathrm{tv}} \leq\ \beta\left(K_{p,n}\right)\ \|\Psi_{p,n}(\mu) - \Psi_{p,n}(\nu)\|_{\mathrm{tv}} \tag{27}$$

with

$$\forall \mu \in \mathbf{M}_1(E)\,, \forall f \in \mathcal{B}_{\mathrm{b}}(E)\,, \qquad \Psi_{p,n}(\mu)(f) = \frac{\mu(g_{p,n}\ f)}{\mu(g_{p,n})}$$

and

$$\sup_{\mu,\nu} \|\Phi_{p,n}(\mu) - \Phi_{p,n}(\nu)\|_{\mathrm{tv}} = \beta\left(K_{p,n}\right) \leq \prod_{q=1}^{n-p} \left[1 - \alpha(S_{p+q}^{(n)})\right] \tag{28}$$

Assume that the transition probability kernels $\{K_n\ ;\ n \geq 1\}$ satisfy the following condition.

$(\mathcal{K})_1$ ***For any time $n \geq 1$, there exist a reference probability measure $\lambda_n \in \mathbf{M}_1(E)$ and a positive number $\epsilon_n \in (0,1]$ so that $K_n(x, \bullet) \sim \lambda_n$ for any $x \in E$ and***

$$\epsilon_n \leq \frac{dK_n(x, \bullet)}{d\lambda_n} \leq \frac{1}{\epsilon_n}$$

Then, we have for any $\mu, \nu \in \mathbf{M}_1(E)$

$$\begin{aligned}
\sum_{n\geq 1} \epsilon_n^2 = \infty \quad &\Longrightarrow \quad \lim_{n\to\infty} \|\Phi_{0,n}(\mu) - \Phi_{0,n}(\nu)\|_{\mathrm{tv}} = 0 \\
\lim_{n\to\infty} \frac{1}{n} \sum_{p=1}^{n} \epsilon_p^2 \stackrel{\mathrm{def.}}{=} \epsilon^2 > 0 \quad &\Longrightarrow \quad \limsup_{n\to\infty} \frac{1}{n} \log \|\Phi_{0,n}(\mu) - \Phi_{0,n}(\nu)\|_{\mathrm{tv}} \leq -\epsilon^2 < 0 \\
\inf_{n\geq 1} \varepsilon_n \stackrel{\mathrm{def.}}{=} \epsilon > 0 \quad &\Longrightarrow \quad \frac{1}{T} \log \sup_{p\geq 0} \|\Phi_{p,p+T}(\mu) - \Phi_{p,p+T}(\nu)\|_{\mathrm{tv}} < -\epsilon^2
\end{aligned} \tag{29}$$

Proof:(*Sketch*) Since, for any $\mu \in \mathbf{M}_1(E)$ we have $\Phi_{p,n}(\mu) = \Psi_{p,n}(\mu) K_{p,n}$ then, using (24) and (26) one proves (28). Under $(\mathcal{K})_1$, the second part of the theorem is a consequence of the fact that $\alpha(S_p^{(n)}) \geq \epsilon_p^2$, for any $0 \leq p \leq n$. To prove these lower bounds we use (25) and the definition of the transitions $S_p^{(n)}$. ∎

Remarks 2.4:

- Theorem 2.3 holds for any sequence of fitness functions $\{g_n\ ;\ n \geq 1\}$. One also observe that the last assertion in the Theorem 2.3 holds true when the transition probability kernels $\{K_n\ ;\ n \geq 1\}$ satisfy $(\mathcal{K})_\varepsilon$.
  In contrast to [4] (see for instance Corollary 1 p. 706 in [4]) the exponential asymptotic property (29) is valid for any time parameter and it does not depend on the form neither on the regularity of the fitness functions.

- In time homogeneous settings (that is $K_n = K$, $g_n = g$) the mapping $\Phi_n = \Phi$ is again time homogeneous. If there exists some fixed point $\mu = \Phi(\mu) \in \mathbf{M}_1(E)$. Theorem 2.3 gives several conditions underwhich μ is unique and any solution $\{\eta_n\ ;\ n \geq 0\}$ of the time homogeneous dynamical system associated to Φ converges exponentially fast to μ as the time parameter $n \to \infty$. For instance if $(\mathcal{K})_1$ holds with $\epsilon_n = \epsilon > 0$ then

$$\forall\, \eta \in \mathbf{M}_1(E), \qquad \sup_{p\geq 0} \|\Phi_{p,p+T}(\eta) - \mu\|_{\text{tv}} \ \leq\ e^{-\epsilon^2 . T}$$

- Let us consider the special case where $\epsilon_n = \epsilon_1 . n^{-\beta/2}$ with $\epsilon_1 > 0$ and $\beta \leq 1$. Theorem 2.3 tells us that (8) is asymptotically stable but we don't know if exponential rates are available in this scale.

- If $(\mathcal{K})_1$ is satisfied for some sequence of numbers $\{\epsilon_n\ ;\ n \geq 1\}$ and a collection of distributions $\{\lambda_n\ ;\ n \geq 1\}$ then we also have that

$$\forall\, x, z \in E,\, \forall\, n \geq 1, \qquad \widehat{\epsilon}_n \ \leq\ \frac{d\widehat{K}_n\,(x, .)}{d\widehat{\lambda}_n}(z) \ \leq\ \frac{1}{\widehat{\epsilon}_n}$$

  where $\widehat{\epsilon}_n \stackrel{\text{def.}}{=} \epsilon_n^2$ and $\widehat{\lambda}_n \in \mathbf{M}_1(E)$ is the distribution on E given for any bounded test function f by

$$\widehat{\lambda}_n(f) \stackrel{\text{def.}}{=} \frac{\lambda_n\ (\ g_{n+1}\ f\ )}{\lambda_n\ (\ g_{n+1}\ )}$$

  This observation shows that Theorem 2.3 remains true if we replace the one step mappings $\{\Phi_n\ ;\ n \geq 1\}$ and the numbers $\{\epsilon_n^2\ ;\ n \geq 1\}$ by the mappings $\{\widehat{\Phi}_n\ ;\ n \geq 1\}$ and the numbers $\{\epsilon_n^4\ ;\ n \geq 1\}$. It is also worth to notice that this results is again independent of the fitness functions.

- By definition of the one step mappings $\{\widehat{\Phi}_n\ ;\ n \geq 1\}$ we have the formula

$$\forall\, \mu \in \mathbf{M}_1(E),\, \forall\, n \geq 1, \qquad \widehat{\Phi}_n(\mu) \ =\ \Psi_{n+1}\left(\Phi_{1,n}(\mu K_1)\right)$$

  Thus, by definition of the mappings $\{\Psi_n\ ;\ n \geq 1\}$, if $(\mathcal{G})$ is satisfied then one can check that for any $\mu, \nu \in \mathbf{M}_1(E)$ and $n \geq 1$

$$\left\|\widehat{\Phi}_{0,n}(\mu) - \widehat{\Phi}_{0,n}(\nu)\right\|_{\text{tv}} \leq 2\ a_{n+1}^2 \ \ \|\Phi_{1,n}(\mu K_1) - \Phi_{1,n}(\nu K_1)\|_{\text{tv}}$$

  where $\widehat{\Phi}_{0,n}$ are the composite mappings

$$\widehat{\Phi}_{0,n} = \widehat{\Phi}_n \circ \widehat{\Phi}_{n-1} \circ \ldots \circ \widehat{\Phi}_1$$

Therefore Theorem 2.3 can also be used to derive several conditions for the asymptotic stability of (23).
In contrast to the previous remark the resulting error bound depends on the sets of numbers $\{a_n\ ;\ n \geq 1\}$ and $\{\epsilon_n^2\ ;\ n \geq 1\}$ instead of $\{\epsilon_n^4\ ;\ n \geq 1\}$. This fact is important for instance if $\epsilon_n = \epsilon_1 . n^{-1/2}$ and $\sup_n a_n < \infty$. In this specific situation the estimate resulting from this approach will improve the one we would obtain using the previous remark and it allows to conclude that (23) is asymptotically stable even if $\sum_n \epsilon_n^4 < \infty$.

Next condition relax $(\mathcal{K})_1$.

$(\mathcal{K})_2$ ***For any time*** $p \geq 0$ ***there exist some*** $m \geq 1$, $\lambda_p \in \mathbf{M}_1(E)$ ***and*** $\epsilon_p > 0$ ***such that for any*** $x \in E$

$$\varepsilon_p \leq \frac{dK_p^{(m)}(x, \bullet)}{d\lambda_p} \leq \frac{1}{\varepsilon_p} \qquad \text{where} \quad K_p^{(m)} \stackrel{\text{def.}}{=} K_{p+1} \ldots K_{p+m} \tag{30}$$

Under $(\mathcal{K})_2$ and $(\mathcal{G})$ one can check that for any $0 \leq p+m \leq n$

$$\left(\frac{\varepsilon_p}{a_{p+1,p+m}}\right)^2 \ \alpha\left(K_p\right) \leq \alpha\left(S_p^{(n)}\right) \leq \left(\frac{a_{p+1,p+m}}{\varepsilon_p}\right)^2 \ \alpha\left(K_p\right) \tag{31}$$

We now turn to still another way to produce useful estimates.

By definition of the integral operators $\{S_p^{(n)}\ ;\ 0 \leq p \leq n\}$ we also have for any positive Borel test function $f : E \to \mathbb{R}_+$

$$\frac{1}{a_{p+2,p+m}^2} \ \frac{K_p^{(m)}(g_{p+m,n}f)}{K_p^{(m)}(g_{p+m,n})} \leq S_{p+1}^{(n)} \ldots S_{p+m}^{(n)} f \leq a_{p+2,p+m}^2 \ \frac{K_p^{(m)}(g_{p+m,n}f)}{K_p^{(m)}(g_{p+m,n})}$$

From (30) it is clear that for any $0 < p+m \leq n$

$$\left(\frac{\varepsilon_p}{a_{p+2,p+m}}\right)^2 \leq \alpha\left(S_{p+1}^{(n)} \ldots S_{p+m}^{(n)}\right) \leq \left(\frac{a_{p+2,p+m}}{\varepsilon_p}\right)^2 \tag{32}$$

In contrast to (31) the estimate (32) does not depend anymore on the ergodic coefficients $\{\alpha(K_p)\ ;\ p \geq 1\}$.

Nevertheless the estimate induced by (31) may improve the one we would obtain using (32). For instance, in time homogeneous settings (that is $\varepsilon_p = \varepsilon$, $a_p = a$, $K_p = K$, for all $p \geq 0$) the bounds (32) lead to

$$\left(1 - \alpha\left(S_{p+1}^{(n)} \ldots S_{p+m}^{(n)}\right)\right) \leq 1 - \frac{\varepsilon^2}{a^{2(m-1)}} \leq \exp - \frac{\varepsilon^2}{a^{2(m-1)}} \tag{33}$$

and (31) implies that

$$\left(1 - \alpha\left(S_{p+1}^{(n)} \ldots S_{p+m}^{(n)}\right)\right) \leq \prod_{q=1}^{m} \left(1 - \alpha\left(S_{p+q}^{(n)}\right)\right) \leq \left(1 - \frac{\varepsilon^2}{a^{2m}} \alpha(K)\right)^m$$

and therefore

$$\left(1-\alpha\left(S^{(n)}_{p+1}\ldots S^{(n)}_{p+m}\right)\right)\leq \exp -\left(m\frac{\varepsilon^2}{a^{2m}}\alpha(K)\right) \tag{34}$$

One concludes that (34) improves (33) as soon as $m.\alpha(K)\geq a^2$ and therefore

$$\log\left(1-\alpha\left(S^{(n)}_{p+1}\ldots S^{(n)}_{p+m}\right)\right)\leq -\frac{\epsilon^2}{a^{2m}}\ \max\left(m.\alpha(K),a^2\right)$$

Theorem 2.5 *Assume that* $(\mathcal{G})$ *and* $(\mathcal{K})_2$ *hold for some* $m\geq 1$ *and some sequence of numbers* $\{\epsilon_p\ ;\ p\geq 0\}$. *For any* $\mu,\nu\in\mathbf{M}_1(E)$ *and* $n\geq m$ *we have*

$$\|\Phi_{0,n}(\mu)-\Phi_{0,n}(\nu)\|_{\rm tv}\leq\prod_{p=1}^{n-m}\left(1-\frac{\varepsilon_p^2}{a^2_{p+1,p+m}}\alpha(K_p)\right)$$

and

$$\|\Phi_{0,n}(\mu)-\Phi_{0,n}(\nu)\|_{\rm tv}\leq\prod_{p=0}^{[n/m]-1}\left(1-\frac{\varepsilon^2_{pm}}{a^2_{pm+2,pm+m}}\right)$$

In addition, if $\inf_n\varepsilon_n\stackrel{\rm def.}{=}\varepsilon>0$, $\inf_n\alpha(K_n)\stackrel{\rm def.}{=}\alpha>0$ *and* $\sup_n a_n\stackrel{\rm def.}{=}a<\infty$ *then for any* $u\geq 1$ *and* $T\geq u.m$

$$\frac{1}{T}\log\sup_{p\geq 0}\|\Phi_{p,p+T}(\mu)-\Phi_{p,p+T}(\nu)\|_{\rm tv}<-\frac{1}{v.m}\ \left(\frac{\varepsilon}{a^m}\right)^2\ \max\left(m.\alpha,a^2\right)$$

where $1/u+1/v=1$.

Next we present an easily verifiable sufficient condition for $(\mathcal{K})_2$. It also gives some connections between $(\mathcal{K})_1$ and $(\mathcal{K})_2$. It will be used in section 5 to check that some classes of Gaussian transitions satisfy $(\mathcal{K})_2$ for $m=2$. To clarify the presentation the transition probability kernels K_n here are assumed to be time homogeneous (that is $K_n=K$ for all $n\geq 0$).

$(\mathcal{K})_3$ ***There exist a subset*** $A\in\mathbf{B}(E)$***, a reference probability measure*** $\lambda\in\mathbf{M}_1(E)$ ***and a positive number*** $\epsilon\in(0,1)$ ***such that*** $\lambda(A)\geq\varepsilon$ ***and***

$$\forall\ x\in E,\forall\ z\in A,\qquad \epsilon\ \leq\ \frac{dK(x,\bullet)}{d\lambda}(z)\ \leq\ \frac{1}{\epsilon}$$

In addition there exist a decomposition $A^c=B_1\cup\ldots\cup B_m$***,*** $m\geq 1$ ***and*** $2m$ ***reference probability measures*** $\lambda_1,\ldots,\lambda_m,\gamma_1,\ldots,\gamma_m\in\mathbf{M}_1(E)$ ***such that for any*** $1\leq k\leq m$

$$\forall\ x\in B_k,\forall\ z\in E,\qquad \epsilon\ \leq\ \frac{dK(x,\bullet)}{d\lambda_k}(z)\ \leq\ \frac{1}{\epsilon}$$

and

$$\forall\ x\in E,\forall\ z\in B_k,\qquad \gamma_k(B_k)\geq\varepsilon\qquad\text{and}\qquad\frac{dK(x,\bullet)}{d\gamma_k}(z)\ \geq\ \epsilon$$

Under $(\mathcal{G})$ and $(\mathcal{K})_3$ one concludes that $(\mathcal{K})_2$ holds for $m = 2$ (the last condition above is even useless for that).

Furthermore under $(\mathcal{K})_3$ one can also prove the following bounds

$$\alpha(K) \geq m.\varepsilon^2 \quad \text{and} \quad \alpha\left(S_p^{(n)}\right) \geq \frac{\varepsilon^4}{a_{p+1}^2}$$

Corollary 2.6 *Assume that* $(\mathcal{G})$ *and* $(\mathcal{K})_3$ *hold. Then, for any* $\mu, \nu \in \mathbf{M}_1(E)$ *we have that*

$$\sum_{n\geq 1} a_n^{-2} = \infty \implies \lim_{n\to\infty} \|\Phi_{0,n}(\mu) - \Phi_{0,n}(\nu)\|_{\text{tv}} = 0$$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{p=1}^{n} a_p^{-2} \stackrel{\text{def.}}{=} a^{-2} \implies \limsup_{n\to\infty} \frac{1}{n} \log \|\Phi_{0,n}(\mu) - \Phi_{0,n}(\nu)\|_{\text{tv}} \leq -\frac{\varepsilon^4}{a^2}$$

$$\sup_n a_n = a < \infty \implies \frac{1}{T} \log \sup_{p\geq 0} \|\Phi_{p,p+T}(\mu) - \Phi_{p,p+T}(\nu)\|_{\text{tv}} \leq -\frac{\varepsilon^4}{a^2}$$

A more interesting and stronger result for transition probability kernels $\{K_n \,;\, n \geq 1\}$ which do not satisfy $(\mathcal{K})_2$ is given next. We will use the following condition.

$(\mathcal{KG})$ ***There exists some*** $p_0 \geq 0$ ***such that***

$$\forall\ p \geq p_0, \qquad \delta_p \stackrel{\text{def.}}{=} \inf_{n\geq p} \inf_{x,y\in E} \frac{g_{p,n}(x)}{g_{p,n}(y)} > 0 \tag{35}$$

Let us observe that under $(\mathcal{KG})$ one has a uniform control of the fitness functions $\{g_{p,n} \,;\, 0 \leq p \leq n\}$ in the sense that for any $n \geq p \geq p_0$ and $x, y \in E$

$$\delta_p \leq \frac{g_{p,n}(x)}{g_{p,n}(y)} \leq \frac{1}{\delta_p}$$

In contrast to previous conditions $(\mathcal{KG})$ depends on both the fitness functions $\{g_n \,;\, n \geq 1\}$ and on the transition probability kernels $\{K_n \,;\, n \geq 1\}$. Note that $(\mathcal{KG})$ is clearly satisfied when the fitness functions are constant or more generally if the fitness functions become constant after a given step p_0 (that is $g_p = 1$ for any $p \geq p_0$). In the same vein $(\mathcal{KG})$ is satisfied if we are allowed to choose the constants $\{a_n \,;\, n \geq 1\}$ such that $\sum_{n\geq 1} \log a_n < \infty$. In this situation we would have

$$\delta_p \geq \exp - \left(2 \sum_{q\geq 1} \log a_{p+q}\right) > 0.$$

Another way of viewing condition $\sum_{n\geq 1} \log a_n < \infty$ is to say that the sequence of functions $\{g_n \,;\, n \geq 1\}$ satisfy

$$\sum_{n\geq 1} \sup_{x\in E} |\log g_n(x)| < \infty$$

which clearly implies that g_n tends to the unit function 1 as n tends to infinity.

It is also noteworthy that if $(\mathcal{KG})$ holds for $p_0 = 0$ then condition $(\mathcal{G})$ is directly satisfied. More precisely in this situation we would get the bounds

$$\forall\, x, y \in E,\ \forall\, p \geq 0, \qquad \delta_p \ \leq\ \frac{g_{p+1}(x)}{g_{p+1}(y)}\ \leq\ \frac{1}{\delta_p}$$

To see that $(\mathcal{KG})$ also relax $(\mathcal{K})_2$ it suffices to note that $(\mathcal{K})_2$ implies that for any $0 \leq p+m \leq n$

$$\frac{g_{p,n}(x)}{g_{p,n}(y)} \geq a_{p+1,p+m}^{-2}\ \frac{K_{p+1}\ldots K_{p+m}\,(g_{p+m,n})\,(x)}{K_{p+1}\ldots K_{p+m}\,(g_{p+m,n})\,(y)} \geq \epsilon_p^2\ a_{p+1,p+m}^{-2} > 0$$

Since for any $p \leq n \leq p+m$ and $x, y \in E$

$$\frac{g_{p,n}(x)}{g_{p,n}(y)} \geq a_{p+1,p+m}^{-2}$$

we finally have the lower bounds

$$\forall\, p \geq 0, \qquad \delta_p \ \geq\ \epsilon_p^2\ a_{p+1,p+m}^{-2}\ >\ 0$$

Under $(\mathcal{KG})$ and using (27) and the decomposition

$$\Psi_{p,n}(\mu)(f) - \Psi_{p,n}(\nu)(f) = \frac{\nu(g_{p,n})}{\mu(g_{p,n})}\ \mu\left[\frac{g_{p,n}}{\nu(g_{p,n})}\ \left(f - \Psi_{p,n}(\nu)(f)\right)\right]$$

we also have that

$$\forall n \geq p \geq p_0\,, \qquad \|\Psi_{p,n}(\mu) - \Psi_{p,n}(\nu)\|_{\mathrm{tv}} \leq \frac{2}{\delta_p^2}\ \|\mu - \nu\|_{\mathrm{tv}}$$

Using the same lines of reasoning as before one can prove that for any Borel test function $f : E \to \mathbb{R}_+$ and $m \geq 1$ and $n \geq p+m$

$$\begin{aligned} S_{p+1}^{(n)} \ldots S_{p+m}^{(n)} f \ &\geq\ \delta_{p+1} \ldots \delta_{p+(m-1)}\ \frac{K_p^{(m)}(g_{p+m,n} f)}{K_p^{(m)}(g_{p+m,n})} \\ &\geq\ \delta_{p+1} \ldots \delta_{p+m}\ \ K_p^{(m)} f \end{aligned}$$

from which one deduces the following deeper result.

Theorem 2.7 *If $(\mathcal{KG})$ is satisfied for some parameter $p_0 \geq 0$ then we have for any $n \geq p \geq p_0$*

$$\|\Phi_{p,n}(\mu) - \Phi_{p,n}(\nu)\|_{\mathrm{tv}} \leq \frac{2}{\delta_p^2}\ \beta\left(K_{p,n}\right)\ \|\mu - \nu\|_{\mathrm{tv}}$$

and for any $m \geq 1$

$$\sup_{\mu,\nu} \|\Phi_{0,n}(\mu) - \Phi_{0,n}(\nu)\|_{\mathrm{tv}} \leq \prod_{q=0}^{[(n-p_0)/m]-1} \left(1 - \delta_{p_0+q.m}^{(m)}\ \alpha\left(K_{p_0+q.m}^{(m)}\right)\right) \qquad (36)$$

where $[a]$ *denotes the integer part of* $a \in \mathbb{R}$, $\delta_p^{(m)} \stackrel{\text{def.}}{=} \delta_{p+1} \ldots \delta_{p+m}$ *and*

$$K_p^{(m)} \stackrel{\text{def.}}{=} K_{p+1} \ldots K_{p+m}.$$

In addition if $p_0 = 0$ *and* $\inf_p \delta_p \stackrel{\text{def.}}{=} \delta > 0$ *then for* $m = 1$ *(36) leads to*

$$\frac{1}{n} \log \sup_{\mu,\nu} \|\Phi_{0,n}(\mu) - \Phi_{0,n}(\nu)\|_{\text{tv}} \leq -\frac{\delta}{n} \sum_{p=1}^{n} \alpha\left(K_p\right) \tag{37}$$

Remarks 2.8:

- The bound (36) is sharp in the sense that if the fitness functions g_n are constant then one may choose $m = n, p_0 = 0$. In this situation (36) reads

$$\begin{aligned} \sup_{\mu,\nu} \|\mu K_1 \ldots K_n - \nu K_1 \ldots K_n\|_{\text{tv}} &\leq 1 - \alpha\left(K_1 \ldots K_n\right) = \beta\left(K_1 \ldots K_n\right) \\ &= \sup_{x,y} \|\delta_x K_1 \ldots K_n - \delta_y K_1 \ldots K_n\|_{\text{tv}} \end{aligned}$$

- In view of (37) and under $(\mathcal{KG})$, condition $\sum_{n\geq 1} \alpha\left(K_n\right) = \infty$ is a sufficient condition for the asymptotic stability of the nonlinear semigroup

$$\{\Phi_{p,n}\,;0 \leq p \leq n\}$$

  This condition is a familiar necessary and sufficient condition for the temporally inhomogeneous semigroup associated with the transitions $\{K_n\ ;\ n \geq 1\}$ to be strongly ergodic (see for instance [53], part I, p. 76).

- In nonlinear filtering settings condition $\sum_n \log a_n < \infty$ is related to the form of the observation noise source. An example of observation process satisfying this condition will be given in the end of section 5. Roughly speaking the assumptions $(\mathcal{K})_i$, $i = 1, 2, 3$ say that the signal process is sufficiently mixing and condition $\sum_n \log a_n < \infty$ says that the observation process is sufficiently noisy.

2.2 Asymptotic Behavior of the Particle Systems

2.2.1 Introduction

In this section we investigate the asymptotic behavior of the IPS as the number of particles tends to infinity. In the first subsection 2.2.2 we discuss $\mathbb{L}^p$-mean error bounds and a uniform convergence theorem with respect to the time parameter. A Glivenko-Cantelli Theorem is described in section 2.2.3. Subsection 2.2.4 is concerned with fluctuations of IPS. In the last subsection 2.2.5 we present large deviation principles.

All of the above properties will of course be stated under appropriate regularity conditions on the transition probability kernels $\{K_n\ ;\ n \geq 1\}$ and on the fitness functions $\{g_n\ ;\ n \geq 1\}$.

Assumption $(\mathcal{G})$ is the only assumption needed on the fitness functions. Unless otherwise stated we will always assume that $(\mathcal{G})$ holds.

This condition has a clear interpretation in nonlinear filtering settings (see for instance [33, 35, 36, 37] and section 5). It can be regarded as a technical assumption and several results presented here can be proved without this condition. To illustrate this remark we will give in the beginning of section 2.2.2 a very basic convergence result that does not depend on $(\mathcal{G})$. To guide the reader we now give some comments on the assumptions needed on the transition probability kernels $\{K_n \,;\, n \geq 1\}$.

The uniform convergence result presented in section 2.2.2 is based on the asymptotic stability properties of the limiting measure valued process (8) we have studied in section 2.1.2. This part will then be related to assumptions $(\mathcal{K})_1$,$(\mathcal{K})_2$ and $(\mathcal{K})_3$.

The Glivenko-Cantelli and Donsker Theorems presented in section 2.2.3 and section 2.2.4 extend the corresponding statements in the classical theory of empirical processes. They are simply based on $(\mathcal{G})$. The idea here is to consider a given random measure μ^N as a stochastic process indexed by a collection $\mathcal{F}$ of measurable functions $f : E \to \mathbb{R}$. If μ^N is an empirical measure then, the resulting $\mathcal{F}$-indexed collection $\{\mu^N(f);\, f \in \mathcal{F}\}$ is usually called the $\mathcal{F}$-empirical process associated with the empirical random measures μ^N. The semi-metric commonly used in such a context is the Zolotarev semi-norm defined by

$$\forall\, \mu,\nu \in \mathbf{M}_1(E), \qquad \|\mu - \nu\|_{\mathcal{F}} \;=\; \sup\{|\mu(f) - \nu(f)|;\, f \in \mathcal{F}\}$$

(see for instance [95]). In order to control the behavior of the supremum $\|\eta_n^N - \eta_n\|_{\mathcal{F}}$ as $N \to \infty$, we will impose conditions on the class $\mathcal{F}$ that are classically used in the statistical theory of empirical processes for independent samples. To avoid technical measurability conditions, and in order not to obscure the main ideas, we will always assume the class $\mathcal{F}$ to be countable and uniformly bounded. Our conclusions also hold under appropriate separability assumptions on the empirical process (see [50]).

The Glivenko-Cantelli and Donsker Theorems are uniform versions of the law of large numbers and the central limit theorem for empirical measures. In the classical theory of independent random variables, these properties are usually shown to hold under entropy conditions on the class $\mathcal{F}$. Namely, to measure the size of a given class $\mathcal{F}$, one considers the covering numbers $N(\varepsilon, \mathcal{F}, L_p(\mu))$ defined as the minimal number of $L_p(\mu)$-balls of radius $\varepsilon > 0$ needed to cover $\mathcal{F}$. With respect to classical theory, we will need assumptions on these covering numbers uniformly over all probability measures μ. Classically also, this supremum can be taken over all discrete probability measures. Since we are dealing with interacting particle schemes, we however need to strengthen the assumption and take the corresponding supremum over all probability measures. Several examples of classes of functions satisfying the foregoing uniform entropy conditions are discussed in the book [50].

Denote thus by $\mathcal{N}(\varepsilon, \mathcal{F})$, $\varepsilon > 0$, and by $I(\mathcal{F})$ the uniform covering numbers and entropy integral given by

$$\mathcal{N}(\varepsilon, \mathcal{F}) \;=\; \sup\{N(\varepsilon, \mathcal{F}, L_2(\mu));\, \mu \in \mathbf{M}_1(E)\},$$

$$I(\mathcal{F}) \;=\; \int_0^1 \sqrt{\log \mathcal{N}(\varepsilon,\mathcal{F})}\, d\varepsilon.$$

which will be assumed to be finite.

The fluctuations and the large deviations principles for the particle density profiles $\{\eta_n^N\ ;\ n \geq 0\}$ will be simply based on $(\mathcal{G})$. In fact this assumption on the fitness functions will be used to check the asymptotic tightness property in Donsker's Theorem and an exponential tightness property in large deviation settings.

The study of fluctuations and large deviations on path space require more attention. It becomes more transparent if we introduce a more general abstract formulation. Namely, we will assume that $\{\xi_n\ ;\ n\geq 0\}$ is the IPS approximating model (13) associated with a given sequence of continuous functions $\{\Phi_n\ ;\ n\geq 1\}$. We will use the following assumption.

$(\mathcal{P})_0$ ***For any time*** $n \geq 1$ ***there exists a reference probability measure*** $\lambda_n \in \mathbf{M}_1(E)$ ***such that***

$$\forall\ \mu \in \mathbf{M}_1(E), \qquad \Phi_n(\mu) \;\sim\; \lambda_n$$

This condition might seems difficult to check in general. In fact if the functions $\{\Phi_n\ ;\ n\geq 1\}$ are given by (8) then $(\mathcal{P})_0$ holds if and only if the transition probability kernels $\{K_n\ ;\ n\geq 1\}$ satisfy the following condition.

$(\mathcal{K})_0$ ***For any time*** $n \geq 1$ ***there exists a reference probability measure*** $\lambda_n \in \mathbf{M}_1(E)$ ***so that***

$$\forall\ x \in E, \qquad K_n(x,\bullet) \;\sim\; \lambda_n$$

We shall see in section 5 that this condition covers many typical examples of nonlinear filtering problems (see also [35, 37, 36]).
The main reason for our needing to make the assumption $(\mathcal{P})_0$ for the analysis of the fluctuations on path space is that we want to use a reference product probability measure. We also notice that there is no loss of generality in choosing $\lambda_n = \eta_n$. This particular choice of reference probability measure will be used for technical reasons in the study of the fluctuations on path space. Roughly speaking this is the appropriate and natural choice for studying the weak convergence of the resulting Hamiltonian function under the underlying product measure (see for instance Lemma 2.24 and its proof given on page 52).
The main simplification due to assumption $(\mathcal{P})_0$ is the following continuity property. For any $T\in\mathbb{N}$, we denote $P_T^{(N)}$ the law of

$$\xi_{[0,T]} \stackrel{\text{def.}}{=} (\xi_0^i,\xi_1^i\ldots,\xi_T^i)_{1\leq i\leq N}$$

on the path space $(\Sigma_T^N, \mathbf{B}(\Sigma_T)^{\otimes N})$ where

$$\Sigma_T^N \stackrel{\text{def.}}{=} \underbrace{\Sigma_T\times\ldots\times\Sigma_T}_{N-\text{times}} \qquad \mathbf{B}(\Sigma_T)^{\otimes N} \stackrel{\text{def.}}{=} \underbrace{\mathbf{B}(\Sigma_T)\otimes\ldots\otimes\mathbf{B}(\Sigma_T)}_{N-\text{times}}$$

and $\Sigma_T \stackrel{\text{def.}}{=} E^{T+1}$. Then $P_T^{(N)}$ is absolutely continuous with respect to the product measure $\eta_{[0,T]}^{\otimes N}$ where

$$\eta_{[0,T]} \stackrel{\text{def.}}{=} \eta_0 \otimes \ldots \otimes \eta_T$$

Next we denote by μ_n the marginal at the time $n \in \{0, \ldots, T\}$ of a measure $\mu \in \mathbf{M}_1(\Sigma_T)$ and with some obvious abusive notations by $m(x)$ the empirical measure on path space associated with a configuration $x \in \Sigma_T^N$, that is

$$m(x) \stackrel{\text{def.}}{=} \frac{1}{N} \sum_{i=1}^{N} \delta_{(x_0^i, \ldots, x_T^i)} \in \mathbf{M}_1(\Sigma_T),$$

Under $(\mathcal{P})_0$, it is easily seen that

$$\frac{dP_T^{(N)}}{d\eta_{[0,T]}^{\otimes N}}(x) = \exp H_T^{(N)}(x) \qquad \eta_{[0,T]}^{\otimes N} - \text{a.e.}$$

where $H^{(N)} : \Sigma_T^N \to \mathbb{R}$ is the function defined by

$$H_T^{(N)}(x) = N \sum_{n=1}^{T} \int \log \frac{d\Phi_n(m_{n-1}(x))}{d\eta_n} \, dm_n(x)$$

In addition, if we consider the function $\Phi_{[0,T]} : \mathbf{M}_1(\Sigma_T) \to \mathbf{M}_1(\Sigma_T)$ so that

$$\Phi_{[0,T]}(\mu) = \eta_0 \otimes \Phi_1(\mu_0) \otimes \ldots \otimes \Phi_T(\mu_{T-1})$$

then we see that $H_T^{(N)}$ can also be rewritten in the following form

$$H_T^{(N)}(x) = N \int_{\Sigma_T} \log \frac{d\Phi_{[0,T]}(m(x))}{d\eta_{[0,T]}} \; dm(x) \tag{38}$$

Therefore, the density of $P_T^{(N)}$ only depends on the empirical measure m and we find ourselves exactly in the setting of mean field interacting particles with regular Laplace density. The study of the fluctuations for mean field interacting particle systems via precise Laplace method is now extensively developed (see for instance [3],[13], [66], [79] [109] and references therein).

Various methods are based on the fact that the law of mean field interacting processes can be viewed as a mean field Gibbs measure on path space (see (38)). In such a setting, precise Laplace's method can be developed (see for instance [3], [66], [79]). In [66], the study of the fluctuations for mean field Gibbs measures was extended to analytic potentials which probably includes our setting.
However the analysis of the fluctuations on the path space presented in section 2.2.4 is more related to Shiga/Tanaka's paper [98]. In this article, the authors restrict themselves to dynamics with independent initial data so that the partition function of the corresponding Gibbs measure is equal to one. This simplifies considerably the analysis. In fact, the proof then mainly relies on a simple formula on multiple Wiener integrals and Dynkin-Mandelbaum Theorem [54] on symmetric statistics. Also the pure jump McKean-Vlasov process studied in [98] is rather close to our model.

Using the above notations and, under $(\mathcal{P})_0$, if we denote by $Q_T^{(N)}$ the law of the empirical measures

$$\eta_{[0,T]}^N \stackrel{\text{def.}}{=} m\left(\xi_{[0,T]}\right) = \frac{1}{N}\sum_{i=1}^N \delta_{\left(\xi_0^i,\ldots,\xi_T^i\right)} \in \mathbf{M}_1(\Sigma_T),$$

then $Q_T^{(N)}$ is absolutely continuous with respect to the distribution

$$R_T^{(N)} \in \mathbf{M}_1(\mathbf{M}_1(\Sigma_T))$$

given by

$$R_T^{(N)} F = \int_{\Sigma_T^N} F(m(x))\, \lambda_0^{\otimes N}(dx_0)\ldots\lambda_T^{\otimes N}(dx_T)$$

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

$$R_T^{(N)} F = \int_{\Sigma_T^N} F(m(x))\, R_T^{\otimes N}(dx) \quad \text{with} \quad R_T = \lambda_0 \otimes \ldots \otimes \lambda_T$$

Arguing as before, it is also easily seen that

$$\frac{dQ_T^{(N)}}{dR_T^{(N)}} = \exp\left(NF_T\right) \qquad R_T^{(N)} - \text{a.s.} \tag{39}$$

where $F_T : \mathbf{M}_1(\Sigma_T) \to \mathbb{R}$ is the function defined by

$$F_T(\mu) = \sum_{n=1}^T \int_E \log \frac{d\Phi_n(\mu_{n-1})}{d\lambda_n}\, d\mu_n = \int_{\Sigma_T} \log \frac{d\Phi_{[0,T]}(\mu)}{dR_T}\, d\mu \tag{40}$$

The above formulation will be used in section 2.2.5 to obtain large deviation principles for the law of the empirical measures on path space.

It is worth observing immediately that $R_T^{(N)}$ is the law of an empirical measure associated with N independent path-valued random variables. This observation, together with the above formulation clearly shows that Varadhan's Lemma combined with Sanov's Theorem and cut-off techniques are natural tools for deriving the desired large deviation principle (see [35], [49, 51] or [106] and references therein).

We have no examples in this section 2.2. This choice is deliberate. We will give in section 5 a glossary of assumptions and detailed examples for each specific condition.

2.2.2 $\mathbb{L}^p$-mean errors

As announced in the introduction we start with a very basic but reassuring convergence result which holds without assumption $(\mathcal{G})$ on the fitness functions.

Proposition 2.9 *For any time $n \geq 0$ and $p \geq 1$, there exists some finite constant $C_n^{(p)} < \infty$ such that*

$$\forall\ f \in \mathcal{B}_{\mathrm{b}}(E), \qquad \mathbb{E}\left(|\eta_n^N f - \eta_n f|^p\right)^{\frac{1}{p}} \ \leq\ \frac{1}{\sqrt{N}}\ \|f\|\ C_n^{(p)} \tag{41}$$

In particular, for any $f \in \mathcal{B}_{\mathrm{b}}(E)$ and $n \geq 0$, $\{\eta_n^N f\ ;\ N \geq 1\}$ converges almost surely to $\eta_n f$ as N tends to ∞.

Proof: The proof of (41) is simply based on Marcinkiewicz-Zygmund's inequality (cf. [100] p. 498). More precisely, by definition of the N-particle system (13), for any $n \geq 0$ and $f \in \mathcal{B}_{\mathrm{b}}(E)$ we have that

$$\mathbb{E}\left(|\eta_n^N f - \Phi_n(\eta_{n-1}^N) f|^p\right)^{\frac{1}{p}} \leq \frac{1}{\sqrt{N}}\ \|f\|\ B_p \tag{42}$$

where B_p is a finite constant which only depends on the parameter $p \geq 1$ and where we have used the convention $\Phi_0(\eta_{-1}^N) = \eta_0$. We end up with (41) by induction on the time parameter. Since the result clearly holds for $n = 0$, we next assume that it holds at rank $(n-1)$ for some constant $C_{n-1}^{(p)}$.
Using Lemma 2.2 and the induction hypothesis at rank $(n-1)$ on can check that

$$\mathbb{E}\left(|\eta_n^N f - \eta_n f|^p\right)^{\frac{1}{p}} \leq \mathbb{E}\left(|\eta_n^N f - \Phi_n(\eta_{n-1}^N) f|^p\right)^{\frac{1}{p}} + \frac{1}{\sqrt{N}} \frac{2\|g_n\|\|f\|}{\eta_{n-1}(g_n)}\ C_{n-1}^{(p)}$$

Consequently, in view of (42) the result holds at rank n with

$$C_n^{(p)} = B_p + \frac{2\|g_n\|}{\eta_{n-1}(g_n)}\ C_{n-1}^{(p)}$$

The last assertion is a clear consequence of Borel-Cantelli Lemma. ■

One important drawback of the above inductive proof is that it does not present any information about the "exact values" of the $\mathbb{L}^p$-mean errors. For instance, $C_n^{(p)}$ is roughly of the form $B_p(n+1)\prod_{l=1}^n\left(\frac{2\|g_l\|}{\eta_{l-1}(g_l)}\right)$ and therefore tends to ∞ as $n \to \infty$. In order to get an idea of the exact values of these bounds some numerical simulations have been done in [41]. Precise asymptotic estimates will also be given in section 3.3.2.

When $(\mathcal{G})$ holds then (41) can be proved without using an induction. This approach is based on a natural decomposition of the errors. Since this decomposition will also be used to prove a uniform convergence result with respect to time, Glivenko-Cantelli and Donsker's Theorem and elsewhere we have chosen to present this alternative proof. The decomposition we will use now is the following:

$$\forall f \in \mathcal{B}_{\mathrm{b}}(E)\,,\ \forall n \geq 0\,, \qquad \eta_n^N f - \eta_n f = \sum_{p=0}^{n} \left[\Phi_{p,n}(\eta_p^N) f - \Phi_{p,n}\left(\Phi_p(\eta_{p-1}^N)\right) f\right] \tag{43}$$

with the convention $\Phi_0(\eta_{-1}^N) = \eta_0$. Using Lemma 2.2 we see that each term

$$\left|\Phi_{p,n}(\eta_p^N) f - \Phi_{p,n}\left(\Phi_p(\eta_{p-1}^N)\right) f\right|, \qquad 0 \leq p \leq n, \tag{44}$$

is bounded by

$$a_{p,n}^2 \quad \left[\, \left|\eta_p^N(f_{p,n}) - \Phi_p\left(\eta_{p-1}^N\right)(f_{p,n})\right| + \|f\| \; \left|\eta_p^N(\overline{g}_{p,n}) - \Phi_p\left(\eta_{p-1}^N\right)(\overline{g}_{p,n})\right| \, \right] \tag{45}$$

with

$$f_{p,n} = \overline{g}_{p,n} \;\; K_{p,n}(f) \qquad \text{and} \qquad \overline{g}_{p,n} = \frac{g_{p,n}}{a_{p,n}}$$

so that $\|f_{p,n}\| \leq \|f\|$ and $\|(\overline{g}_{p,n})\| \leq 1$. Using Marcinkiewicz-Zygmund's inequality it is then easy to conclude that, for any $p \geq 1$, $n \geq 0$ and $f \in \mathcal{B}_{\rm b}(E)$

$$E\left(|\eta_n^N f - \eta_n f|^p\right)^{\frac{1}{p}} \leq \frac{B_p}{\sqrt{N}} \;\; \|f\| \quad \sum_{q=0}^n a_{q,n}^2$$

where B_p is a finite universal constant. Here again, a clear deficiency in the preceding result is the degeneracy of the constants when the time parameter is growing, that is $\sum_{q=0}^n a_{q,n}^2$ tends to ∞ as $n \to \infty$. To connect this problem with the stability properties discussed in section 2.1.2 we note that each term (44) can be rewritten as follows

$$\left|\eta_p^N \left(\, \tilde{g}_{p,n}^N \;\, \tilde{K}_{p,n}^N(f) \, \right) / \eta_p^N \left(\, \tilde{g}_{p,n}^N \, \right)\right| \quad \text{with} \quad \tilde{g}_{p,n}^N = \frac{g_{p,n}}{\Phi_p(\eta_{p-1}^N)(g_{p,n})}, \qquad 0 \leq p \leq n$$

and

$$\begin{aligned} \tilde{K}_{p,n}^N(f)(x) \;\; &= \;\; K_{p,n}\left(f - \Phi_{p,n}\left(\Phi_p(\eta_{p-1}^N)\right)(f)\right)(x) \\ &= \;\; \int \left(K_{p,n}f(x) - K_{p,n}f(y)\right) \tilde{g}_{p,n}^N(y)\Phi_p(\eta_{p-1}^N)(dy) \end{aligned}$$

Since $\Phi_p(\eta_{p-1}^N)\left(\, \tilde{g}_{p,n}^N \;\, \tilde{K}_{p,n}^N(f) \, \right) = 0$ and $\|\tilde{K}_{p,n}^N(f)\| \leq \beta(K_{p,n}) \; \|f\|$ we use the same line of arguments as before to prove that

$$E\left(|\eta_n^N f - \eta_n f|^p\right)^{\frac{1}{p}} \leq \frac{B_p}{\sqrt{N}} \;\; \|f\| \quad \sum_{q=0}^n \; \sup_{x,y} \left|\frac{g_{q,n}(x)}{g_{q,n}(y)}\right|^2 \beta(K_{q,n})$$

Next we would like to extend these results in two different ways. First we would like to be able to turn this result into a statement uniform in f varying in a suitable class of functions $\mathcal{F} \subset \mathcal{B}_{\rm b}(E)$. Second we would like to obtain a uniform $\mathbb{L}^p$-error bound with respect to the time parameter without any assumptions on the mutation transition kernels K_n but only on the stability properties of the limiting system (8). We will also use the following extension of Marcinkiewicz-Zygmund's inequality to empirical processes [50].

Lemma 2.10 *Let $\left(X^1, \ldots, X^N\right)$ be independent $E-$valued random variables with common law $P^{(N)}$ and let $\mathcal{F}$ be a countable sequence of functions $f : E \to \mathbb{R}$ such that $\|f\| \leq 1$. Then, for any $p \geq 1$ there exists a universal constant B_p such that*

$$\mathbb{E}\left(\|\frac{1}{N}\sum_{i=1}^N \delta_{X^i} - P^{(N)}\|_{\mathcal{F}}^p \right)^{\frac{1}{p}} \leq \frac{B_p}{\sqrt{N}} \; I(\mathcal{F}) \tag{46}$$

Theorem 2.11 *Let $\mathcal{F}$ be a countable collection of functions f such that $\|f\| \leq 1$ and satisfying the entropy condition $I(\mathcal{F}) < \infty$. Assume moreover that the limiting dynamical system (8) is asymptotically stable in the sense that*

$$\lim_{T\to\infty} \sup_{\mu,\nu\in \mathbf{M}_1(E)} \sup_{p\geq 0} \|\Phi_{p,p+T}(\mu) - \Phi_{p,p+T}(\nu)\|_{\mathcal{F}} = 0 \tag{47}$$

When the fitness functions $\{g_n \; ; \; n \geq 1\}$ satisfy $(\mathcal{G})$ with $\sup_{n\geq 1} a_n \stackrel{\text{def.}}{=} a < \infty$ then we have the following uniform convergence result with respect to time

$$\lim_{N\to\infty} \sup_{n\geq 0} E\left(\left\|\eta_n^N - \eta_n\right\|_{\mathcal{F}}\right) = 0 \tag{48}$$

In addition, let us assume that the limiting dynamical system (8) is exponentially asymptotically stable in the sense that there exist some positive constant $\gamma > 0$ and $T_0 \geq 0$ such that,

$$\forall\ \mu,\nu \in \mathbf{M}_1(E),\ \forall\ T \geq T_0, \qquad \sup_{p\geq 0} \|\Phi_{p,p+T}(\mu) - \Phi_{p,p+T}(\nu)\|_{\mathcal{F}} \leq e^{-\gamma T} \tag{49}$$

Then we have for any $p \geq 1$, the uniform $\mathbb{L}^p$-error bound given by

$$\sup_{n\geq 0} \mathbb{E}\left(\left\|\eta_n^N - \eta_n\right\|_{\mathcal{F}}^p\right)^{\frac{1}{p}} \leq \frac{C_p\ e^{\gamma'}}{N^{\frac{\alpha}{2}}}\ \ I(\mathcal{F}) \tag{50}$$

for any $N \geq 1$ so that

$$T(N) \stackrel{\text{def}}{=\!=} \left[\frac{1}{2}\frac{\log N}{\gamma+\gamma'}\right] + 1 \geq T_0$$

where C_p is a universal constant which only depends on $p \geq 1$ and α and γ' are given by

$$\alpha = \frac{\gamma}{\gamma+\gamma'} \quad \text{and} \quad \gamma' = 1 + 2\log a.$$

Proof: We use again the decomposition (43). There is no loss of generality to assume that $1 \in \mathcal{F}$. Then, by the same line of arguments as before and using the extended version of Marcinkiewicz-Zygmund's inequality one can prove that for any $0 \leq q \leq n$ and $p \geq 1$

$$\mathbb{E}\left(\left\|\Phi_{q,n}(\eta_q^N) - \Phi_{q,n}\left(\Phi_q(\eta_{q-1}^N)\right)\right\|_{\mathcal{F}}^p \left|\eta_{q-1}^N\right.\right)^{\frac{1}{p}} \leq \frac{B_p}{\sqrt{N}}\ \ a_{q,n}^2\ \ I(\mathcal{F}_{q,n})$$

where B_p is a universal constant and

$$\mathcal{F}_{q,n} \stackrel{\text{def.}}{=} \overline{g}_{q,n}\ K_{q,n}(\mathcal{F}) \stackrel{\text{def.}}{=} \{\overline{g}_{q,n}\ K_{q,n}(f)\ ;\ f\in\mathcal{F}\} \qquad \overline{g}_{q,n} = \frac{g_{q,n}}{a_{q,n}} \tag{51}$$

Now, using the fact that $I(\mathcal{F}_{q,n}) \leq I(\mathcal{F})$ (cf. Lemma 2.3, p. 9, [42]) and $a_{q,n} \leq a_{0,n} \leq a^n$ one concludes that

$$\mathbb{E}\left(\left\|\Phi_{q,n}(\eta_q^N) - \Phi_{q,n}\left(\Phi_q(\eta_{q-1}^N)\right)\right\|_{\mathcal{F}}^p\right)^{\frac{1}{p}} \leq \frac{B_p}{\sqrt{N}}\ \ a^{2n}\ \ I(\mathcal{F})$$

and therefore for any $T \geq T_0$

$$\sup_{0\leq n\leq T} \mathbb{E}\left(\|\eta_n^N - \eta_n\|_{\mathcal{F}}^p\right)^{\frac{1}{p}} \leq \frac{B_p}{\sqrt{N}}\ (T+1)\ a^{2T}\ \ I(\mathcal{F}) \tag{52}$$

Similarly, for any $q \geq 0$ and $T \geq T_0$ we have

$$\begin{aligned} &\|\eta_{q+T}^N - \eta_{q+T}\|_{\mathcal{F}} \\ &\leq \sum_{r=q+1}^{q+T} \left\|\Phi_{r,q+T}(\eta_r^N) - \Phi_{r,q+T}\left(\Phi_r(\eta_{r-1}^N)\right)\right\|_{\mathcal{F}} + \left\|\Phi_{q,q+T}(\eta_q^N) - \Phi_{q,q+T}(\eta_q)\right\|_{\mathcal{F}} \end{aligned}$$

Under our assumptions, this implies that

$$\|\eta_{q+T}^N - \eta_{q+T}\|_{\mathcal{F}} \leq \sum_{r=q+1}^{q+T} \left\|\Phi_{r,q+T}(\eta_r^N) - \Phi_{r,q+T}\left(\Phi_r(\eta_{r-1}^N)\right)\right\|_{\mathcal{F}} + e^{-\gamma T}$$

Thus, using the same line of arguments as before, one gets for any $T \geq T_0$

$$\sup_{q\geq 0} \mathbb{E}\left(\|\eta_{q+T}^N - \eta_{q+T}\|_{\mathcal{F}}^p\right)^{\frac{1}{p}} \leq e^{-\gamma T} + \frac{B_p}{\sqrt{N}}\ (T+1)\ a^{2T}\ \ I(\mathcal{F}) \tag{53}$$

Combining (52) and (53) leads to a uniform $\mathbb{L}^p$-error bounds with respect to time in the form of the inequality

$$\forall\ T \geq T_0, \qquad \sup_{n\geq 0} \mathbb{E}\left(\|\eta_n^N - \eta_n\|_{\mathcal{F}}^p\right)^{\frac{1}{p}}\ \ \leq\ \ e^{-\gamma T} + \frac{B'_p}{\sqrt{N}}\ e^{\gamma' T}\ \ I(\mathcal{F})$$

where $\gamma' = 1 + 2\log a$ and $B'_p > 0$ is a universal constant. Obviously, if we choose $N \geq 1$ and

$$T = T(N) \stackrel{\mathrm{def}}{=} \left[\frac{1}{2}\frac{\log N}{\gamma + \gamma'}\right] + 1 \geq T_0$$

where $[r]$ denotes the integer part of $r \in \mathbb{R}$, we get that

$$\sup_{n\geq 0} \mathbb{E}\left(\left\|\eta_n^N - \eta_n\right\|_{\mathcal{F}}^p\right)^{\frac{1}{p}} \leq \frac{1}{N^{\alpha/2}}\ \left(1 + e^{\gamma'} B'_p\, I(\mathcal{F})\right)$$

where $\alpha = \gamma/(\gamma + \gamma')$. This ends the proof of the theorem. ∎

Remarks 2.12:

- The critical exponent resulting from the proof of Theorem 2.11 is sharp in the following sense: if the transition probability kernels $\{K_n\ ;\ n \geq 1\}$ are given by

$$K_n(x, dz) = \lambda_n(dz), \qquad \lambda_n \in \mathbf{M}_1(E)$$

  then we see that $\xi_n = (\xi_n^1, \ldots, \xi_n^N)$ is a sequence of independent random variables with common law λ_n. In this situation the uniform upper bound (50) holds for any choice of $\gamma \in \mathbb{R}_+$. Letting $\gamma \to \infty$ the critical exponent tends to $1/2$ which is the characteristic exponent of the weak law of large numbers.

- Several conditions for exponential asymptotic stability of the limiting measure system (8) are given in section 2.1.2. The exponential stability condition (49) can be easily related to the conditions $(\mathcal{K})_i$, $i = 1, 2, 3$ discussed in section 2.1.2.

- The proof of Theorem 2.11 can be used to treat polynomial asymptotically stable limiting dynamical systems (8). The resulting uniform $\mathbb{L}^p$-error bound has roughly the form $(\log N)^{-\beta}$ for some $\beta > 0$.

- The above result can also be used to obtain reliability intervals which are valid uniformly with respect to the time parameter (cf. [36]).

- In nonlinear filtering settings, the fitness functions and therefore the sequence $\{a_{q,n}\ 0 \leq q \leq n\}$ depend on the observation process. The above result can be used to give quenched and/or averaged uniform $\mathbb{L}^p$-mean error bounds with respect to time (cf. section 4.4 and [36]).

2.2.3 Glivenko-Cantelli Theorem

Let $\mathcal{F}$ be a countable collection of functions f such that $\|f\| \leq 1$. Upon carefully examining the proof of Theorem 2.11, we have already proved that for any $p \geq 1$ and for any time $n \geq 0$

$$\mathbb{E}\left(\|\eta_n^N - \eta_n\|_{\mathcal{F}}^p\right)^{\frac{1}{p}} \leq \frac{B_p}{\sqrt{N}}\ (n+1)\ a_{0,n}^2\ \ I(\mathcal{F})$$

Then, as an immediate consequence of Borel-Cantelli Lemma we have that

$$I(\mathcal{F}) < \infty \Longrightarrow \lim_{N\to\infty} \|\eta_n^N - \eta_n\|_{\mathcal{F}} = 0 \qquad \text{P-a.e.}$$

Our aim is now to show that this almost sure convergence remains true if we replace the entropy condition $I(\mathcal{F}) < \infty$ by a boundness condition of the covering numbers, namely $\mathcal{N}(\varepsilon, \mathcal{F}) < \infty$, for any $\varepsilon > 0$.

As usually we make use of the decomposition (43). There is no loss of generality in assuming that $1 \in \mathcal{F}$. Then, by Lemma 2.2, we get for any $0 \leq q \leq n$

$$\|\Phi_{q,n}(\mu) - \Phi_{q,n}(\nu)\|_{\mathcal{F}} \leq 2\, a_{q,n}^2\ \|\mu - \nu\|_{\mathcal{F}_{q,n}} \tag{54}$$

where the class $\mathcal{F}_{q,n}$ is the class of functions defined in (51). It easily follows that, for every $\varepsilon > 0$,

$$\begin{aligned} &P\left(\left\|\eta_n^N - \eta_n\right\|_{\mathcal{F}} > \varepsilon\right) \\ &\qquad \leq (n+1) \sup_{0\leq q\leq n} P\left(\left\|\Phi_{q,n}(\eta_q^N) - \Phi_{q,n}(\Phi_p(\eta_{q-1}^N))\right\|_{\mathcal{F}} > \frac{\varepsilon}{n+1}\right). \end{aligned}$$

Using (54), this implies that

$$P\left(\left\|\eta_n^N - \eta_n\right\|_{\mathcal{F}} > \varepsilon\right) \leq (n+1) \sup_{0\leq q\leq n} P\left(\left\|\eta_q^N - \Phi_q(\eta_{q-1}^N)\right\|_{\mathcal{F}_{q,n}} > \frac{\varepsilon}{\sigma_n}\right) \tag{55}$$

where $\sigma_n = 2(n+1)a_{0,n}^2$.

The Glivenko-Cantelli Theorem may then be stated as follows.

Theorem 2.13 *Assume that $\mathcal{F}$ is a countable collection of functions f such that $\|f\| \leq 1$ and $\mathcal{N}(\varepsilon, \mathcal{F}) < \infty$ for any $\varepsilon > 0$. Then, for any time $n \geq 0$, $\|\eta_n^N - \eta_n\|_{\mathcal{F}}$ converges almost surely to 0 as $N \to \infty$.*

Theorem 2.13 is based on the following standard lemma in the theory of empirical processes (see for instance [50] or [42]).

Lemma 2.14 *Let $\{X^i; 1 \leq i \leq N\}$ be independent random variables with common law $P^{(N)}$ and let $\mathcal{F}$ be a countable collection of functions f such that $\|f\| \leq 1$. Then, for any $\varepsilon > 0$ and $\sqrt{N} \geq 4\,\varepsilon^{-1}$ we have that*

$$P\left(\|\frac{1}{N}\sum_{i=1}^{N} \delta_{X^i} - P^{(N)})\|_{\mathcal{F}} > 8\,\varepsilon\right) \leq 8\,\mathcal{N}(\varepsilon, \mathcal{F})\, e^{-N\varepsilon^2/2}$$

Proof of Theorem 2.13: Let us fixed n throughout the argument. Using (55) and Lemma 2.14 one easily gets that, for $\sqrt{N} \geq 4\,\varepsilon_n^{-1}$ where $\varepsilon_n = \varepsilon/8\sigma_n$,

$$P\left(\left\|\eta_n^N - \eta_n\right\|_{\mathcal{F}} > \varepsilon\right) \leq 8(n+1)\, e^{-N\varepsilon_n^2/2} \sup_{0\leq q\leq n} \mathcal{N}(\varepsilon_n, \mathcal{F}_{q,n}).$$

Since $\mathcal{N}(\varepsilon, \mathcal{F}_{q,n}) \leq \mathcal{N}(\varepsilon, \mathcal{F})$ for each $\varepsilon > 0$, and $0 \leq q \leq n$ (cf. Lemma 2.3, p. 9, [42]) one concludes that

$$P\left(\left\|\eta_n^N - \eta_n\right\|_{\mathcal{F}} > \varepsilon\right) \leq 8(n+1)\mathcal{N}(\varepsilon_n, \mathcal{F})\, e^{-N\varepsilon_n^2/2}$$

as soon as $\sqrt{N} \geq 4\,\varepsilon_n^{-1}$. The end of proof of Theorem 2.13 is an immediate consequence of the Borel-Cantelli Lemma. ∎

The proof of Theorem 2.13 gives an exponential rate of convergence but this result is only valid for a number of particles larger than some value depending on the time parameter. Our aim is now to refine this exponential bound in the case of uniformly bounded classes $\mathcal{F}$ with polynomial covering numbers. More precisely we will use the following assumption

($\mathcal{P}oly.$) ***There exist some constants C and V such that***

$$\forall 0 < \varepsilon < C\,, \qquad \mathcal{N}(\varepsilon, \mathcal{F}) \leq \left(\frac{C}{\varepsilon}\right)^V$$

Several examples of classes of functions satisfying this condition are discussed in [50]. For instance Vapnik-Cervonenkis classes $\mathcal{F}$ of index $V(\mathcal{F})$ and envelope function $F = 1$ satisfy ($\mathcal{P}oly.$) with $V = 2(V(\mathcal{F})-1)$ and a constant C that only depends on V.

Theorem 2.15 *Let $\mathcal{F}$ be a countable class of measurable functions $f : E \to [0,1]$ satisfying ($\mathcal{P}oly.$) for some constants C and V. Then, for any $n \geq 0$, $\delta > 0$ and $N \geq 1$,*

$$P\left(\sqrt{N}\left\|\eta_n^N - \eta_n\right\|_{\mathcal{F}} > \delta\sigma_n\right) \leq (n+1)\left(\frac{D\,\delta}{\sqrt{V}}\right)^V e^{-2\,\delta^2}$$

where D is a constant that only depends on C and $\sigma_n = 2(n+1)\prod_{q=1}^{n} a_q^2$.

Proof: We use again the decomposition (43). Using the same notations as in the proof of Theorem 2.13 and Theorem 2.11, we have

$$P\left(\left\|\Phi_{q,n}(\eta_q^N)-\Phi_{q-1,n}(\eta_{q-1}^N)\right\|_{\mathcal{F}}>\frac{\varepsilon}{n+1}\right)\leq P\left(\left\|\eta_q^N-\Phi_q(\eta_{q-1}^N)\right\|_{\mathcal{F}_{q,n}}>\varepsilon_n\right) \quad (56)$$

where $\varepsilon_n=\varepsilon/\sigma_n$ and $\sigma_n=2(n+1)a_{0,n}^2$. It is also convenient to note that each class $\mathcal{F}_{q,n}$, $0\leq q\leq n$, satisfies $(\mathcal{P}oly.)$. Indeed, the class $\mathcal{F}_{q,n}$ is again a countable class of functions $f:E\to[0,1]$ and using Lemma 2.3 in [42] we also have, for every $\varepsilon>0$ and $0\leq q\leq n$

$$\mathcal{N}(\varepsilon,\mathcal{F}_{q,n})\leq\mathcal{N}(\varepsilon,\mathcal{F})\leq\left(\frac{C}{\varepsilon}\right)^V.$$

We are now in position to apply the exponential bounds of [107] (see also [50]). More precisely, by recalling that η_p^N is the empirical measure associated with N conditionally independent random variables with common law $\Phi_q(\eta_{q-1}^N)$, we get

$$P\left(\left\|\eta_q^N-\Phi_q(\eta_{q-1}^N)\right\|_{\mathcal{F}_{q,n}}>\varepsilon_n\mid\eta_{q-1}^N\right)\leq\left(\frac{D\sqrt{N}\varepsilon_n}{\sqrt{V}}\right)^V e^{-2\,(\sqrt{N}\varepsilon_n)^2}$$

where D is a constant that only depends on C. The remainder of the proof is exactly as in the proof of Theorem 2.13. Using (56), one gets finally

$$P\left(\left\|\eta_n^N-\eta_n\right\|_{\mathcal{F}}>\varepsilon\right)\leq(n+1)\left(\frac{D\sqrt{N}\,\varepsilon_n}{\sqrt{V}}\right)^V e^{-2\,(\sqrt{N}\varepsilon_n)^2}.$$

If we denote $\delta=\sqrt{N}\varepsilon/\sigma_n$ we obtain the desired inequality and the theorem is thus established. ■

2.2.4 Central Limit Theorems

The study of the fluctuations for the IPS scheme (13) is decomposed into three parts.

- In the first one we present central limit theorems (**CLT**) for a class of processes arising naturally in the study of the convergence of the particle scheme.
- The second part concerns a Donsker's Theorem for the particle density profiles. The identification of the covariance function is based on the convergence results presented in the first part.
- The last part of this section presents a technique for obtaining fluctuations for the empirical distributions on path space.

CLT for Processes

One of the best approaches for obtaining fluctuations of the particle density profiles is through a study of the convergence of some suitably chosen processes. To describe these processes, it is convenient to introduce some additional notations. For any $\mathbb{R}^d$-valued function $f=(f^1,\ldots,f^d)$, $f^i\in\mathcal{B}_b(E)$, $1\leq i\leq d$, and for any integral operator K on E and $\mu\in\mathbf{M}_1(E)$ we will slightly abuse notations and we write

$$\mu K(f)\stackrel{\text{def.}}{=}\left(\mu K(f^1),\ldots,\mu K(f^d)\right)$$

Let $F^N=\{F_n^N\ ;\ n\geq 0\}$ be the natural filtration associated with the N-particle system $\{\xi_n^{(N)}\ ;\ n\geq 0\}$. The first class of processes which arises naturally in our context are the $\mathbb{R}^d$-valued and F^N-martingale $\{M_n^{(N)}(f)\ ;\ n\geq 0\}$ defined by

$$\forall\ n\geq 0,\qquad M_n^{(N)}(f)\ =\ \sum_{p=0}^{n}\ \left[\eta_p^N(f_p)-\Phi_p(\eta_{p-1}^N)(f_p)\right] \tag{57}$$

with the usual convention $\Phi_0(\eta_{-1}^N)=\eta_0$ and where $f:(p,x)\in\mathbb{N}\times E\mapsto f_p(x)\in\mathbb{R}^d$ is a bounded measurable function. Using the above notations, the j-th, $1\leq j\leq d$, component of the martingale $\{M_n^{(N)}(f)\ ;\ n\geq 0\}$ is the F^N-martingale defined by

$$\forall\ n\geq 0,\qquad M_n^{(N)}(f^j)\ =\ \sum_{p=0}^{n}\ \left[\eta_p^N(f_p^j)-\Phi_p(\eta_{p-1}^N)(f_p^j)\right]$$

Most of the results presented here are based on the following CLT for the martingale (57).

Lemma 2.16
For any bounded measurable function $f:(p,x)\in\mathbb{N}\times E\mapsto f_p(x)\in\mathbb{R}^d$ and $d\geq 1$, the $\mathbb{R}^d$-valued and F^N-martingale $\{\sqrt{N}\ M_n^{(N)}(f)\ ;\ n\geq 0\}$ converges in law to an $\mathbb{R}^d$-valued and Gaussian martingale $\{M_n(f)\ ;\ n\geq 0\}$ such that for any $1\leq i,j\leq d$

$$\forall\ n\geq 0,\qquad \langle M(f^i),M(f^j)\rangle_n\ =\ \sum_{p=0}^{n}\ \eta_p\left(\left(f_p^i-\eta_p(f_p^i)\right)\left(f_p^j-\eta_p(f_p^j)\right)\right)$$

Proof: The proof is based on ideas of J. Jacod (see [40] for another presentation in more general settings). To use the CLT for triangular arrays of $\mathbb{R}^d$-valued random variables (Theorem 3.33, p. 437 in [68]) we first rewrite the martingale $\sqrt{N}\ M_n^{(N)}(f)$ in the following form

$$\sqrt{N}\ M_n^{(N)}(f)=\sum_{i=1}^{N}\sum_{p=0}^{n}\ \frac{1}{\sqrt{N}}\left(\ f_p(\xi_p^i)-\Phi_p\left(\eta_{p-1}^N\right)(f_p)\ \right)$$

If we denote by $[a]$ the integer part of $a\in\mathbb{R}$ and $\{a\}=a-[a]$ this yields

$$\sqrt{N}\ M_n^{(N)}(f)=\sum_{k=1}^{(n+1)N}\ U_k^N(f)$$

where for any $1\leq k\leq(n+1)N$,

$$U_k^N(f)=\frac{1}{\sqrt{N}}\left(\ f_p(\xi_p^i)-\Phi_p\left(\eta_{p-1}^N\right)(f_p)\ \right)\quad\text{with}\quad i=N\{\frac{k}{N}\}\quad\text{and}\quad p=\left[\frac{k}{N}\right]$$

so that $k = pN + i$. Our aim is now to describe the limiting behavior of the martingale $\sqrt{N}\ M^{(N)}(f)$ in terms of the process

$$X_t^N(f) \stackrel{\text{def.}}{=} \sum_{k=1}^{[Nt]+N} U_k^N(f)$$

To this end, we denote $\mathcal{F}_k^N$ the σ-algebra generated by the random variables ξ_p^j for any pair-index (j,p) such that $pN + j \leq k$. By definition of the IPS transitions (13) and using the fact that $\left[\frac{[Nt]}{N}\right] = [t]$ one gets that for any $i,j \leq d$

$$\sum_{k=1}^{[Nt]+N} E\left(U_k^N(f^i)U_k^N(f^j)\,\middle|\mathcal{F}_{k-1}^N\right)$$

$$= C_{[t]}^N(f^i,f^j) + \frac{[Nt]-N[t]}{N}\ \left(C_{[t]+1}^N(f^i,f^j) - C_{[t]}^N(f^i,f^j)\right)$$

where, for any $n \geq 0$ and $1 \leq i,j \leq d$

$$C_n^N\left(f^i,f^j\right) = \sum_{p=0}^{n} \Phi_p(\eta_{p-1}^N)\left(\ \left(\ f_p^i - \Phi_p(\eta_{p-1}^N)f_p^i\ \right)\left(\ f_p^j - \Phi_p(\eta_{p-1}^N)f_p^j\ \right)\ \right)$$

This implies that for any $1 \leq i,j \leq d$

$$\sum_{k=1}^{[Nt]+N} E\left(U_k^N(f^i)U_k^N(f^j)\,\middle|\mathcal{F}_{k-1}^N\right) \xrightarrow[N\to\infty]{P} C_t(f^i,f^j)$$

with

$$\forall\ n \geq 0, \qquad C_n\left(f^i,f^j\right) \;=\; \sum_{p=0}^{n} \eta_p\left(\ \left(\ f_p^i - \eta_p f_p^i\ \right)\left(\ f_p^j - \eta_p f_p^j\ \right)\ \right)$$

and

$$\forall\ t \in \mathbb{R}_+, \qquad C_t(f^i,f^j) \;=\; C_{[t]}\left(f^i,f^j\right) + \{t\}\ \left(C_{[t]+1}\left(f^i,f^j\right) - C_{[t]}\left(f^i,f^j\right)\right)$$

Since $\left\|U_k^N(f)\right\| \leq \frac{2}{\sqrt{N}}\|f_{\left[\frac{k}{N}\right]}\|$ for any $1 \leq k \leq [Nt]+N$, the conditional Linderberg condition is clearly satisfied and therefore one concludes that the $\mathbb{R}^d$-valued martingale $\{X_t^N(f)\ ;\ t \in \mathbb{R}_+\}$ converges in law to a continuous Gaussian martingale

$$\{X_t(f)\ ;\ t \in \mathbb{R}_+\}$$

such that, for any $1 \leq i,j \leq d$

$$\forall\ t \in \mathbb{R}_+, \qquad \langle X(f^i), X(f^j)\rangle_t \;=\; C_t(f^i,f^j)$$

Recalling that $X_{[t]}^N(f) = \sqrt{N}\ M_{[t]}^{(N)}(f)$ the proof of the lemma is completed. ∎

A first consequence of Lemma 2.16 is another CLT for a martingale process related to the "un-normalized" approximating measures $\{\gamma_n^N\ ;\ n \geq 0\}$ defined in (17).

Proposition 2.17 *For any $T \geq 0$ and $f = (f^1, \ldots, f^d) \in \mathcal{B}_{\rm b}(E)^d$, the $\mathbb{R}^d$-valued process*

$$\Gamma_n^N(f) = \gamma_n^N(Q_{n,T}f) - \gamma_n(Q_{n,T}f) \qquad 0 \leq n \leq T \tag{58}$$

is an F^N-martingale such that, for any $1 \leq i,j \leq d$ and $0 \leq n \leq T$

$$\begin{aligned}&\langle \Gamma^N(f^i), \Gamma^N(f^j)\rangle_n \\ &= \frac{1}{N}\sum_{p=0}^{n} \left(\gamma_p^N(1)\right)^2 \ \Phi_p(\eta_{p-1}^N)\left(\left[Q_{p,T}f^i - \Phi_p(\eta_{p-1}^N)Q_{p,T}f^i\right]\right. \\ &\qquad\qquad\qquad\qquad \left.\times \left[Q_{p,T}f^j - \Phi_p(\eta_{p-1}^N)Q_{p,T}f^j\right]\ \right)\end{aligned} \tag{59}$$

Moreover, the F^N-martingale $\{\sqrt{N}\Gamma_n^N(f)\ ;\ 0 \leq n \leq T\}$ converges in law to an $\mathbb{R}^d$-valued and Gaussian martingale $\{\Gamma_n(f)\ ;\ 0 \leq n \leq T\}$ such that for any $1 \leq i,j \leq d$ and $0 \leq n \leq T$

$$\langle \Gamma(f^i), \Gamma(f^j)\rangle_n = \sum_{p=0}^{n} \left(\gamma_p(1)\right)^2\ \eta_p\left(\ \left[Q_{p,T}f^i - \eta_p Q_{p,T}f^i\right]\left[Q_{p,T}f^j - \eta_p Q_{p,T}f^j\right]\ \right)$$

Proof: For any $\varphi = (\varphi^1, \ldots, \varphi^d) \in \mathcal{B}_{\rm b}(E)^d$ we have the decomposition

$$\gamma_n^N(\varphi) - \gamma_n(\varphi) = \sum_{p=0}^{n}\ \left[\gamma_p^N(Q_{p,n}\varphi) - \gamma_{p-1}^N(Q_pQ_{p,n}\varphi)\right]$$

with the usual convention $\gamma_{-1}^N(Q_0) = \gamma_0(= \eta_0)$. By definition of $\{\gamma_n^N\ ;\ n \geq 0\}$ this can also be written in the following form

$$\gamma_n^N(\varphi) - \gamma_n(\varphi) = \sum_{p=0}^{n}\ \gamma_p^N(1)\left[\eta_p^N(Q_{p,n}\varphi) - \Phi_p(\eta_{p-1}^N)(Q_{p,n}\varphi)\right]$$

Therefore one gets (58) by choosing $\varphi = Q_{n,T}f$ and (59) is a clear consequence of the above decomposition. We turn now to the proof of the convergence of the F^N-martingale $\{\sqrt{N}\Gamma_n^N(f)\ ;\ 0 \leq n \leq T\}$. If we put for any $0 \leq n \leq T$

$$\overline{\Gamma}_n^N(f) \stackrel{\text{def.}}{=} \sum_{p=0}^{n}\ \gamma_p(1)\left[\eta_p^N(Q_{p,T}(f)) - \Phi_p(\eta_{p-1}^N)(Q_{p,T}(f))\right]$$

and if we denote $\|a\| = \sum_{i=1}^d |a^i|$, for any $a \in \mathbb{R}^d$, then, under our assumptions,

$$\mathbb{E}\left(\sup_{0\leq n\leq T}\left\|\overline{\Gamma}_n^N(f) - \Gamma_n^N(f)\right\|\right) \leq \frac{C_T(f)}{N} \tag{60}$$

for some finite constant $C_T(f) < \infty$ which does not depend on the parameter N. Lemma 2.16 clearly implies that the F^N-martingale $\{\sqrt{N}\,\overline{\Gamma}_n^N(f)\ ;\ 0 \leq n \leq T\}$ converges in law to the desired Gaussian martingale $\{\Gamma_n(f)\ ;\ 0 \leq n \leq T\}$ and (60) clearly ends the proof of the proposition. ∎

Corollary 2.18
For any time $T \geq 0$, the sequence of random fields $\{W_T^{\gamma,N}(f)\ ;\ f \in \mathcal{B}_{\rm b}(E)\}$ where

$$W_T^{\gamma,N}(f) \stackrel{def.}{=} \sqrt{N}\left(\gamma_T^N(f) - \gamma_T(f)\right)$$

converges in law as $N \to \infty$, in the sense of convergence of finite dimensional distributions, to a centered Gaussian field $\{W_T^\gamma(f)\ ;\ f \in \mathcal{B}_{\rm b}(E)\}$ satisfying for any $f, h \in \mathcal{B}_{\rm b}(E)$

$$\mathbb{E}(W_T^\gamma(f) W_T^\gamma(h)) = \sum_{p=0}^{T} (\gamma_p(1))^2\ \eta_p(\ [Q_{p,T}f - \eta_p Q_{p,T}f]\,[Q_{p,T}h - \eta_p Q_{p,T}h]\ )$$

The analysis of the fluctuations for the particle density profiles $\{\eta_n^N\ ;\ n \geq 0\}$ is more delicate mainly because the limiting measure valued process (8) is not linear. In fact many ideas work equally well when we replace the semigroup $\{Q_{p,n}\ ;\ 0 \leq p \leq n\}$ by the "normalized" one $\{\overline{Q}_{p,n}\ ;\ 0 \leq p \leq n\}$ given by

$$\forall\ 0 \leq p \leq n,\ \forall\ f \in \mathcal{B}_{\rm b}(E), \qquad \overline{Q}_{p,n}f \stackrel{\rm def.}{=} \frac{Q_{p,n}f}{\eta_p(Q_{p,n}1)}$$

To see that $\{\overline{Q}_{p,n}\ ;\ 0 \leq p \leq n\}$ is indeed a semigroup we first note that for any $0 \leq p \leq m \leq n$

$$\overline{Q}_{p,n}f = \frac{Q_{p,m}Q_{m,n}f}{\eta_p(Q_{p,n}1)} = \frac{\eta_m(Q_{m,n}1)}{\eta_p(Q_{p,n}1)}\ Q_{p,m}\overline{Q}_{m,n}f$$

Since

$$\eta_m(Q_{m,n}1) = \frac{\eta_p(Q_{p,m}Q_{m,n}1)}{\eta_p(Q_{p,m}1)} = \frac{\eta_p(Q_{p,n}1)}{\eta_p(Q_{p,m}1)}$$

one concludes that for any $0 \leq p \leq m \leq n$

$$\overline{Q}_{p,n}f = \frac{1}{\eta_p(Q_{p,m}1)} Q_{p,m}\overline{Q}_{m,n}f = \overline{Q}_{p,m}\overline{Q}_{m,n}f$$

For any $0 \leq n \leq T$ and $f = (f^1, \ldots, f^d) \in \mathcal{B}_{\rm b}(E)^d$ we write

$$f_{n,T} = \overline{Q}_{n,T}\,(f - \eta_T f) \tag{61}$$

Using this notation, the analog of Proposition 2.17 for the particle density profiles $\{\eta_n^N\ ;\ n \geq 0\}$ is the following result.

Proposition 2.19 *For any $T \geq 0$ and $f = (f^1, \ldots, f^d) \in \mathcal{B}_{\rm b}(E)^d$ the $\mathbb{R}^d$-valued process $\{W_n^N(f_{n,T})\ ;\ 0 \leq n \leq T\}$ given by*

$$W_n^N(f_{n,T}) \stackrel{\rm def.}{=} \sqrt{N}\ \eta_n^N(f_{n,T}) \tag{62}$$

converges in law to an $\mathbb{R}^d$-valued Gaussian martingale $\{W_n(f_{n,T})\ ;\ 0 \leq n \leq T\}$ such that for any $1 \leq i, j \leq d$ and $0 \leq n \leq T$

$$\langle\ W(f^i_{\bullet,T})\,,\ W(f^j_{\bullet,T})\ \rangle_n = \sum_{p=0}^{n}\ \eta_p\left(\ f^i_{p,T}\ f^j_{p,T}\ \right)$$

Proof: For any $\varphi = (\varphi^1, \ldots, \varphi^d) \in \mathcal{B}_{\rm b}(E)^d$ we have the decomposition

$$\eta_n^N(\overline{Q}_{n,T}\varphi) = \eta_0^N(\overline{Q}_{0,T}\varphi) + \sum_{p=1}^{n} \left[\eta_p^N(\overline{Q}_{p,T}\varphi) - \eta_{p-1}^N(\overline{Q}_{p-1,T}\varphi)\right]$$

If we choose $\varphi = (f - \eta_T f)$ with $f = (f^1, \ldots, f^d) \in \mathcal{B}_{\rm b}(E)^d$ this yields

$$\eta_n^N(f_{n,T}) = B_n^{(N)}(f_{\bullet,T}) + M_n^{(N)}(f_{\bullet,T})$$

where

$$B_n^{(N)}(f_{\bullet,T}) = \sum_{p=1}^{n} \left[1 - \eta_{p-1}^N(\overline{Q}_{p-1,p}1)\right] \;\; \Phi_p(\eta_{p-1}^N)(f_{p,T}) \tag{63}$$

$$M_n^{(N)}(f_{\bullet,T}) = \sum_{p=0}^{n} \left[\eta_p^N(f_{p,T}) - \Phi_p(\eta_{p-1}^N) f_{p,T}\right] \tag{64}$$

with the usual convention $\Phi_0(\eta_{-1}^N) = \eta_0$. Since for any $0 \leq p \leq T$

$$\eta_p\left(f_{p,T}\right) = \eta_T(f - \eta_T f) = 0 \quad \text{and} \quad \eta_{p-1}\left(\overline{Q}_{p-1,p}1\right) = \eta_p(1) = 1$$

then (63) can also be written in the following form

$$\begin{aligned} & B_n^{(N)}(f_{\bullet,T}) \\ &= \sum_{p=1}^{n} \left[\eta_{p-1}(\overline{Q}_{p-1,p}1) - \eta_{p-1}^N(\overline{Q}_{p-1,p}1)\right] \;\; \left[\Phi_p(\eta_{p-1}^N) f_{p,T} - \Phi_p(\eta_{p-1}) f_{p,T}\right] \end{aligned}$$

Using Proposition 2.9 and Lemma 2.2 one gets after some tedious but easy calculations

$$\mathbb{E}\left(\sup_{0\leq n\leq T} \|B_n^{(N)}(f_{\bullet,T})\|\right) \leq \frac{C_T}{N}\, \|f\| \tag{65}$$

for some finite constant $C_T < \infty$ which only depends on the parameter T (we recall that $\|f\| = \sum_{i=1}^d \|f^i\|$, for any $f = (f^1, \ldots, f^d)$). Using the same arguments as in the proof of Proposition 2.17 we ends the proof of Proposition 2.19. More precisely, Lemma 2.16 implies that the F^N-martingale

$$\{\sqrt{N}\, M_n^{(N)}(f_{\bullet,T}) \; ; \; 0 \leq n \leq T\}$$

converges in law to the desired Gaussian martingale $\{W_n(f_{\bullet,T}) \; ; \; 0 \leq n \leq T\}$ and (65) completes the proof of the proposition. ■

Corollary 2.20
For any time $T \geq 0$, the sequence of random fields $\{W_T^N(f) \; ; \; f \in \mathcal{B}_{\rm b}(E)\}$ where

$$W_T^N(f) \stackrel{\text{def.}}{=} \sqrt{N}\left(\eta_T^N(f) - \eta_T(f)\right)$$

converges in law as $N \to \infty$, in the sense of convergence of finite dimensional distributions, to a centered Gaussian field $\{W_T(f) \; ; \; f \in \mathcal{B}_{\rm b}(E)\}$ satisfying for any $f, h \in \mathcal{B}_{\rm b}(E)$

$$\mathbb{E}(W_T(f) W_T(h)) = \sum_{p=0}^{T} \eta_p\left[\overline{Q}_{p,T}\,(f - \eta_T f)\, \overline{Q}_{p,T}\,(h - \eta_T h)\right]$$

Donsker Theorem

Before getting into the details it is useful to make a couple of remarks. In the first place we note that the covariance functions can be formulated using the fitness functions $\{g_{p,T}\ ;\ 0 \leq p \leq T\}$ and the transitions $\{K_{p,T}\ ;\ 0 \leq p \leq T\}$ defined in Lemma 2.2. More precisely, since

$$\overline{Q}_{p,T}\,(f - \eta_T f) = \overline{Q}_{p,T}(1)\left(\frac{\overline{Q}_{p,T}(f)}{\overline{Q}_{p,T}(1)} - \eta_T f\right) = \frac{g_{p,T}}{\eta_p(g_{p,T})}\ (K_{p,T} f - \eta_T f)$$

for any $f \in \mathcal{B}_{\rm b}(E)$, we have that

$$\mathbb{E}(W_T(f)W_T(h)) = \sum_{p=0}^{T} \int \left(\frac{g_{p,T}}{\eta_p(g_{p,T})}\right)^2 (K_{p,T} f - \eta_T f)\,(K_{p,T} h - \eta_T h)\ d\eta_p$$

If the transition probability kernels $\{K_n;\ n \geq 1\}$ are trivial, in the sense that,

$$\forall\ 0 \leq p \leq T, \qquad K_p(x, dz) \ = \ \mu_p(dz)\ \in\ \mathbf{M}_1(E)$$

then one can readily check that $\eta_p = \mu_p$ for any $0 \leq p \leq T$ and

$$\forall 0 \leq p < T\,, \qquad K_{p,T}(x, dz) = \mu_T(dz)$$

In this particular situation $\{W_T(f);\ f \in \mathcal{B}_{\rm b}(E)\}$ is the classical μ_T-Brownian bridge. Namely, W_T is the centered Gaussian process with covariance

$$\mathbb{E}(W_T(f)W_T(h)) = \mu_T\left((f - \mu_T f)(h - \mu_T h)\right).$$

The second remark is that the random fields $\{W_T^N(f);\ f \in \mathcal{B}_{\rm b}(E)\}$ can also be regarded as an empirical process indexed by the collection of bounded measurable functions. In this interpretation, Corollary 2.20 simply says that the marginals of the $\mathcal{B}_{\rm b}(E)$-indexed empirical process weakly converge to the marginals of a centered Gaussian process $\{W_T(f);\ f \in \mathcal{B}_{\rm b}(E)\}$. One natural question we may ask is whether there exists a functional convergence result for an $\mathcal{F}$-indexed empirical process $\{W_T^N(f);\ f \in \mathcal{F}\}$ where $\mathcal{F} \subset \mathcal{B}_{\rm b}(E)$.

We recall that weak convergence in $l^\infty(\mathcal{F})$ can be characterized as the convergence of the marginals together with the asymptotic tightness of the process

$$\{W_T^N(f)\ ;\ f \in \mathcal{F}\}$$

As announced in section 2.2.1 the asymptotic tightness is related to the entropy condition $I(\mathcal{F}) < \infty$. The following technical lemma is proved in [42].

Lemma 2.21 *If $\mathcal{F}$ is a countable collection of functions f such that $\|f\| \leq 1$ and $I(\mathcal{F}) < \infty$ then, for any $T \geq 0$, the $\mathcal{F}$-indexed process $\{W_T^N(f);\ f \in \mathcal{F}\}$ is asymptotically tight.*

Theorem 2.22 (Donsker Theorem) *Assume that $\mathcal{F}$ is a countable class of functions such that $\|f\| \leq 1$ for any $f \in \mathcal{F}$ and $I(\mathcal{F}) < \infty$. Then, for any $T \geq 0$, $\{W_T^N(f);\ f \in \mathcal{F}\}$ converges weakly in $l^\infty(\mathcal{F})$ as $N \to \infty$ to a centered Gaussian process $\{W_T(f);\ f \in \mathcal{F}\}$ with covariance function*

$$\mathbb{E}(W_T(f)W_T(h)) = \sum_{p=0}^{T} \int \left(\frac{g_{p,T}}{\eta_p(g_{p,T})}\right)^2 (K_{p,T}(f) - \eta_T(f))(K_{p,T}(h) - \eta_T(h))d\eta_p.$$

Fluctuations on Path Space

In this section we will use notations of section 2.2.1, p. 33 and the following strengthening of $(\mathcal{K})_0$

$(\mathcal{TCL})$ ***For any time $n \geq 1$ there exist a reference probability measure $\lambda_n \in \mathbf{M}_1(E)$ and a $\mathbf{B}(E)$-measurable function φ_n so that $K_n(x, \bullet) \sim \lambda_n$ and***

$$\forall\ p \geq 1, \qquad \left|\log \frac{dK_n(x, \bullet)}{d\lambda_n}(z)\right| \leq \varphi_n(z) \qquad \textit{and} \qquad \int \exp\left(p\ \varphi_n\right) d\lambda_n < \infty$$

As we already noticed the distribution $P_T^{(N)}$ induced by $\xi_{[0,T]}$ on path space

$$(\Sigma_T^N, \mathbf{B}(\Sigma_T)^N)$$

is absolutely continuous with respect to the product measure $\eta_{[0,T]}^{\otimes N}$ and

$$\frac{dP_T^{(N)}}{d\eta_{[0,T]}^{\otimes N}}(x) = \exp H_T^{(N)}(x), \qquad \eta_{[0,T]}^{\otimes N} - \text{a.e.},$$

where $H^{(N)} : \Sigma_T^N \to \mathbb{R}$ is the symmetric function given by

$$H_T^{(N)}(x) = N\ \sum_{n=1}^{T}\ \int\ \log \frac{d\Phi_n(m_{n-1}(x))}{d\eta_n}\ dm_n(x)$$

To clarify the presentation, we simplify the notations suppressing the time parameter T in our notations so that we write η, $P^{(N)}$, W^N, $H^{(N)}$, Σ and Σ^N instead of $\eta_{[0,T]}$, $P_{[0,T]}^{(N)}$, $W_{[0,T]}^N$, $H_T^{(N)}$, Σ_T and Σ_T^N.

In what follows we use $\mathbb{E}_{\eta^{\otimes N}}(\,.\,)$ (resp. $\mathbb{E}_{P^{(N)}}(\,.\,)$) to denote expectations with respect to the measure $\eta^{\otimes N}$ (resp. $P^{(N)}$) on Σ^N and, unless otherwise stated, the sequence $\{x^i\ ;\ i \geq 1\}$ is regarded as a sequence of Σ-valued and independent random variables with common law η.

To get the fluctuations of the empirical measures on path space it is enough to study the limit of

$$\{\mathbb{E}_{P^{(N)}}(\exp{(iW^N(\varphi))})\ ;\ N \geq 1\} \qquad \text{where} \quad W^N = \sqrt{N}\ \left(\eta^N - \eta\right)$$

for functions $\varphi \in L^2(\eta)$. Writing

$$\mathbb{E}_{P^{(N)}}(\exp(iW^N(\varphi))) = \mathbb{E}_{\eta^{\otimes N}}(\exp(iW^N(\varphi) + H^{(N)}(x))),$$

one finds that the convergence of

$$\{\mathbb{E}_{P^{(N)}}(\exp(iW^N(\varphi)))\ ;\ N \geq 1\}$$

follows from the convergence in law and the uniform integrability of

$$\exp(iW^N(\varphi) + H^{(N)}(x))$$

under the product law $\mathbb{E}_{\eta^{\otimes N}}$. The last point is clearly equivalent to the uniform integrability of $\exp H^{(N)}(x)$ under $\mathbb{E}_{\eta^{\otimes N}}$. The proof of the uniform integrability of $\exp H^{(N)}(x)$ then relies on a classical result (see for instance Theorem 5 p. 189 in [100] or Scheffé's Lemma 5.10 p.55 in [112]) which says that, if a sequence of non-negative random variables $\{X_N\ ;\ N \geq 1\}$ converges almost surely towards some random variable X as $N \to \infty$ then we have

$$\lim_{N\to\infty} \mathbb{E}(X_N) = \mathbb{E}(X) < \infty \Longleftrightarrow \{X_N\ ;\ N \geq 1\} \text{ is uniformly integrable}$$

The equivalence still holds if X_N only converges in distribution by Skorohod's Theorem (see for instance Theorem 1 p. 355 in [100]). Since $\mathbb{E}_{\eta^{\otimes N}}(\exp H^{(N)}(x)) = 1$ it is clear that the uniform integrability of

$$\{\exp H^{(N)}(x)\ ;\ N \geq 1\}$$

follows from the convergence in distribution of $H^{(N)}(x)$ towards a random variable H such that $\mathbb{E}(\exp H) = 1$. Thus, it suffices to study the convergence in distribution of

$$\{iW^N(\varphi) + H^{(N)}(x)\ ;\ N \geq 1\}$$

for $L^2(\eta)$ functions φ to conclude.

To state such a result we first need to introduce some notations. Under the assumption $(\mathcal{K})_0$, for any $n \geq 1$ there exists a reference probability measure $\lambda_n \in \mathbf{M}_1(E)$ such that $K_n(x,.) \sim \lambda_n$. In this case we shall use the notation

$$\forall\ (x,z) \in E^2, \qquad k_n(x,z) \ \stackrel{\text{def.}}{=}\ \frac{dK_n(x,.)}{d\lambda_n}(z)$$

For any $x = (x_0, \ldots, x_T)$ and $z = (z_0, \ldots, z_T) \in \Sigma$ set

$$\begin{aligned} q(x,z) &= \sum_{n=1}^{T} q_n(x,z) \qquad \text{with} \quad q_n(x,z) = \frac{g_n(z_{n-1})\ k_n(z_{n-1}, x_n)}{\displaystyle\int_E g_n(u)\ k_n(u, x_n)\eta_{n-1}(du)} \\ a(x,z) &= q(x,z) - \int_\Sigma q(x',z)\ \eta(dx') \end{aligned}$$

One consequence of $(\mathcal{TCL})$ is that the integral operator A given by

$$\forall\ \varphi \in L^2(\Sigma,\eta), \qquad A_T\varphi(x) \ = \ \int a(z,x)\ \varphi(z)\ \eta(dz)$$

is an Hilbert-Schmidt operator on $L^2(\Sigma,\eta)$.

Theorem 2.23 *Assume that condition $(\mathcal{TCL})$ is satisfied. For any $T \geq 0$ the integral operator $I - A_T$ is invertible and the random field*

$$\{W^N_{[0,T]}(\varphi)\ ;\ \varphi \in L^2(\eta_{[0,T]})\}$$

converges as $N \to \infty$ to a centered Gaussian field

$$\{W_{[0,T]}(\varphi)\ ;\ \varphi \in L^2(\eta_{[0,T]})\}$$

satisfying

$$\begin{aligned}
&\mathbb{E}\left(W_{[0,T]}(\varphi_1)W_{[0,T]}(\varphi_2)\right)\\
&= \left((I - A_T)^{-1}(\varphi_1 - \eta(\varphi_1)), (I - A_T)^{-1}(\varphi_2 - \eta(\varphi_2))\right)_{L^2(\eta_{[0,T]})}
\end{aligned}$$

for any $\varphi_1, \varphi_2 \in L^2(\eta_{[0,T]})$, in the sense of convergence of finite dimensional distributions.

The basic tools for studying the convergence in law of $\{H^{(N)}(x)\ ;\ N \geq 1\}$ are the Dynkin-Mandelbaum Theorem on symmetric statistics and Shiga-Tanaka's formula of Lemma 1.3 in [98]. The detailed proof of Theorem 2.23 is given in [37]. Here we merely content ourselves in describing the main line of this approach. Here again we simplify notations suppressing the time parameter T and we write A instead of A_T. Let us first recall how one can see that $I - A$ is invertible. This is in fact classical now (see [3] and [98] for instance). First one notices that, under our assumptions, A^n, $n \geq 2$ and $A\ A^*$ are trace class operators with

$$\begin{aligned}
\mathrm{Trace} A^n &= \int \dots \int a(x^1, x^2) \dots a(x^n, x^1)\ \eta(dx^1) \dots \eta(dx^n)\\
\mathrm{Trace} A A^* &= \int_{\Sigma_T^2} a(x, z)^2\ \eta(dx)\ \eta(dz) = \|a\|^2_{L^2(\eta \otimes \eta)}
\end{aligned}$$

Furthermore by definition of a and the fact that η is a product measure it is easily checked that

$$\forall\ n \geq 2, \qquad \mathrm{Trace} A^n = 0$$

Standard spectral theory (see [101] for instance) then shows that $\det_2(I - A)$ is equal to one and therefore that $I - A$ is invertible.

The identification of the weak limit of $\{H^{(N)}(x)\ ;\ N \geq 1\}$ relies on L^2-techniques and more precisely Dynkin-Mandelbaum construction of multiple Wiener integrals as a limit of symmetric statistics. To state such a result, we first introduce Wiener integrals.

Let $\{I_1(\varphi)\ ;\ \varphi \in L^2(\eta)\}$ be a centered Gaussian field satisfying

$$\mathbb{E}(I_1(\varphi_1)I_1(\varphi_2)) = (\varphi_1, \varphi_2)_{L^2(\eta)}$$

If we set, for each $\varphi \in L^2(\eta)$ and $m \geq 1$

$$h_0^\varphi = 1 \qquad h_m^\varphi(z_1, \dots, z_m) = \varphi(z_1) \dots \varphi(z_m),$$

the multiple Wiener integrals $\{I_m(h_m^\varphi)\ ;\ \varphi \in L^2(\eta)\}$ with $m \geq 1$, are defined by the relation

$$\sum_{m\geq 0} \frac{t^m}{m!}\, I_m(h_m^\varphi) = \exp\left(tI_1(\varphi) - \frac{t^2}{2}\|\varphi\|^2_{L^2(\eta)}\right).$$

The multiple Wiener integral $I_m(\phi)$ for $\phi \in L^2_{\mathrm{sym}}(\eta^{\otimes m})$ is then defined by a completion argument. Theorem 2.23 is therefore a consequence of the following lemma.

Lemma 2.24 ([37])

$$\lim_{N\to\infty} H^{(N)}(x) \overset{\mathrm{law}}{=} \frac{1}{2}I_2(f) - \frac{1}{2}\mathrm{Trace} AA^* \tag{66}$$

where f is given by

$$f(y,z) = a(y,z) + a(z,y) - \int_{\Sigma_T} a(u,y)\, a(u,z)\, \eta(du). \tag{67}$$

In addition, for any $\varphi \in L^2(\eta)$,

$$\lim_{N\to\infty} (H^{(N)}(x) + iW^N(\varphi)) \overset{\mathrm{law}}{=} \frac{1}{2}I_2(f) + iI_1(\varphi) - \frac{1}{2}\mathrm{Trace} AA^*$$

Following the above observations, we get for any $\varphi \in L^2(\eta)$,

$$\begin{aligned} \lim_{N\to\infty} \mathbb{E}_{P^{(N)}}\left(\exp iW^N(\varphi)\right) &= \lim_{N\to\infty} \mathbb{E}_{\eta^{\otimes N}}\left(\exp\left(iW^N(\varphi) + H^{(N)}(x)\right)\right) \\ &= \mathbb{E}\left(\exp\left(iI_1(\varphi) + \frac{1}{2}I_2(f) - \frac{1}{2}\mathrm{Trace} AA^*\right)\right) \end{aligned}$$

Moreover, Shiga-Tanaka's formula of Lemma 1.3 in [98] shows that for any $\varphi \in L^2_{\mathrm{sym}}(\eta)$,

$$\mathbb{E}\left(\exp\left(iI_1(\varphi) + \frac{1}{2}I_2(f) - \frac{1}{2}\mathrm{Trace} AA^*\right)\right) = \exp\left(-\frac{1}{2}\|(I-A)^{-1}\varphi\|^2_{L^2(\eta)}\right) \tag{68}$$

The proof of Theorem 2.23 is thus complete. The proof of Lemma 2.24 entirely relies on a construction of multiple Wiener integrals as a limit of symmetric statistics. For completeness and to guide the reader we present this result.

Let $\{\zeta^i\ ;\ i \geq 1\}$ be a sequence of independent and identically distributed random variables with values in an arbitrary measurable space $(\mathcal{X}, \mathcal{B})$. To every symmetric function $h(z_1, \ldots, z_m)$ there corresponds a statistic

$$\sigma_m^N(h) = \sum_{1\leq i_1<\ldots<i_m\leq N} h(\zeta^{i_1}, \ldots, \zeta^{i_m})$$

with the convention $\sigma_m^N = 0$ for $m > N$. Every integrable symmetric statistic $S(\zeta^1, \ldots, \zeta^N)$ has a unique representation of the form

$$S(\zeta^1, \ldots, \zeta^N) = \sum_{m\geq 0} \sigma_m^N(h_m) \tag{69}$$

where $h_m(z_1,\ldots,z_m)$ are symmetric functions subject to the condition

$$\int h_m(z_1,\ldots,z_{m-1},u)\,\mu(du)=0 \tag{70}$$

where μ is the probability distribution of ζ^1.
We call such functions $\{h_m\;;\;m\geq 0\}$ canonical. Finally we denote by $\mathcal{H}$ the set of all sequences $h=(h_0,h_1(z_1),\ldots,h_m(z_1,\ldots,z_m),\ldots)$ where h_m are canonical and

$$\sum_{m\geq 0}\frac{1}{m!}\mathbb{E}(h_m^2(\zeta^1,\ldots,\zeta^m))<\infty$$

As in [98] we will use repeatedly the following

Theorem 2.25 (Dynkin-Mandelbaum [54])
For $h\in\mathcal{H}$ the sequence of random variables $Z_N(h)=\sum_{m\geq 0}\frac{1}{N^{m/2}}\sigma_m^N(h_m)$ converges in law, as $N\to\infty$, to

$$W(h)=\sum_{m\geq 0}\frac{I_m(h_m)}{m!}$$

Proof of Lemma 2.24: (*Sketched*)
It is first useful to observe that for any $\mu\in\mathbf{M}_1(E)$ and $n\geq 1$ we have that

$$\frac{d\Phi_n(\mu)}{d\eta_n}(x)=\frac{d\Phi_n(\mu)}{d\Phi_n(\eta_{n-1})}(x)=\frac{\mu\left(g_n(\,.\,)k_n(\,.\,,x)\right)}{\eta_{n-1}\left(g_n(\,.\,)k_n(\,.\,,x)\right)}\Bigg/\frac{\mu\left(g_n\right)}{\eta_{n-1}\left(g_n\right)}$$

Therefore the symmetric statistics $H^{(N)}(x)$ can be written in the form

$$H^{(N)}(x)=\sum_{n=1}^{T}\sum_{i=1}^{N}\left[\log\left(\frac{1}{N}\sum_{j=1}^{N}q_n(x^i,x^j)\right)-\log\left(\frac{1}{N}\sum_{j=1}^{N}\overline{q}_n(x^j)\right)\right]$$

where

$$\overline{q}_n(x^j)=g_n(x_{n-1}^j)/\eta_{n-1}(g_n)=\int_\Sigma q_n(y,x^j)\,\eta(dy)$$

By the representation

$$\log z=(z-1)-\frac{(z-1)^2}{2}+\frac{(z-1)^3}{3(\varepsilon z+(1-\varepsilon))^3}$$

which is valid for all $z>0$ with $\varepsilon=\varepsilon(z)$ such that $\varepsilon(z)\in[0,1]$ we obtain the decomposition

$$\begin{aligned}H^{(N)}(x)\;&=\;\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}a(x^i,x^j)-\frac{1}{2}\sum_{n=1}^{T}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{j=1}^{N}q_n(x^i,x^j)-1\right)^2\\&\quad+\frac{N}{2}\sum_{n=1}^{T}\left(\frac{1}{N}\sum_{j=1}^{N}\overline{q}_n(x^j)-1\right)^2+R^{(N)}\end{aligned}\tag{71}$$

where the remainder term $R^{(N)}$ cancels as N tends to ∞. The technical trick is to decompose each term as in (69) in order to identify the limit by applying Theorem 2.25. For instance, the first term can readily be written as follows

$$\frac{1}{N}\sum_{i=1}^{N} a(x^i,x^i) + \frac{1}{N}\sum_{i<j}(a+a^*)(x^i,x^j)$$

with $\int_\Sigma \; a(z,z)\;\eta(dz) = \int_\Sigma \; a(x,z)\;\eta(dz) = \int_\Sigma \; a(z,x)\;\eta(dz) = 0$ for any $x\in\Sigma$ and therefore a clear application of Theorem 2.25 yields that it converges in law as $N\to\infty$ to $\frac{1}{2}I_2(a+a^*)$. ∎

2.2.5 Large Deviations Principles

The LDP presented in this section are not restricted to the situation where the functions $\{\Phi_n\;;\;n\geq 1\}$ have the form (8). In what follows we undertake a more general and abstract formulation and we assume that $\{\xi_n\;;\;n\geq 0\}$ is the IPS approximating model (13) associated with a given sequence of continuous functions $\{\Phi_n\;;\;n\geq 1\}$. The LDP for the IPS approximating model (13) for (8) will then be deduced directly from these results.

Large Deviations on Path Space

To prove large deviations principles (**LDP**) for the laws $\{Q_T^{(N)}\;;\;N\geq 1\}$ of the empirical measures $\eta_{[0,T]}^N$ we will always assume that the continuous functions $\{\Phi_n\;;\;n\geq 1\}$ satisfy $(\mathcal{P})_0$. As it has already seen in section 2.2.1, p. 34 the main simplification due to this assumption is that $Q_T^{(N)}$ is absolutely continuous with respect to the distribution $R_T^{(N)}\in\mathbf{M}_1(\mathbf{M}_1(\Sigma_T))$ of the empirical measure associated with N independent random variables with common law $\lambda_{[0,T]}$. In addition, we have that

$$\frac{dQ_T^{(N)}}{dR_T^{(N)}} = \exp\left(NF_T\right) \qquad R_T^{(N)} - \text{a.s.}$$

where $F_T:\mathbf{M}_1(\Sigma_T)\to\mathbb{R}$ is the function defined by

$$F_T(\mu) = \sum_{n=1}^{T}\int_E \log\frac{d\Phi_n(\mu_{n-1})}{d\lambda_n}\,d\mu_n = \int_{\Sigma_T}\log\frac{d\Phi_{[0,T]}(\mu)}{dR_T}\,d\mu \tag{72}$$

In a first stage for analysis it is reasonable to suppose that

$(\mathcal{P})_1$ ***For any*** $n\geq 1$ ***there exists a reference probability measure*** $\lambda_n\in\mathbf{M}_1(E)$ ***such that for all*** $\mu\in\mathbf{M}_1(E)$, $\Phi_n(\mu)\sim\lambda_n$ ***and the function*** $\mathbf{M}_1(E)^2\ni(\mu,\nu)\to \int\log\frac{d\Phi_n(\nu)}{d\lambda_n}\,d\mu$ ***is bounded continuous.***

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

$$I(\mu|\nu) = \int\log\frac{d\mu}{d\nu}\,d\mu$$

if $\mu << \nu$ and $+\infty$ otherwise, Sanov's Theorem and Varadhan's Lemma yields

Theorem 2.26 *Assume that $\{\Phi_n \; ; \; n \geq 1\}$ is a sequence of continuous functions such that $(\mathcal{P})_1$ holds. Then, for any $T \geq 0$, $\{Q_T^{(N)}, \; N \geq 1\}$ satisfies a LDP with good rate function*

$$J_T(\mu) = I(\mu|\Phi_{[0,T]}(\mu))$$

and $\eta_{[0,T]}$ is the unique minimizer of J_T.

Proof:(*Sketch*) Under the assumptions of the theorem, F_T is bounded continuous so that $\{Q_T^{(N)}, \; N \geq 1\}$ satisfies a LDP with good rate function

$$J_T(\mu) = I(\mu|R_T) - F_T(\mu)$$

according to Sanov's Theorem and Varadhan's Lemma (see [51] for instance). ■

Corollary 2.27 *Assume that the functions $\{\Phi_n \; ; \; n \geq 1\}$ are given by (8) and the transitions probability kernels $\{K_n \; ; \; n \geq 1\}$ are Feller and satisfy the following assumption*

$(\mathcal{K})_1'$ ***For any time $n \geq 1$, there exists a reference probability measure $\lambda_n \in \mathbf{M}_1(E)$ such that $K_n(x, \bullet) \sim \lambda_n$ and***

- ***the function $z \mapsto \log \frac{dK_n(x,\bullet)}{d\lambda_n}(z)$ is Lipschitz, uniformly on the parameter $x \in E$, and for any $z \in E$ the map $x \mapsto \frac{dK_n(x,\bullet)}{d\lambda_n}(z)$ is continuous***
- ***there exists a positive number $\epsilon_n \in (0, 1]$ such that***

$$\epsilon_n \leq \frac{dK_n(x, \bullet)}{d\lambda_n} \leq \frac{1}{\epsilon_n}$$

Then, for any $T \geq 0$, $\{Q_T^{(N)}, \; N \geq 1\}$ satisfies a LDP with good rate function J_T.

Condition $(\mathcal{K})_1'$ is stronger than condition $(\mathcal{K})_1$ which has been used in section 2.1.2, p. 24, as a mixing condition to derive exponential stability properties for the limiting measure valued system (8). In LDP settings this hypothesis is more related to a compactness assumption.

Here we present a way to relax $(\mathcal{P})_1$ based on cut-off arguments.

Let $F_T^M : \mathbf{M}_1(\Sigma_T) \to \mathbb{R}$ be the cut-off transformation of F_T given by

$$F_T^M(\mu) = \sum_{n=1}^{T} \int_E \psi^M \left(\log \frac{d\Phi_n(\mu_{n-1})}{d\lambda_n} \right) \, d\mu_n$$

where

$$\psi^M(x) = x \, 1_{|x| \leq M} + \mathrm{sign}(x) \, M \, 1_{|x| > M}$$

Next assumptions relax $(\mathcal{P})_1$

$(\mathcal{L})_0$ ***For any time*** $n \geq 1$***, there exists a reference probability measure*** $\lambda_n \in \mathbf{M}_1(E)$ ***such that for all*** $\mu \in \mathbf{M}_1(E)$***,*** $\Phi_n(\mu) \sim \lambda_n$ ***and the function***

$$(x, \nu) \mapsto \log \frac{d\Phi_n(\nu)}{d\lambda_n}(x)$$

is uniformly continuous w.r.t. x ***(and uniformly w.r.t.*** ν***) and continuous w.r.t.*** ν***.***

$(\mathcal{L})_1$ ***There exist constants*** $c_T < \infty$***,*** $\alpha_T > 1$ ***such that***

$$R_T^{(N)}\left(e^{\alpha_T N F_T}\right) \leq e^{c_T N}$$

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

$$R_T^{(N)}\left(|F_T - F_T^M| > \epsilon\right) \leq e^{-NL_{T,\epsilon}(M)}. \tag{73}$$

$(\mathcal{L})_2$ ***There exist constants*** $\delta_T > 0$ $C_T < \infty$***,*** $D_T < \infty$ ***and a function*** ϵ_T***,*** $\epsilon_T(M)$ ***is going to zero when*** M ***is going to infinity, such that for any*** $\mu \in \mathbf{M}_1(\Sigma_T)$ ***and*** $M \in \mathbb{R} \cup \{\infty\}$

$$\begin{aligned} I(\mu|R_T) - F_T^M(\mu) &\geq \delta_T\, I(\mu|R_T) - C_T \\ |F_T(\mu) - F_T^M(\mu)| &\leq \epsilon_T(M)\,(I(\mu|R_T) + D_T) \end{aligned}$$

Theorem 2.28 *Assume conditions* $(\mathcal{L})_0$*,* $(\mathcal{L})_1$ *and* $(\mathcal{L})_2$ *are satisfied. Then,* $\{Q_T^{(N)} :\ N \geq 1\}$ *satisfies a LDP with good rate function* J_T*.*

Proof:(*Sketch*) Under $(\mathcal{L})_0$ one first check that F_T^M is bounded continuous. The proof is now based on the ideas of Azencott and Varadhan and amounts to replace the functions F_T (which are a priori nor bounded nor continuous) by the functions F_T^M to get the LDP up to a small error ϵ in the rate function by $(\mathcal{L})_1$ and then pass to the limit $M \to \infty$ by $(\mathcal{L})_2$ to let finally $\epsilon \downarrow 0$. ∎

Conditions $(\mathcal{L})_1$ and $(\mathcal{L})_2$ are hard to work with. It is quite remarkable that an exponential moment condition suffices to check $(\mathcal{L})_1$ and $(\mathcal{L})_2$.

Corollary 2.29 *Suppose the functions* $\{\Phi_n\ ;\ n \geq 1\}$ *satisfy next condition*

$(\mathcal{P})_1'$ ***For any time*** $1 \leq n \leq T$ ***there exists a reference probability measure*** $\lambda_n \in \mathbf{M}_1(E)$ ***such that for all*** $\mu \in \mathbf{M}_1(E)$***,*** $\Phi_n(\mu) \sim \lambda_n$ ***and***

- ***For any*** $1 \leq n \leq T$ ***the function***

$$(x, \nu) \mapsto \log \frac{d\Phi_n(\nu)}{d\lambda_n}(x)$$

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

- ***There exist*** $\mathbf{B}(E)$***-measurable functions*** φ ***and*** ψ ***and constants*** $\alpha, \beta \in]1, \infty]$ ***and*** $\epsilon > 0$ ***such that*** $\frac{1}{\alpha} + \frac{1}{\beta} < 1$ ***and for any*** $1 \leq n \leq T$

$$\left| \log \frac{d\Phi_n(\mu)}{d\lambda_n}(x) \right| \leq \varphi(x) + \mu(\psi)$$

and

$$\int \exp\left(\alpha \varphi^{1+\epsilon}\right) d\lambda_n \vee \int \exp\left(\beta \psi^{1+\epsilon}\right) d\lambda_n < \infty \tag{74}$$

Then, $\{Q_T^{(N)} : N \geq 1\}$ *satisfies the LDP with good rate function* J_T*.*

Corollary 2.30 *Assume that the functions* $\{\Phi_n \ ; \ n \geq 1\}$ *are given by (8) and the transitions probability kernels* $\{K_n \ ; \ n \geq 1\}$ *are Feller and satisfy the following assumption*

$(\mathcal{K})_1''$ ***For any time*** $1 \leq n \leq T$ ***there exists a reference probability measure*** $\lambda_n \in \mathbf{M}_1(E)$ ***such that*** $K_n(x, \bullet) \sim \lambda_n$ ***and***

- ***For any time*** $1 \leq n \leq T$ ***the function***

$$z \mapsto \log \frac{dK_n(x, \bullet)}{d\lambda_n}(z)$$

is Lipschitz, uniformly on the parameter $x \in E$***, and for any*** $z \in E$ ***the map***

$$x \mapsto \frac{dK_n(x, \bullet)}{d\lambda_n}(z)$$

is continuous.

- ***There exist a*** $\mathbf{B}(E)$***-measurable function*** φ ***and constants*** $\alpha > 1$ ***and*** $\epsilon > 0$ ***such that for any time*** $1 \leq n \leq T$

$$\left| \log \frac{dK_n(x, \bullet)}{d\lambda_n}(z) \right| \leq \varphi(z) \qquad \text{and} \qquad \int \exp\left(\alpha \varphi^{1+\epsilon}\right) d\lambda_n < \infty$$

Then, $\{Q_T^{(N)} : N \geq 1\}$ *satisfies a LDP with good rate function* J_T*.*

Large Deviations for the Particle Density Profiles

The large deviations results on path space rely largely on the existence of a family of reference distributions $\{\lambda_n : n \geq 1\}$ satisfying condition $(\mathcal{P})_0$ and therefore does not apply to some filtering problems (see section 5). To remove this assumption we shall be dealing with the law $\{P_n^N \, ; n \geq 1\}$, of the particle density profiles $\{\eta_n^N \, ; n \geq 1\}$.

Theorem 2.31 *Assume that the continuous functions* $\{\Phi_n \; ; \; n \geq 1\}$ *satisfy the following condition*

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

$$\mu \tilde{M}(A^c) < \delta \Longrightarrow \Phi_n(\mu)M(A^c) < \epsilon$$

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

Then, for any $n \geq 0$, $\{P_n^{(N)} \; : \; N \geq 1\}$ *obeys a LDP with convex good rate function* H_n *given by*

$$\left\{ \begin{array}{lcl} H_n(\mu) & = & \sup_{V \in \mathcal{C}_b(E)} \left(\mu(V) + \inf_{\nu \in \mathbf{M}_1(E)} \left(H_{n-1}(\nu) - \log\left(\Phi_n(\nu)(e^V)\right)\right)\right), \quad n \geq 1 \\ H_0(\mu) & = & I(\mu|\eta_0) \end{array} \right.$$

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

Proof:(*Sketch*) First we check that $(\mathcal{ET})$ insures that for any time $n \geq 0$ the sequence

$$\{P_n^N : \; N \geq 1\}$$

is exponentially tight (cf. Proposition 2.5 in [35]). To get the desired LDP we proceed by induction on the time parameter. For $n = 0$ the result is a trivial by Sanov's Theorem so we assume that it holds for $(n-1)$. Observe that the moment generating function at rank n is given for any $V \in \mathcal{C}_b(E)$ by

$$\mathbb{E}\left(\exp\left(N\eta_n^N(V)\right)\right) \;=\; \mathbb{E}\left(\left(\Phi_n(\eta_{n-1}^N)(e^V)\right)^N\right) = \mathbb{E}\left(\exp\left(NG_n(\eta_{n-1}^N)\right)\right)$$

with

$$G_n(\eta) \stackrel{\text{def}}{=} \log[\Phi_n(\eta)(e^V)]$$

Since G_n is bounded continuous then Varadhan's Lemma (see for instance [106], Theorem 2.6 p 24) and the induction hypothesis at rank $(n-1)$ imply that

$$\lim_{N\to\infty} \frac{1}{N} \log \int \exp\left(N\mu(V)\right) \, P_n^N(d\mu) = \Lambda_n(V)$$

with

$$\Lambda_n(V) = -\inf_{\mu \in \mathbf{M}_1(E)} \left(H_{n-1}(\mu) - \log\left(\Phi_n(\mu)(e^V)\right)\right)$$

Using the exponential tightness property we are now in position to apply Baldi's Theorem (cf. Theorem 4.5.20 p. 139 and Corollary 4.6.14 in [49]). More precisely, it remains to check that Λ_n is finite valued and Gateaux differentiable. The first point is obvious. To check that Λ_n is differentiable we introduce for $\nu \in \mathbf{M}_1(E)$,

$$I_n^V(\nu) := H_{n-1}(\nu) - \log(\Phi_n(\nu)(e^V)).$$

After some calculations one finds that for any $v \in \mathcal{C}_b(E)$, $\|v\| \leq 1$

$$D\Lambda_n(V)[v] = \sup_{\{\nu : I_n^V(\nu) \leq -\Lambda_n(V)\}} \frac{\int v e^V d\Phi_n(\nu)}{\int e^V d\Phi_n(\nu)}.$$

■

To see that $(\mathcal{ET})$ holds if the functions $\{\Phi_n \ ; \ n \geq 1\}$ are given by (8) we simply notice that, under the assumption $(\mathcal{G})$, for any $n \geq 1$ and for any compact set $A \subset E$,

$$\forall \ \eta \in \mathbf{M}_1(E), \qquad \Phi_n(\eta)(A^c) \ \leq \ a_n^2 \ \eta K_n(A^c)$$

Corollary 2.32 *If the functions $\{\Phi_n \ ; \ n \geq 1\}$ are given by (8) and the transitions probability kernels $\{K_n \ ; \ n \geq 1\}$ are Feller then for any $n \geq 0$, $\{P_n^N \ : \ N \geq 1\}$ obeys a LDP with convex good rate function H_n.*

2.3 Branching Particle Systems Variants

The research literature abounds with variations on the IPS model (13). Each of these variants is intended to make the selection and/or the mutation transition more efficient in some sense. As a guide to their usage, this section presents a survey of what is currently the state of the art.

The analysis of the convergence of all these alternative schemes is actually not complete. We therefore gather together here several discussions and results which we hope will assist in giving an illustration of how the analysis developed in section 2.2 may be applied.

The first scheme discussed in section 2.3.1 concerns a genetic type algorithm with periodic selection times. The key idea here is to use the stability properties of the limiting system to produce a more efficient IPS. We will show that the resulting algorithm can be reduced to the latter through a suitable state space basis. In this specific situation all the results developed in section 2.2 may be applied. It will also be shown that the convergence exponent in the uniform convergence result 2.11 improves the one obtained for the generic IPS.

Section 2.3.2 presents a way to produce an IPS whose mutation transitions depends on the fitness functions. In nonlinear filtering settings this scheme is often referred as an IPS with conditional mutations. In this situation the fitness functions and therefore the mutation transitions depend on the observations record so that the particles are more likely to track the signal process.

In section 2.3.3 we present "less time-consuming" selection transitions such as the *Remainder Stochastic Sampling* and other BIPS alternatives including *Bernoulli, Poisson and Binomial* branching distributions. These selection transitions can be related to the classical weighted bootstrap theory as well as genetic algorithms theory (the book [10] includes a useful survey on these two subjects).

In filtering settings the choice of the mutation transitions $\{K_n \; ; \; n \geq 1\}$ is dictated by the form of the signal. It may happen that these transitions are degenerated (i.e. deterministic) and therefore the IPS approximating models will not work in practice since after a finite number of steps we would get a single deterministic path. The last section 2.3.4 presents a way to regularize such degenerated mutation transitions. This regularization step has been introduced in [59]. We shall indicate how the results of section 2.2 are applied.

In practice the most efficient IPS approximating model is the one obtained by combining conditional mutations and periodic selections. The choice of the best selection scheme is more an art form. There is a balance between *time-consuming* and *efficiency*. The less time consuming selections seems to be *Remainder Stochastic Sampling* and *Bernoulli branching selections*. In the last case the size of the system is not fixed but random and explosion is possible (cf. [36, 38] and 4) section 2.3.3).

The most important and unsolved problem is to understand and compare the influence of the various selections schemes on the long time behavior of the BIPS approximating models. The interested reader will find that although we have restricted ourselves to the relatively less complicated generic IPS model (13) most of our general techniques apply across these more complex BIPS variants.

There are many open problems such as that of finding the appropriate analog of the results of section 2.2 for the BIPS approximating models presented in section 2.3.3. This study has been started in [22] and in [25] but many open problems such as the fluctuations remain to be answered. A related question will be to find a criterion for determining the "optimal branching" transition. This problem will probably lead to difficult optimization problems since this criterion should be related to the long time behavior of the BIPS approximating schemes.

2.3.1 Periodic Interactions/Selections

The IPS with periodic selections discussed here has been introduced in [36] as a way to combine the stability properties of the limiting system with the long time behavior of the particle scheme.

The prediction/mutation of the former include exploration paths of a given length $T \geq 1$ and the selection transition is used every T steps and it involves T fitness functions. Our immediate goal is to show that the former genetic algorithm can be reduced to the latter through a suitable state space basis. To this end we need to introduce some additional notations.

To each $p \in \{1, \ldots, T\}$ we associate a sequence of meshes $\{t_n^{(T,p)} \; ; \; n \geq 0\}$ by setting

$$t_0^{(T,p)} = 0 \qquad \text{and} \qquad \forall\, n \geq 1, \qquad t_n^{(T,p)} = (n-1)T + p$$

If we write $\Delta_n = t_n - t_{n-1}$ for any $n \geq 1$ we clearly have that

$$\Delta_1 = p \qquad \text{and} \qquad \forall\, n > 1, \qquad \Delta_n = T$$

The parameter T is called the selection period, n will denote the time step and the parameter p will only be used to cover all the time space basis $\mathbb{N}$. The construction below will depend on the pair parameter (T,p) but to clarify the presentation we simplify the notations suppressing the pair parameter (T,p) so that we simply note t_n instead of $t_n^{(T,p)}$.

We also notice that the distributions given by

$$\forall n \geq 0\,, \qquad \mu_n = \eta_{t_n} \times K_{t_n+1} \times \ldots \times K_{t_{n+1}-1} \in \mathbf{M}_1(E^{\Delta_{n+1}})$$

are solutions of the measure valued process

$$\mu_n = \Phi_n^{(p)}(\mu_{n-1}) \qquad n \geq 1 \tag{75}$$

where $\Phi_n^{(p)} : \mathbf{M}_1(E^{\Delta_n}) \to \mathbf{M}_1(E^{\Delta_{n+1}})$ is the continuous function given by

$$\forall\, \mu \in \mathbf{M}_1(E^{\Delta_n}), \qquad \Phi_n^{(p)}(\mu) = \Psi_n^{(p)}(\mu)\mathcal{K}_n^{(p)}$$

and $\mathcal{K}_n^{(p)}$ and $\Psi_n^{(p)}$ are defined as follows.

- $\Psi_n^{(p)} : \mathbf{M}_1(E^{\Delta_n}) \to \mathbf{M}_1(E^{\Delta_n})$ is the continuous function defined for any test function $f \in \mathcal{B}_b(E^{\Delta_n})$ by setting

$$\Psi_n^{(p)}(\mu)(f) = \frac{\mu(g_n^{(p)}\, f)}{\mu(g_n^{(p)})} \quad \text{with} \quad g_n^{(p)}(x) = \prod_{q=1}^{\Delta_n} g_{t_{n-1}+q}(x_q).$$

- $\mathcal{K}_n^{(p)}$ is the transition probability kernel from E^{Δ_n} to $E^{\Delta_{n+1}}$ given by

$$\mathcal{K}_n^{(p)}((x_1,\ldots,x_{\Delta_n}), d(z_1,\ldots,z_{\Delta_{n+1}})) = K_{t_n}(x_{\Delta_n}, dz_1) \times \ldots$$
$$\ldots \times K_{t_{n+1}-1}(z_{\Delta_{n+1}-1}, dz_{\Delta_{n+1}})$$

The IPS associated with (75) is now defined as a Markov chain $\{\zeta_n\,;\, n \geq 0\}$ with product state spaces $\{(E^{\Delta_{n+1}})^N\,;\, n \geq 0\}$ where N is the number of particles and $\{\Delta_{n+1}\,;\, n \geq 0\}$ the selection periods.
The initial particle system $\zeta_0 = (\zeta_0^1, \ldots, \zeta_0^N)$ takes values in $(E^{\Delta_1})^N = (E^p)^N$ and it is given by

$$P(\zeta_0 \in dx) \;=\; \prod_{i=1}^N \mu_0(dx^i)$$

and the transition of the chain is now given by

$$\begin{aligned} P(\zeta_n \in dx | \zeta_{n-1} = z) &= \prod_{i=1}^N \Phi_n^{(p)}\left(\frac{1}{N}\sum_{j=1}^N \delta_{z^j}\right)(dx^i) \\ &= \prod_{i=1}^N \sum_{j=1}^N \frac{g_n^{(p)}(z^j)}{\sum_{k=1}^N g_n^{(p)}(z^k)} K_n^{(p)}(z^i, dx^i) \end{aligned}$$

where $dx = dx^1 \times \ldots \times dx^N$ is an infinitesimal neighborhood of the point $x = (x^1, \ldots, x^N) \in (E^{\Delta_{n+1}})^N$ and for any $1 \leq i \leq N$, $z^i = (z^i_1, \ldots, z^i_{\Delta_n}) \in E^{\Delta_n}$.

If we denote

$$\forall\ n \geq 0, \qquad \zeta_n \ = \ (\xi_{t_n}, \ldots, \xi_{t_{n+1}-1})$$

we see that the former algorithm is indeed a genetic type algorithm with T-periodic selection/updating transitions. Between the dates t_n and t_{n+1} the particles evolves randomly according to the transition probability kernel of the signal and the selection mechanism takes place at each time t_n , $n \geq 1$.

As announced in the beginning of the section, this IPS scheme with periodic selection times is reduced to the one presented in section 1.3 through a suitable state space basis so that the analysis given in section 2.2 applies to this situation.

The uniform results with respect to time given in Theorem 2.11 can be improved by choosing a suitable period which depends on the stability properties of the limiting system and on the number of particles. More precisely, if we denote by $\eta^N_{t_n}$ the particle density profiles given by

$$\eta^N_{t_n} = \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi^i_{t_n}}$$

then we have the the following theorem.

Theorem 2.33 *Assume that the limiting dynamical system (8) is exponentially asymptotically stable in the sense that there exist some positive constant $\gamma > 0$ such that for any function $f \in \mathcal{B}_{\rm b}(E)$ with $\|f\| \leq 1$*

$$\forall\ \mu, \nu \in \mathbf{M}_1(E),\ \forall\ T \geq 0, \qquad \sup_{p\geq 0} |\Phi_{p,p+T}(\mu)(f) - \Phi_{p,p+T}(\nu)(f)| \ \leq \ e^{-\gamma T}$$

If the selection period is chosen so that $T = T(N) \stackrel{\rm def}{=} \left[\frac{1}{2}\frac{\log N}{\gamma+\gamma'}\right] + 1$ then for any $f \in \mathcal{B}_{\rm b}(E)$, $\|f\| \leq 1$, we have the uniform bound given by

$$\sup_{n\geq 0} \mathbb{E}\left(|\eta^N_{t_n} f - \eta_{t_n} f|\right) \leq \frac{4e^{2\gamma'}}{N^{\beta/2}} \qquad \text{with} \qquad \beta = \frac{\gamma}{\gamma+\gamma'}, \qquad \gamma' = 2\log a \qquad (76)$$

Remarks 2.34:

- Although we have not yet checked more general $\mathbb{L}^p$-bounds or uniform convergence results over a suitable class of functions, the proof of Theorem 2.33 essentially follows the same arguments as in the proof of Theorem 2.11.
- It is also worth observing that the choice of the selection period depends on the stability properties of the limiting system as well as on the number of particles.
- Another remark is that the critical exponent β resulting from the proof of Theorem 2.33 is now sharp in the following sense: if the fitness functions are constant then, without loss of generality, we may chose $a = 1$. In this specific situation the critical exponent $\beta = 1$ which is again the characteristic exponent of the weak law of large numbers.

- Our last remark is that periodic selections are very efficient and have a specific interpretation in nonlinear filtering settings. We recall that in this situation the fitness functions are related to the observation process. Roughly speaking the selection transition evaluates the population structure and allocates reproductive opportunities in such a way that these particles which better match with the current observation are given more chance to "reproduce". This stabilizes the particles around certain values of the real signal in accordance with its noisy observations.

  It often appears that a single observation data is not really sufficient to distinguish in a clearly manner the relative fitness of individuals. For instance this may occurs in high noise environment. In this sense the IPS with periodic selections allows particles to learn the observation process between the selection dates in order to produce more effective selections.

2.3.2 Conditional Mutations

The IPS with conditional mutations is defined in a natural way as the IPS approximating model associated to the measure valued process (23) presented in section 2.1.1, p. 22. To clarify the presentation it is convenient here to change the time parameter in the fitness functions $\{g_n \ ; \ n \geq 1\}$ so that we will write g_n instead of g_{n+1}. Using these notations (23) is defined by the recursion

$$\forall\, n \geq 1, \qquad \widehat{\eta}_n \;=\; \widehat{\Phi}_n(\widehat{\eta}_{n-1}) \tag{77}$$

where $\widehat{\Phi}_n(\eta) = \widehat{\Psi}_n(\eta)\widehat{K}_n$ and for any $f \in \mathcal{B}_{\rm b}(E)$,

$$\widehat{\Psi}_n(\eta)(f) \stackrel{\rm def.}{=} \frac{\eta(\widehat{g}_n f)}{\eta(\widehat{g}_n)}, \qquad \widehat{K}_n f \stackrel{\rm def.}{=} \frac{K_n(g_n f)}{K_n(g_n)}, \qquad \widehat{g}_n \stackrel{\rm def.}{=} K_n(g_n).$$

As noticed in section 2.1.1 this model has the same form as (8). It is then clear that all the results developed in section 2.1.2 and section 2.2 can be translated in these settings.

In contrast to (8) we also note that the prediction transitions here depends on the fitness functions and therefore the corresponding IPS approximating model will involve mutation transitions which also depend on these functions. More precisely, let $\widehat{\zeta}_n = \left(\widehat{\zeta}_n^1, \ldots, \widehat{\zeta}_n^N\right) \in E^N$ be the N-IPS associated with (77) and defined as in (13) by the transition probability kernels

$$P\left(\widehat{\zeta}_n \in dx \,|\, \widehat{\zeta}_{n-1} = z\right) = \prod_{p=1}^{N} \widehat{\Phi}_n(m(z))(dx^p)$$

where $dx \stackrel{\rm def}{=} dx^1 \times \cdots \times dx^N$ is an infinitesimal neighborhood of the point $x = (x^1, \ldots, x^N) \in E^N$, $z = (z^1, \ldots, z^N) \in E^N$. As before we see that the above transition involves two mechanisms

$$\widehat{\zeta}_{n-1} \xrightarrow{\text{Selection}} \tilde{\zeta}_n \xrightarrow{\text{Mutation}} \widehat{\zeta}_n$$

which can also be modelled as follows

$$\begin{aligned} P(\tilde{\zeta}_n \in dx \,|\, \widehat{\zeta}_{n-1} = z) &= \prod_{p=1}^{N} \sum_{i=1}^{N} \frac{\widehat{g}_n(z^i)}{\sum_{j=1}^{N} \widehat{g}_n(z^j)}\, \delta_{z^i}(dx^p) \\ P(\widehat{\zeta}_n \in dz \,|\, \tilde{\zeta}_n = x) &= \prod_{p=1}^{N} \widehat{K}_n(x^p, dz^p) \end{aligned}$$

As we already noticed in section 2.1.1 the fitness functions $\{\widehat{g}_n \; ; \; n \geq 0\}$ and the transitions $\{\widehat{K}_n \; ; \; n \geq 1\}$ involve integrations over the whole state space E so that another level of approximation is in general needed.

Nevertheless let us work out an example in which these two objects have a simple form

Example 2.35 *Let us suppose that $E = \mathbb{R}$ and the fitness functions $\{g_n \; ; \; n \geq 1\}$ and the transitions $\{K_n \; ; \; n \geq 1\}$ are given by*

$$g_n(x) = \; e^{-\frac{1}{2\, r_n}(y_n - c_n . x)^2} \qquad K_n(x, dz) = \frac{1}{\sqrt{2\pi \; q_n}}\; e^{-\frac{1}{2\, q_n}(z - a_n(x))^2}$$

where $a_n : \mathbb{R} \to \mathbb{R}$, $q_n > 0$ for any $n \geq 1$ and $r_n > 0$, $c_n, y_n \in \mathbb{R}$, for any $n \geq 0$. In this situation one gets easily

$$\begin{aligned} \widehat{K}_n(x, dz) &= \frac{1}{\sqrt{2\;\pi\;|s_n|}} \exp\left(-\frac{1}{2\;|s_n|}\left(z - \left[a_n(x) + s_n\, c_n\, r_n^{-1}\,(y_n - c_n\; a_n(x))\right]\right)^2\right) \\ \text{and } \; \widehat{g}_n(x) &= \frac{1}{\sqrt{2\;\pi\;|q_n|\,|r_n|/|s_n|}} \exp\left(-\frac{1}{2\;|q_n|\,|r_n|/|s_n|}\,(y_n - c_n\; a_n(x))^2\right) \end{aligned}$$

with $s_n = (q_n^{-1} + c_n\; r_n^{-1}\; c_n)^{-1}$.

One idea to approximate the transition $\widehat{\zeta}_{n-1} \to \widehat{\zeta}_n$ is to introduce an auxiliary branching mechanism.

$$\widehat{\zeta}_{n-1} \xrightarrow{\text{Branching}} \zeta_n \xrightarrow{\text{Selection/Mutation}} \widehat{\zeta}_n$$

The branching transition is defined inductively as follows.
At each time $(n-1)$ each particle $\widehat{\zeta}^i_{n-1}$ branches independently of each other into M-auxiliary particles with common law $K_n\left(\widehat{\zeta}^i_{n-1}, .\right)$, that is for any $1 \leq i \leq N$,

$$\widehat{\zeta}^i_{n-1} \xrightarrow{\text{Branching}} \zeta^{(i)}_n \stackrel{\text{def.}}{=} \left(\zeta^{i,1}_n, \ldots, \zeta^{i,M}_n\right)$$

where $\left(\zeta^{i,1}_n, \ldots, \zeta^{i,M}_n\right)$ are (conditionally) M-independent particles with common law $K_n\left(\widehat{\zeta}^i_{n-1}, .\right)$.
At the end of this branching step the system consists in $N \times M$ particles

$$\zeta_n \stackrel{\text{def.}}{=} \left(\zeta^{(1)}_n, \ldots, \zeta^{(N)}_n\right) \in \underbrace{E^M \times \ldots \times E^M}_{N\text{-times}}$$

If the parameter M is sufficiently large then, in some sense, the empirical measures associated with each sub-group of M particles is an approximating measure of $K_n\left(\widehat{\zeta}^i_{n-1},.\right)$, that is

$$\forall\ f\in\mathcal{B}_{\mathrm{b}}(E),\qquad K_n^{(M)}(f)(\zeta_n^{(i)})\quad\underset{M\to\infty}{\longrightarrow}\quad K_n(f)(\widehat{\zeta}^i_{n-1})\tag{78}$$

where $K_n^{(M)}$ is the transition probability kernel from E^M into E given for any $x\in E^M$ and $f\in\mathcal{B}_{\mathrm{b}}(E)$ by the formula

$$K_n^{(M)}(x,.)\stackrel{\text{def.}}{=}\frac{1}{M}\sum_{i=1}^M\delta_{x^i}\qquad\text{and}\qquad (K_n^{(M)}f)(x)=\int_E f(z)\ K_n^{(M)}(x,dz)$$

Using the above notations we also have, in some sense, that

$$\widehat{g}_n^{(M)}(\zeta_n^{(i)})\stackrel{\text{def.}}{=}K_n^{(M)}(g_n)(\zeta_n^{(i)})\quad\underset{M\to\infty}{\longrightarrow}\quad K_n(g_n)(\widehat{\zeta}^i_{n-1})=\widehat{g}_n(\widehat{\zeta}^i_{n-1})\tag{79}$$

and for any $f\in\mathcal{B}_{\mathrm{b}}(E)$

$$\widehat{K}_n^{(M)}(f)(\zeta_n^{(i)})\stackrel{\text{def.}}{=}\frac{K_n^{(M)}(g_n\ f)\left(\zeta_n^{(i)}\right)}{K_n^{(M)}(g_n)(\zeta_n^{(i)})}\quad\underset{M\to\infty}{\longrightarrow}\quad\widehat{K}_n(f)(\widehat{\zeta}^i_{n-1})\tag{80}$$

Finally if we combine (79) and (80) one gets an M-approximation of the desired transition

$$\sum_{i=1}^N\frac{\widehat{g}_n^{(M)}(\zeta_n^{(i)})}{\sum_{j=1}^N\widehat{g}_n^{(M)}(\zeta_n^{(i)})}\ \widehat{K}_n^{(M)}\left(\zeta_n^{(i)},.\right)\quad\underset{M\to\infty}{\longrightarrow}\quad\widehat{\Phi}_n\left(\frac{1}{N}\sum_{i=1}^N\delta_{\widehat{\zeta}^i_{n-1}}\right)\tag{81}$$

The next particle system $\widehat{\zeta}_n=\left(\widehat{\zeta}^1_n,\ldots,\widehat{\zeta}^N_n\right)$ simply consists in N conditionally independent particles with common law the left hand side of (81).

Our new BIPS is now defined by the following Markov Model

$$\widehat{\zeta}_{n-1}\xrightarrow{\text{Branching}}\zeta_n=\left(\zeta_n^{(1)},\ldots,\zeta_n^{(N)}\right)\xrightarrow{\text{Selection}}\tilde{\zeta}_n=\left(\tilde{\zeta}_n^{(1)},\ldots,\tilde{\zeta}_n^{(N)}\right)\xrightarrow{\text{Mutation}}\widehat{\zeta}_n$$

with the following transitions

- **Branchings**: The branching transition

  $$\widehat{\zeta}_{n-1}\in E^N\longrightarrow\zeta_n=\left(\zeta_n^{(1)},\ldots,\zeta_n^{(N)}\right)\in(E^M)^N$$

  is defined by

  $$P\left(\zeta_n\in dx^{(1)}\times\ldots dx^{(N)}\,\middle|\,\widehat{\zeta}_{n-1}=z\right)=\prod_{i=1}^N(K_n(z^i,\cdot))^{\otimes M}(dx^{(i)})$$

  where $dx^{(i)}=dx^{i,1}\times\ldots\times dx^{i,M}$ is an infinitesimal neighborhood of the point $x^{(i)}=(x^{i,1},\ldots,x^{i,M})\in E^M$, $z=(z^1,\ldots,z^N)\in E^N$.

- **Selection**: The selection transition

$$\zeta_n = \left(\zeta_n^{(1)}, \ldots, \zeta_n^{(N)}\right) \in (E^M)^N \longrightarrow \tilde{\zeta}_n = \left(\tilde{\zeta}_n^{(1)}, \ldots, \tilde{\zeta}_n^{(N)}\right) \in (E^M)^N$$

is defined by

$$\begin{aligned} & P\left(\tilde{\zeta}_n \in dy^{(1)} \times \ldots \times dy^{(N)} \middle| \zeta_n = (x^{(1)}, \ldots, x^{(N)})\right) \\ & = \prod_{i=1}^N \sum_{p=1}^N \frac{\widehat{g}_n^{(M)}(x^{(p)})}{\sum_{q=1}^N \widehat{g}_n^{(M)}(x^{(q)})}\ \delta_{x^{(p)}}(dy^{(i)}) \end{aligned}$$

- **Mutation:** The mutation transition

$$\tilde{\zeta}_n = \left(\tilde{\zeta}_n^{(1)}, \ldots, \tilde{\zeta}_n^{(N)}\right) \in (E^M)^N \longrightarrow \widehat{\zeta}_n \in E^N$$

is defined by

$$P\left(\widehat{\zeta}_n \in dz \middle| \tilde{\zeta}_n = (y^{(1)}, \ldots, y^{(N)})\right) = \prod_{i=1}^N \widehat{K}_n^{(M)}\left(y^{(i)}, dz^i\right)$$

This algorithm has been introduced in [33]. In this work exponential rates of convergence and $\mathbb{L}^p$-mean error bounds are discussed. The LDP associated with such branching strategy and comparisons with the rates presented in section 2.2.5 are described in [35].

2.3.3 Branching Selections

Roughly speaking, the selection transition is intended to improve the quality of the system by given individuals of "higher quality" to be copied into the next generation. In other words, selection focuses the evolution of the system on promising regions in the state space by allocating reproductive opportunities in such a way that those particles which have a higher fitness are given more chances to give an offspring than those which have a poorer fitness. They are number of ways to approximate the updating transitions but they are all based on the same natural idea. Namely, how to approximate an updated empirical measure of the following form

$$\Psi_n\left(\frac{1}{N}\sum_{i=1}^N \delta_{x^i}\right) = \sum_{i=1}^N \frac{g_n(x^i)}{\sum_{j=1}^N g_n(x^j)}\ \delta_{x^i} \tag{82}$$

by a new probability measure with atoms of size integers multiples of $1/N$?

In the generic IPS approximating model (13) this approximation is done by sampling N-independent random variables $\{\widehat{x}^i\ ;\ 1 \leq i \leq N\}$ with common law (82) and the corresponding approximating measure is given by

$$\frac{1}{N}\sum_{i=1}^N \delta_{\widehat{x}^i} = \sum_{i=1}^N \frac{M^i}{N}\ \delta_{x^i}$$

where

$$\left(M^1,\ldots,M^N\right) \stackrel{\text{def.}}{=} \text{Multinomial}\left(N, \frac{g_n(x^1)}{\sum_{j=1}^N g_n(x^j)}, \ldots, \frac{g_n(x^N)}{\sum_{j=1}^N g_n(x^j)}\right)$$

Using these notations the random and $\mathbb{N}$-valued random variables can be regarded as random number of offsprings created at the positions $(x^1,\ldots,x^N)$.

The above question is strongly related to weighted bootstrap and genetic algorithms theory (see for instance [10] and references therein). In this connection the above multinomial approximating strategy can be viewed as a weighted Efron bootstrap.
It is well known that sampling according to a multinomial may be "time consuming" mainly because it requires a sorting of the population. As in classical bootstrap literature the other idea consists in using independent random numbers $\left(M^1,\ldots,M^N\right)$ distributed according a suitably chosen branching law. In what follows we present an abstract BIPS approximating model which enables a unified description of several classes of branching laws that can be used in practice including Bernoulli, binomial and Poisson distributions.

Abstract BIPS Model

The abstract BIPS model will be a two step Markov chain

$$(N_n,\xi_n) \xrightarrow{Branching} (\widehat{N}_n,\widehat{\xi}_n) \xrightarrow{Mutation} (N_{n+1},\xi_{n+1}) \tag{83}$$

with product state space

$$\mathcal{E} = \bigcup_{\alpha\in\mathbb{N}} (\{\alpha\}\times E^\alpha)$$

with the convention $E^\alpha = \{\triangle\}$ a cemetery if $\alpha = 0$. We will note

$$\mathcal{F} = \{F_n, \widehat{F}_n \ : \ n \geq 0\}$$

the canonical filtration associated to (83) so that

$$F_n \subset \widehat{F}_n \subset F_{n+1}$$

The points of the set E^α, $\alpha \geq 0$ are called particle systems and are mostly denoted by the letters x and z. The parameter $\alpha \in \mathbb{N}$ represents the size of the system. The initial number of particles $N_0 \in \mathbb{N}$ is a fixed non-random number which represents the precision parameter of the BIPS algorithm.

The evolution in time of the BIPS is defined inductively as follows.

- At the time $n = 0$:
  The initial particle system $\xi_0 = (\xi_0^1,\ldots,\xi_0^{N_0})$ consists in N_0 independent and identically distributed particles with common law ε_0.

- Evolution in time:
  At the time n, the particle system ξ_n consists in N_n particles.
  If $N_n = 0$ the particle system died and we let $\widehat{N}_n = 0$ and $N_{n+1} = 0$.
  Otherwise the branching correction is defined as follows

1. **Branching Correction:**
   When $N_n > 0$ we associate to $\xi_n = (\xi_n^1, \ldots, \xi_n^{N_n}) \in E^{N_n}$ the weight vector $W_n = (W_n^1, \ldots, W_n^{N_n}) \in \mathbb{R}^{N_n}$ given by

$$\sum_{i=1}^{N_n} W_n^i \, \delta_{\xi_n^i} = \Psi_{n+1}(m(\xi_n)) \qquad \text{where} \quad m(\xi_n) = \frac{1}{N_n} \sum_{i=1}^{N_n} \delta_{\xi_n^i}$$

   Then, each particle ξ_n^i, $1 \leq i \leq N_n$, branches into a random number of offsprings M_n^i, $1 \leq i \leq N_n$ and the mechanism is chosen so that for any $f \in \mathcal{B}_{\mathrm{b}}(E)$

$$\mathbb{E}\left(\sum_{i=1}^{N_n} M_n^i f(\xi_n^i) \, | F_n \right) = N_n \Psi_{n+1}(m(\xi_n)) f \tag{84}$$

   and there exists a finite constant $C < \infty$ so that

$$\mathbb{E}\left(\left[\sum_{i=1}^{N_n} M_n^i f(\xi_n^i) - N_n \Psi_{n+1}(m(\xi_n)) f \right]^2 |F_n \right) \leq C \, N_n \|f\|^2 \tag{85}$$

   At the end of this stage the particle system $\widehat{\xi}_n$ consists in $\widehat{N}_n = \sum_{i=1}^{N_n} M_n^i$ particles denoted by

$$\widehat{\xi}_n^i = \xi_n^k \qquad 1 \leq k \leq N_n \qquad \sum_{l=1}^{k-1} M_n^l + 1 \leq i \leq \sum_{l=1}^{k-1} M_n^l + M_n^k \tag{86}$$

2. **Mutation transition:**

   If $\widehat{N}_n = 0$ the particle system dies and $N_{n+1} = 0$.
   Otherwise, each particle moves independently of each other starting off from the parent particle branching site $\widehat{\xi}_n^i$ with law $K_{n+1}(\widehat{\xi}_n^i, dx)$, $1 \leq i \leq \widehat{N}_n$, for any $1 \leq i \leq \widehat{N}_n$. During this transition the total number of particles doesn't change $(N_{n+1} = \widehat{N}_n)$ and the mechanism can be summarized as follows, for any $\alpha \geq 0$ and $z \in E^\alpha$

$$P\left(\xi_{n+1} \in dx | \widehat{\xi}_n = z, \widehat{N}_n = \alpha \right) = \prod_{i=1}^{\alpha} K_{n+1}(z^i, dx^i)$$

   where $dx = dx^1 \times \ldots \times dx^\alpha$ is an infinitesimal neighborhood of $x \in E^\alpha$ with the conventions $dx = \{\triangle\}$ and $\prod_{i=1}^{\alpha} = 1$ if $\alpha = 0$.

The approximation of the flow of distributions $\{\eta_n \; ; \; n \geq 0\}$ by the particle density profiles

$$\eta_n^N \stackrel{\text{def.}}{=} \frac{1}{N_0} \sum_{i=1}^{N_n} \delta_{\xi_n^i}$$

is guaranteed by the following theorem.

Theorem 2.36 *If the branching selection law satisfy (84) and (85) then, the total mass process $N = (N_n)_{n \geq 0}$ is a non-negative integer valued martingale with respect to the filtration $F = (F_n)_{n \geq 0}$ with the following properties*

$$\forall\, n \geq 0, \qquad \mathbb{E}\left(\sup_{0 \leq k \leq n} \left(\frac{N_k}{N_0} - 1 \right)^2 \right) \;\leq\; \frac{C\,n}{N_0} \qquad \textit{and} \qquad P\,(N_n = 0) \leq \frac{C\,n}{N_0} \tag{87}$$

In addition, for any $n \geq 0$ and $f \in \mathcal{B}_{\mathrm{b}}(E)$, $\|f\| \leq 1$ we have that

$$\mathbb{E}\left[\left(\eta_n^N(f) - \eta_n(f) \right)^2 \right] \leq \frac{B_n}{N_0}$$

for some finite constant B_n which only depends on the time parameter n.

Branching Selections

Here we present several examples of branching laws which satisfy conditions (84) and (85). The first one is known as the *Remainder Stochastic Sampling* in genetic algorithms literature. It has been presented for the first time in [6, 7]. From a pure practical point of view this sampling technique seems to be the more efficient since it is extremely time saving and if the BIPS model is only based on this branching selection scheme then the size of the system remains constant.

1) Remainder Stochastic Sampling

In what follows we denote by $[a]$ (resp. $\{a\} = a - [a]$) the integer part (resp. the fractional part) of $a \in \mathbb{R}$.
At each time $n \geq 0$, each particle ξ_n^i branches directly into a fixed number of offsprings

$$\forall\, 1 \leq i \leq N_n, \qquad \overline{M}_n^i \;\; df \;\; [N_n W_n^i]$$

so that the intermediate population consists in $\overline{N}_n \stackrel{\text{def.}}{=} \sum_{i=1}^{N_n} \overline{M}_n^i$ particles. To prevent extinction and to keep the size of the system fixed it is convenient to introduce in this population $\tilde{N}_n$ additional particles with

$$\tilde{N}_n \stackrel{\text{def.}}{=} N_n - \overline{N}_n = \sum_{i=1}^{N_n} \{N_n W_n^i\}$$

One natural way to do this is to introduce the additional sequence of branching numbers

$$\left(\tilde{M}_n^1, \ldots, \tilde{M}_n^{N_n} \right) \stackrel{\text{def.}}{=} \text{Multinomial}\left(\tilde{N}_n, \frac{\{N_n W_n^1\}}{\sum_{j=1}^{N_n} \{N_n W_n^j\}}, \ldots, \frac{\{N_n W_n^{N_n}\}}{\sum_{j=1}^{N_n} \{N_n W_n^j\}} \right) \tag{88}$$

More precisely, if each particle ξ_n^i again produces a number of $\tilde{M}_n^i$ additional offsprings, $1 \leq i \leq N_n$, then the total size of the system is kept constant.

At the end of this stage, the particle system $\widehat{\xi}_n$ again consists in N_n particles denoted by

$$\begin{array}{rclll}
\widehat{\xi}_n^i &=& \xi_n^k & 1 \le k \le N_n & \sum_{l=1}^{k-1} \overline{M}_n^l + 1 \le i \le \sum_{l=1}^{k-1} \overline{M}_n^l + \overline{M}_n^k \\
\widehat{\xi}_n^{\overline{N}_n+i} &=& \xi_n^k & 1 \le k \le N_n & \sum_{l=1}^{k-1} \tilde{M}_n^l + 1 \le i \le \sum_{l=1}^{k-1} \tilde{M}_n^l + \tilde{M}_n^k
\end{array}$$

The multinomial random numbers (88) can also be defined as follows

$$\tilde{M}_n^k = \mathrm{Card}\left\{1 \le j \le \tilde{N}_n \ ; \ \tilde{\xi}_n^j = \xi^k\right\} \qquad 1 \le k \le N_n$$

where $(\tilde{\xi}_n^1, \ldots, \tilde{\xi}_n^{\tilde{N}_n})$ are $\tilde{N}_n$ independent random variables with common law

$$\sum_{i=1}^{N_n} \frac{\{N_n W_n^i\}}{\sum_{j=1}^{N_n} \{N_n W_n^j\}} \, \delta_{\xi_n^i}$$

It is easily checked that (84) is satisfied and (85) holds for $C = 1$.

Let us now present some classical examples of independent branching numbers that satisfy the non bias condition (84) and the $\mathbb{L}^2$-condition (85).

2) Bernoulli branching numbers:

The Bernoulli branching numbers were introduced in [23, 21] and further developed in [25].

They are defined as a sequence $M_n = (M_n^i, \ 1 \le i \le N_n)$ of conditionally independent random numbers with respect to F_n with distribution given for any $1 \le i \le N_n$ by

$$P(M_n^i = k | F_n) = \left\{ \begin{array}{rl} \{N_n W_n^i\} & \text{if} \quad k = [N_n W_n^i] + 1 \\ 1 - \{N_n W_n^i\} & \text{if} \quad k = [N_n W_n^i] \end{array} \right.$$

In addition it can be seen from the relation $\sum_{i=1}^{N_n} (N_n W_n^i) = N_n$ that at least one particle has one offspring (cf. [23] for more details). Therefore using the above branching correction the particle system never dies.

It is also worth observing that the Bernoulli branching numbers are defined as in the *Remainder Stochastic Sampling* by replacing the multinomial remainder branching law (88) by a sequence of N_n independent Bernoulli random variables $\left(\tilde{M}_n^1, \ldots, \tilde{M}_n^{N_n}\right)$ given by

$$P(\tilde{M}_i^{N_n} = 1 | F_n) = 1 - P(\tilde{M}_i^{N_n} = 0 | F_n) = \{N_n W_n^i\}$$

3) Poisson branching numbers:

The Poisson branching numbers are defined as a sequence $M_n = (M_n^i,\ 1 \le i \le N_n)$ of conditionally independent random numbers with respect to F_n with distribution given for any $1 \le i \le N_n$ by

$$\forall\, k \ge 0, \qquad P(M_n^i = k|F_n) \;=\; \exp\left(-N_n W_n^i\right)\ \frac{(N_n W_n^i)^k}{k!}$$

4) Binomial branching numbers:

These numbers are defined as a sequence $M_n = (M_n^i,\ 1 \le i \le N_n)$ of conditionally independent random numbers with respect to F_n with distribution given for any $1 \le i \le N_n$ by

$$\forall\, 0 \le k \le N_n, \qquad P(M_n^i = k|F_n) \;=\; \binom{N_n}{k}\ (W_n^i)^k\ (1 - W_n^i)^{N_n - k}$$

All of these models are described in full details in [25]. In particular it is shown that the BIPS with multinomial branching laws arises by conditioning a BIPS with Poisson branching laws. The law of large numbers and large deviations for the BIPS model with Bernoulli branching laws are studied in [25] and [22].

The convergence analysis of these BIPS approximating schemes is still in progress. They are many open problems such as that of finding the analog of the Donsker and Glivenko-Cantelli Theorems as well as the study of their long time behavior. This last question is maybe the most important one. The main difficulty here is that the total size process $\{N_n\ ;\ n \ge 0\}$ is an F-martingale with predictable quadratic variation

$$A_n = N_0^2 + \sum_{p=1}^{n}\ \mathbb{E}\left(|N_p - N_{p-1}|^2/F_{p-1}\right) = N_0^2 + \sum_{p=0}^{n-1}\sum_{i=1}^{N_p}\ \mathbb{E}\left((M_p^i - N_p W_p^i)^2/F_p\right)$$

and therefore a uniform convergence result presented in section 2.2.2 will take place only if

$$\sup_{n\ge 0} E\left(A_n^2\right) = \sum_{p=1}^{\infty}\mathbb{E}\left(|N_p - N_{p-1}|^2\right) < \infty$$

The following simple example shows that even for the minimum variance Bernoulli branching law one cannot expects to obtain the analog of the uniform convergence results as those presented in section 2.2.2. Let us assume that the state space $E = \{0, 1\}$, the fitness functions $\{g_n\ ;\ n \ge 1\}$ and the transition kernels $\{K_n\ ;\ n \ge 1\}$ are time homogeneous and given by

$$g(1) = 3g(0) > 0 \qquad K(x, dz) = \nu(dz) \stackrel{\text{def}}{=} \frac{1}{2}\delta_0(dz) + \frac{1}{2}\delta_1(dz)$$

In this case one can check that for sufficiently large N_0

$$\mathbb{E}\left((\eta_n^N(1) - \eta_n(1))^2\right) \ge \frac{n}{5N_0} \xrightarrow[n \to \infty]{} \infty$$

In contrast to the situation described above in this simple case the IPS approximating model (13) will simply consist at each time in N_0 i.i.d. particles with common law ν and

$$\forall\, n \geq 0, \qquad \mathbb{E}\left((\eta_n^N(f) - \eta_n(f))^2\right) \;\leq\; \frac{1}{N_0}$$

for any bounded test function such that $\|f\| \leq 1$.

2.3.4 Regularized Mutations

The regularization of the mutation transition discussed in this section has been presented in [59]. Hereafter we briefly indicate why it is sometimes necessary to add a regularization step and how the previous analysis applies to this situation.

In nonlinear filtering settings the mutation transitions $\{K_n \;;\; n \geq 1\}$ are given by the problem at hand. More precisely they are the transitions of the un-known signal process. If $E = \mathbb{R}^d$, it may happen that some coordinates of the signal are deterministic. For instance if the signal is purely deterministic the IPS approximating scheme (13) does not work since after a finite number of steps we would get a single deterministic path.

As noticed in [41] the standard regularization technique used in practice consists in adding a "small noise" in the deterministic parts of the signal. When $E = \mathbb{R}^d$, $d \geq 1$, the introduction of such a "small noise" in the signal dynamics structure consists in replacing the transitions $\{K_n \;;\; n \geq 1\}$ by the regularized ones

$$K_n^{(\alpha)} = R^{(\alpha)} K_n \qquad \text{or} \qquad K_n^{(\alpha)} = K_n R^{(\alpha)}$$

where $R^{(\alpha)}$ is a new transition probability kernel on E and defined for any $f \in \mathcal{B}_{\rm b}(E)$ by

$$R^{(\alpha)}(f)(x) \stackrel{\text{def.}}{=} \int \frac{1}{\alpha^d}\, \theta\left(\frac{y-x}{\alpha}\right)\, f(y)\, dy$$

where $\theta : \mathbb{R}^d \to (0, \infty)$ is a Borel bounded function such that

$$\int \theta(y)\, dy = 1 \qquad \text{and} \qquad \sigma \stackrel{\text{def.}}{=} \int |y|\, \theta(y)\, dy < \infty$$

The regularized limiting measure valued system is now defined by

$$\eta_n^{(\alpha)} = \Phi_n^{(\alpha)}\left(\eta_{n-1}^{(\alpha)}\right) \qquad n \geq 1 \tag{89}$$

with $\eta_0^{(\alpha)} = \eta_0$ and where $\Phi_n^{(\alpha)} : \mathbf{M}_1(E) \to \mathbf{M}_1(E)$ is defined as in (8) by replacing the transitions $\{K_n \;;\; n \geq 1\}$ by $\{K_n^{(\alpha)} \;;\; n \geq 1\}$.

Let us denote by

$$\xi_n^{(\alpha)} = \left(\xi_n^{(\alpha),1}, \ldots, \xi_n^{(\alpha),N}\right)$$

the N-IPS approximating model associated with (89) and by $\eta_n^{(\alpha),N}$ the empirical measure associated with the system $\xi_n^{(\alpha)}$ and defined as usual by

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

It is transparent from the above construction that the convergence results presented in section 2.2.2 can be applied. For instance for any bounded test function $f \in \mathcal{B}_{\mathrm{b}}(E)$, $\|f\| \leq 1$, and $p \geq 1$ and $n \geq 0$

$$\mathbb{E}\left(\left|\eta_n^{(\alpha),N}(f) - \eta_n^{(\alpha)}(f)\right|^p\right)^{\frac{1}{p}} \leq \frac{B_p}{\sqrt{N}}\,(n+1)\; a_{0,n}^2$$

where B_p is a universal constant which only depends on the parameter p.
Of course we still have to check that the flow of distributions $\{\eta_n^{(\alpha)}\ ;\ n \geq 1\}$ is close to the desired flow $\{\eta_n\ ;\ n \geq 1\}$ as α is close to zero. To this end we introduce some additional notations.

We denote by Lip_1 the set of globally Lipschitz functions with Lipschitz norm less than 1 that is

$$|f(x) - f(y)| \leq |x - y|$$

and $\|f\| \leq 1$. We will also use the following assumption

$(\mathcal{R})$ ***For any time*** $n \geq 1$***, there exist some constants*** $C_n^1, C_n^2 < \infty$ ***such that***

$$g_n \in C_n^1 \cdot \mathrm{Lip}_1 \qquad \text{and} \qquad K_n(\mathrm{Lip}_1) \subset C_n^2 \bullet \mathrm{Lip}_1.$$

Lemma 2.37 *Under* $(\mathcal{R})$*, for any* $n \geq 1$ *there exists some constant* $C_n < \infty$ *such that for any* $f \in \mathrm{Lip}_1$

$$\left|\eta_n^{(\alpha)}(f) - \eta_n(f)\right| \leq C_n\; \alpha\;\; \sigma \tag{90}$$

Proof: For any $f \in \mathrm{Lip}_1$ and $x \in \mathbb{R}^d$ we clearly have

$$\left|R^{(\alpha)}(f)(x) - f(x)\right| \leq \int |f(x + \alpha\; y) - f(x)|\;\; \theta(y)\; dy \leq \alpha\;\; \sigma$$

and therefore

$$\sup_{f \in \mathrm{Lip}_1} \left\|R^{(\alpha)}(f) - f\right\| \leq \alpha\;\; \sigma$$

Under our assumptions this implies that for any $\eta \in \mathbf{M}_1(E)$

$$\sup_{f \in \mathrm{Lip}_1} \left|\Phi_n^{(\alpha)}(\eta)(f) - \Phi_n(\eta)(f)\right| \leq\; \alpha\;\; \sigma\; A_n$$

for some constant $A_n < \infty$. Let us prove (90) by induction on the parameter n. For $n = 0$ the result is trivial with $C_0 = 0$ so we assume that it holds for $(n-1)$. Using Lemma 2.2 we have the decomposition

$$\begin{aligned}
&\Phi_n(\eta_{n-1}^{(\alpha)})(f) - \Phi_n(\eta_{n-1})(f) \\
&= \frac{1}{\eta_{n-1}(g_n)} \\
&\quad \times \left[(\eta_{n-1}^{(\alpha)}(g_n K_n(f)) - \eta_{n-1}(g_n K_n(f))) + \Phi_n(\eta_{n-1}^{(\alpha)})(f)\ (\eta_{n-1}(g_n) - \eta_{n-1}^{(\alpha)}(g_n)) \right]
\end{aligned}$$

for any $f \in \mathrm{Lip}_1$. There is no loss of generality to assume that $C_n^{(1)} \geq a_n$ and that $C_n^{(2)} \geq 1$. Thus one gets

$$\begin{aligned}
&\Phi_n(\eta_{n-1}^{(\alpha)})(f) - \Phi_n(\eta_{n-1})(f) \\
&= \frac{C_n^{(1)} + a_n C_n^{(2)}}{\eta_{n-1}(g_n)} \left[(\eta_{n-1}^{(\alpha)}(f_1) - \eta_{n-1}(f_1)) \right] \\
&\qquad + \Phi_n(\eta_{n-1}^{(\alpha)})(f)\ \frac{C_n^{(1)}}{\eta_{n-1}(g_n)} \left[(\eta_{n-1}(f_2) - \eta_{n-1}^{(\alpha)}(f_2)) \right]
\end{aligned}$$

with

$$f_1 = \frac{1}{C_n^{(1)} + a_n C_n^{(2)}}\ g_n K_n(f) \qquad \text{and} \qquad f_2 = \frac{1}{C_n^{(1)}}\ g_n$$

so that $f_1, f_2 \in \mathrm{Lip}_1$. Using the induction hypothesis we arrive at

$$\sup_{f \in \mathrm{Lip}_1} \left\| \Phi_n(\eta_{n-1}^{(\alpha)})(f) - \Phi_n(\eta_{n-1})(f) \right\| \leq \frac{2(C_n^{(1)} + a_n C_n^{(2)})}{\eta_{n-1}(g_n)}\ C_{n-1}\ \sigma\ \alpha$$

Therefore if we combine the above results one gets finally

$$\sup_{f \in \mathrm{Lip}_1} |\eta_n^{(\alpha)}(f) - \eta_n(f)| \ \leq\ C_n\ \sigma\ \alpha$$

with

$$C_n = A_n + \frac{2(C_n^{(1)} + a_n C_n^{(2)})}{\eta_{n-1}(g_n)}\ C_{n-1}.$$

∎

A direct consequence of the above lemma is the following estimate

$$\mathbb{E}\left(\left| \eta_n^{(\alpha),N}(f) - \eta_n(f) \right|^p \right)^{\frac{1}{p}} \leq \frac{B_p}{\sqrt{N}}\ (n+1)\ a_{0,n}^2 + C_n\ \sigma\ \alpha$$

The approximation of $\{\eta_n\ ;\ n \geq 0\}$ by the regularized IPS is now guaranteed by the following proposition.

Proposition 2.38 *Assume that the condition $(\mathcal{R})$ is satisfied. Then for any $n \geq 0$ and $p \geq 1$ there exists some constant $C_{p,n} < \infty$ such that*

$$\alpha(N) = 1/\sqrt{N} \Longrightarrow \sup_{f \in \mathrm{Lip}_1} \mathbb{E}\left(\left| \eta_n^{(\alpha),N}(f) - \eta_n(f) \right|^p \right)^{\frac{1}{p}} \leq \frac{C_{p,n}}{\sqrt{N}}$$

3 The Continuous Time Case

We will try here to retranscribe some results obtained in previous parts for discrete time settings to continuous time models. Generally speaking, the same behaviors are expected, but the technicalities are more involved, even only for the definitions. Furthermore, this case has been less thoroughly investigated, and it seems that a genuine continuous time interacting particle system approximation has only recently been introduced in [45].

This last paper will be our reference for this part and we will keep working in the same spirit, but in the details our approach will be different in order to obtain numerous improvements and to prove new results: weak propagation of chaos valid for all bounded measurable functions and related upper bounds in terms of the supremum norm, as well as uniform convergence results with respect to time and central limit theorem and exponential upper bounds for the fluctuations.

Heuristically the main difference between the two time models is that for discrete time, in the selection step all the particles are allowed to change, whereas for continuous time, only one particle may change at a (random) selection time, but the length of the interval between two selection times is of order $1/N$ (N being as above the number of particles). So in some mean sense, in one unit time interval, "every particle should have had a good chance to change".
This is a first weak link between discrete and continuous time. But, even if this may not be clear at first reading, there are stronger relations between the formal treatment of the discrete and continuous times, and in our way to trying to understand the subject, they have influenced each other. In order to point out these connections, we have tried to use the same notations in both set-ups.

To finish this opening introduction, we must precise that the main results obtained here (the weak propagation of chaos and to some extend the central limit theorem) can also be deduced from the approach of Graham and Méléard [65], valid for a more general set-up, except they have put more restrictions on the state space, which is assumed to be $\mathbb{R}^d$, for some $d \geq 1$ (but perhaps this point is unessential).
Nevertheless, we have preferred to introduce another method, may be more immediate (e.g. without any reference to Boltzmann trees or Sobolev imbeddings ...), because we have taken into account that our models are simpler, since we are not in the need of considering the broader situation of [65].
More precisely, in our case, we have a nice a priori explicit expression (3) and (4) for the (deterministic) limiting objects, making them appear as a ratio of linear terms with respect to η_0, which is hidden in $\mathbb{E}$ as the initial distribution. This structure is more tractable than the information one would get by merely considering the nonlinear equation of evolution looking like (11) satisfied by the family $(\eta_t)_{t\geq 0}$. So we can make use of some associated nonnegative Feynman-Kac semigroups to obtain without difficulty the desired results.

3.1 Hypotheses on the Limiting Process

To accomplish the program described above it is first convenient to define more precisely the objects arising in formula (4). Our next objective is to introduce several kinds of assumptions needed in the sequel and to prove some preliminary results. As before, the metric space (E, r) is supposed to be Polish, and we denote by $\mathbf{B}(E)$ the σ-field of Borel subsets of E.

3.1.1 Definitions and Weak Regularity Assumption

Here we introduce the basic definitions and the weak hypothesis under which we will work. It is already largely weaker than the one considered in [45], and we believe that maybe it can even be removed (we have been making some recent progress in this direction, but at the expense of readability, considering tensorized empirical measures ...), but at least this weak assumption make clear the regularity problem one is to encounter when following the approach presented in this paper.

The simplest object to be explained is the family $(U_t)_{t\geq 0}$ of non-negative functions. We will assume that the mapping

$$U\,:\,\mathbb{R}_+\times E\ni(t,x)\;\;\mapsto\;\;U_t(x)\,\in\mathbb{R}_+$$

is $\mathbf{B}(\mathbb{R}_+)\otimes\mathbf{B}(E)$-measurable (where $\mathbf{B}(\mathbb{R}_+)$ is the usual σ-field of the Borel subsets of $\mathbb{R}_+$) and locally bounded in the sense that for all $T>0$, its restriction to $[0,T]\times E$ is bounded.

Next we need to define the E-valued time-inhomogeneous Markov process X arising in the right hand side of (4). In our settings the more efficient and convenient way seems to be in terms of a martingale problem (cf [55] for a general reference), since our method will be largely based on properties of martingales (as it was already true for discrete time, so the set-up we are now presenting is quite a natural generalization of the previous one):

For $t\geq 0$, let $D([t,+\infty[,E)$ be the set of all càdlàg paths from $[t,+\infty[$ to E. We denote by $(X_s)_{s\geq t}$ the process of canonical coordinates on $D([t,+\infty[,E)$, which generates on this space the σ-algebra $\mathcal{D}_{t,+\infty}=\sigma(X_s\,:\,s\geq t)$. We will also use as customary the notation $\mathcal{D}_{t,s}=\sigma(X_u\,:\,t\leq u\leq s)$, for $0\leq t\leq s$.

Let $\mathcal{A}$ be a dense sub-algebra of $\mathcal{C}_{\mathrm{b}}(E)$ which is supposed to contain $\mathbb{I}$ (that is the function taking everywhere the value 1).

A linear operator L_0 from the domain $\mathcal{A}$ to $\mathcal{C}_{\mathrm{b}}(E)$ will be called a pregenerator, if for all $x\in E$, there exists a probability $\mathbb{P}_x$ on $(D([0,+\infty[,E),\,\mathcal{D}_{0,+\infty})$ such that

- $X_0\circ\mathbb{P}_x=\delta_x$, the Dirac mass in x, and
- for all $\varphi\in\mathcal{A}$, the process

$$\left(\varphi(X_s)-\varphi(X_0)-\int_0^s L_0(\varphi)(X_u)\,du\right)_{s\geq 0}$$

is a $(\mathcal{D}_{0,s})_{s\geq 0}$-martingale under $\mathbb{P}_x$.

Let $(L_t)_{t\geq 0}$ be a measurable family of pregenerators: for each $t\geq 0$, $L_t\,:\,\mathcal{A}\to\mathcal{C}_{\mathrm{b}}(E)$ is a pregenerator, and for each $\varphi\in\mathcal{A}$ fixed,

$$\mathbb{R}_+\times E\ni(t,x)\;\;\mapsto\;\;L_t(\varphi)(x)$$

is $\mathcal{B}(\mathbb{R}_+)\otimes\mathcal{E}$-measurable. For the sake of simplicity, we will furthermore impose that the above function is locally bounded.

Our first hypothesis is

(H1) ***For all*** $(t,x)\in\mathbb{R}_+\times E$***, there exists a unique probability*** $\mathbb{P}_{t,x}$ ***on*** $(D([t,+\infty[,E),\,\mathcal{D}_{t,+\infty})$ ***such that***

- $X_t\circ\mathbb{P}_{t,x}=\delta_x$***, and***
- ***for all*** $\varphi\in\mathcal{A}$***, the process***

$$\left(\varphi(X_s)-\varphi(X_t)-\int_t^s L_u(\varphi)(X_u)\,du\right)_{s\geq t}$$

is a $(\mathcal{D}_{t,s})_{s\geq t}$***-martingale under*** $\mathbb{P}_{t,x}$.
Furthermore, it is assumed that for all $A\in\mathcal{D}_{0,+\infty}$***, the mapping***

$$\mathbb{R}_+\times E\ni(t,x)\ \mapsto\ \mathbb{P}_{t,x}(\theta_t(A))$$

is measurable (where $(\theta_t)_{t\geq 0}$ ***denotes the family of usual time shifts on*** $D([0,+\infty[,E)$***).***

Combining the measurability and the uniqueness assumption (H1), one can check that

$$\big((X_t)_{t\geq 0},\,(\mathbb{P}_{t,x})_{(t,x)\in\mathbb{R}_+\times E}\big) \tag{91}$$

is a strong Markov process (this uniqueness condition will only be needed here, so it can be removed if we rather suppose that the previous object is a Markov process). One can also define a probability $\mathbb{P}_{\eta_0}$ on $(D([0,+\infty[,E),\mathcal{D}_{0,+\infty})$, for any distribution $\eta_0\in M_1(E)$, by

$$\forall\ A\in\mathcal{D}_{0,+\infty},\qquad \mathbb{P}_{\eta_0}(A)\ =\ \int_E \mathbb{P}_{0,x}(A)\,\eta_0(dx)$$

which is easily seen to be the unique solution to the martingale problem associated to $(L_t)_{t\geq 0}$ whose initial law is η_0 (from now on, $\mathbb{E}_{\eta_0}$ will stand for the expectation relative to $\mathbb{P}_{\eta_0}$, the probability $\eta_0\in\mathbf{M}_1(E)$ being fixed).

The previous martingale problem can be extended to a time-space version as follows:

Let $\mathbf{A}$ be the set of absolutely continuous functions $g:\mathbb{R}_+\to\mathbb{R}$ with bounded derivative in the sense that there exists a bounded measurable function $g':\mathbb{R}_+\to\mathbb{R}$, such that for all $t\geq 0$,

$$g(t)\ =\ g(0)+\int_0^t g'(s)\,ds$$

On $\mathbf{A}\otimes\mathcal{A}$, we define the operator L given on functions of the form $f=g\otimes\varphi$, with $g\in\mathbf{A}$ and $\varphi\in\mathcal{A}$, by

$$\forall\ t\geq 0,\,\forall\ x\in E,\qquad L(f)(t,x)\ =\ g'(t)\varphi(x)+g(t)L_t(\varphi)(x) \tag{92}$$

Then we have the standard result:

Lemma 3.1 *Let $(t,x) \in \mathbb{R}_+ \times E$ be fixed. Under $\mathbb{P}_{t,x}$, for each $f \in \mathbf{A} \otimes \mathcal{A}$, the process $(M_s(f))_{s \geq t}$ defined by*

$$\forall\ s \geq t, \qquad M_s(f) \ = \ f(s, X_s) - f(t, X_t) - \int_t^s L(f)(u, X_u)\, du$$

is a square integrable martingale and its increasing process has the form

$$\forall\ s \geq t, \qquad \langle M(f) \rangle_s \ = \ \int_t^s \Gamma\,(f,f)\,(u, X_u)\, du$$

where Γ is the "carré du champ" bilinear operator associated to the pregenerator L and defined by

$$\forall\ f,g \in \mathbf{A} \otimes \mathcal{A}, \qquad \Gamma\,(f,g) \ = \ L(fg) - f\ L(g) - g\ L(f) \tag{93}$$

We can consider, for $s \geq 0$, the "carré du champ" bilinear operator Γ_s associated to the pregenerator L_s, which is naturally defined by

$$\forall\ \phi, \varphi \in \mathcal{A}, \qquad \Gamma_s(\phi, \varphi) \ = \ L_s(\,\phi\varphi\,) - \phi\, L_s(\varphi) - \varphi\, L_s(\phi)$$

We easily see that for all $f, g \in \mathbf{A} \otimes \mathcal{A}$,

$$\forall\ (s,x) \in \mathbb{R}_+ \times E, \qquad \Gamma(f,g)(s,x) \ = \ \Gamma_s\,(f(s,.), g(s,.))\,(x)$$

In Lemma 3.1, the fact that $\mathbf{A} \otimes \mathcal{A}$ is an algebra is crucial in order to describe the increasing process. But the domain $\mathbf{A} \otimes \mathcal{A}$ is rather too small for our purposes. We extend it in the following way: for $T > 0$ fixed, we denote by $\mathcal{B}_b([0,T] \times E)$ the set of all measurable bounded functions $f\,:\,[0,T] \times E \to \mathbb{R}$ and by $\mathcal{B}_T$ the vector space of applications $f \in \mathcal{B}_b([0,T] \times E)$ for which there exists a function $\bar{L}(f) \in \mathcal{B}_b([0,T] \times E)$ such that the process $(M_t(f))_{0 \leq t \leq T}$ defined by

$$\forall\ 0 \leq t \leq T, \qquad M_t(f) \ = \ f(t, X_t) - f(0, X_0) - \int_0^t \bar{L}(f)(s, X_s)\, ds$$

is a $\mathbb{P}_{\eta_0}$-martingale.

In this article and unless otherwise stated, all martingales will be implicitly assumed to be càdlàg (a.s.). Note that those coming from (H1) or from Lemma 3.1 are automatically càdlàg. Let us furthermore introduce $\mathcal{A}_T$ the subset of function $f \in \mathcal{B}_T$ for which there exists a non-negative mapping $\bar{\Gamma}(f,f) \in \mathcal{B}_{\rm b}([0,T] \times E)$ such that the increasing process associated with $(M_t(f))_{0 \leq t \leq T}$ has the form

$$\forall\ 0 \leq t \leq T, \qquad \langle M(f) \rangle_t \ = \ \int_0^t \bar{\Gamma}\,(f,f)\,(s, X_s)\, ds$$

Remark that $\bar{L}(f)$ and $\bar{\Gamma}(f,f)$ may not be uniquely defined by $f \in \mathcal{A}_T$ (but we will keep abusing of these notations), nevertheless this is not really important, since for martingale problems, one can consider multi-valued operators, cf [55].

Next we will need some regularity conditions on the function U and the family of probability measures $(\mathbb{P}_{t,x})_{(t,x)\in\mathbb{R}_+\times E}$. These are expressed in the following assumption (in the subsection 3.1.2, we will give separate and stronger hypotheses on U and $(\mathbb{P}_{t,x})_{(t,x)\in\mathbb{R}_+\times E}$ which insure that (H2) is satisfied):

(H2) ***For all*** $T > 0$ ***and*** $\varphi \in \mathcal{A}$ ***fixed, the application***

$$F_{T,\varphi} \,:\, [0,T]\times E \ni (t,x) \;\;\mapsto\;\; \mathbb{E}_{t,x}\left[\exp\left(\int_t^T U_s(X_s)\,ds\right)\varphi(X_T)\right] \tag{94}$$

belongs to $\mathcal{A}_T$.

We notice that $F_{T,\varphi}$ satisfies almost surely the first assumption hidden in (H2), that is $F_{T,\varphi}\in\mathcal{B}_T$. To see this claim let us denote for any $T>0$ and $\varphi\in\mathcal{A}$ fixed,

$$\forall\, 0\leq t\leq T, \qquad N_t(T,\varphi) \;=\; F_{T,\varphi}(t,X_t)$$

The Markov property of X implies that the process $(\widetilde{M}_t)_{0\leq t\leq T}$ defined by

$$\begin{aligned}\forall\, 0\leq t\leq T, \qquad \widetilde{M}_t \;&=\; \exp\left(\int_0^t U_s(X_s)\,ds\right) N_t(T,\varphi)\\ &=\; \mathbb{E}_{\eta_0}\left[\exp\left(\int_0^T U_s(X_s)\,ds\right)\varphi(X_T)\middle|\,\mathcal{D}_{0,t}\right]\end{aligned}$$

is a martingale. So we have

$$N_t(T,\varphi) \;=\; \exp\left(-\int_0^t U_s(X_s)\,ds\right)\widetilde{M}_t \tag{95}$$

$$\begin{aligned}\;=\; N_0(T,\varphi) + \int_0^t \exp\left(-\int_0^s U_u(X_u)\,du\right)\,d\widetilde{M}_s\\ -\int_0^t U_s(X_s)N_s(T,\varphi)\,ds\end{aligned} \tag{96}$$

from which it follows that

$$(\widehat{M}_t)_{0\leq t\leq T} \;\stackrel{\text{def.}}{=}\; \left(N_t(T,\varphi)-N_0(T,\varphi)+\int_0^t U_s(X_s)N_s(T,\varphi)\,ds\right)_{0\leq t\leq T}$$

is a martingale. Note that it is not necessarily càdlàg, and this could be annoying for some calculations afterwards (in (95) there was already a little difficulty, nevertheless it is possible to consider there a càdlàg modification, and end up with the fact that $(\widehat{M}_t)_{0\leq t\leq T}$ is a (contingently non-càdlàg) martingale). If (H2) is fulfilled, a classical uniqueness argument for semimartingale expansion will show that $(\widehat{M}_t)_{0\leq t\leq T}$ is almost surely càdlàg. Thus, we can take

$$\forall\, 0\leq t\leq T,\,\forall\, x\in E, \qquad \bar{L}(F_{T,\varphi})(t,x) \;=\; -U_t(x)F_{T,\varphi}(t,x)$$

In addition if the mapping $F_{T,\varphi}$ defined in (94) belongs to $\mathbf{A}_T\otimes\mathcal{A}$, where $\mathbf{A}_T$ is the set of restrictions to $[0,T]$ of functions from $\mathbf{A}$, then we conclude that one can choose

$$\bar{\Gamma}(F_{T,\varphi},F_{T,\varphi}) \;=\; \Gamma(F_{T,\varphi},F_{T,\varphi})$$

(where the action of Γ on $(\mathbf{A}_T \otimes \mathcal{A})^2$ is defined similarly as the one on $(\mathbf{A} \otimes \mathcal{A})^2$).

But our setting is too weak to insure this property, leading us to make the assumption (H2), which will serve as an ersatz.

One way to check this condition is to follow the procedure given below:

Lemma 3.2 *Let $f \in \mathcal{B}_T$ such that there exists a sequence $(f_n)_{n\geq 0}$ of elements of $\mathbf{A}_T \otimes \mathcal{A}$ satisfying, as n and m tend to infinity, $f_n \rightharpoonup f$ and*

$$\Gamma(f_n - f_m, f_n - f_m) \rightharpoonup 0,$$

where $\rightharpoonup$ stands for the bounded pointwise convergence on $[0,T] \times E$, and

$$\sup_{n\in\mathbb{N},\, 0\leq t\leq T} \|L(f_n)(t,\cdot)\| < +\infty$$

Then we have that $f \in \mathcal{A}_T$.

Proof: By dominated convergence, we obtain that

$$\lim_{n,m\to\infty} \mathbb{E}_{\eta_0}[\langle M(f_n - f_m)\rangle_T] \;=\; 0$$

in other words

$$\lim_{n,m\to\infty} \mathbb{E}_{\eta_0}[(M_T(f_n) - M_T(f_m))^2] \;=\; 0$$

It is now quite standard to deduce from this Cauchy convergence that there exists a martingale $(M_t)_{0\leq t\leq T}$ (that we can choose càdlàg) such that

$$\lim_{n\to\infty} \mathbb{E}_{\eta_0}[\sup_{0\leq t\leq T} (M_t(f_n) - M_t)^2] \;=\; 0$$

Since for any $0 \leq t \leq T$ $f_n(t, X_t) - f_n(0, X_0)$ converges in $\mathbb{L}^2(\mathbb{P}_{\eta_0})$ to $f(t, X_t) - f(0, X_0)$ as $n \to \infty$ one concludes that

$$\int_0^t L(f_n)(s, X_s)\, ds - \int_0^t \bar{L}(f)(s, X_s)\, ds$$

converges in $\mathbb{L}^2(\mathbb{P}_{\eta_0})$ to

$$M_t(f) - M_t \;=\; f(t, X_t) - f(0, X_0) - \int_0^t \bar{L}(f)(s, X_s)\, ds - M_t$$

This shows that the latter process (for $0 \leq t \leq T$) is predictable, as a limit of predictable processes, and we already know that it is also a martingale. Furthermore, from our hypotheses, there exists a finite constant $C_T > 0$, such that for all $n \in \mathbb{N}$ and all $0 \leq t_1 \leq t_2 \leq T$,

$$\left| \int_{t_1}^{t_2} L(f_n)(s, X_s)\, ds - \int_{t_1}^{t_2} \bar{L}(f)(s, X_s)\, ds \right| \;\leq\; C_T(t_2 - t_1)$$

so in the end (considering a subsequence), $(M_t(f) - M_t)_{t\geq 0}$ will also be of bounded variations. It is now well known that up to an evanescent set, it is the null process, that is $(M_t)_{0\leq t\leq T} = (M_t(f))_{0\leq t\leq T}$.

On the other hand since L_s is a pregenerator, for any $s \geq 0$, then Γ_s satisfies

$$\forall\, \varphi \in \mathcal{A}, \forall\, x \in E, \qquad \Gamma_s(\varphi,\varphi)(x) \;\geq\; 0$$

(see (98) below).

This implies that

$$\forall\, n,m \geq 0, \forall\, (s,x) \in [0,T] \times E,$$

$$\left|\sqrt{\Gamma(f_n,f_n)(s,x)} - \sqrt{\Gamma(f_m,f_m)(s,x)}\right| \;\leq\; \sqrt{\Gamma(f_n - f_m, f_n - f_m)(s,x)} \tag{97}$$

Therefore there exists a function $\bar{\Gamma}(f,f) : [0,T] \times E \to \mathbb{R}_+$ such that

$$\Gamma(f_n,f_n) \rightharpoonup \bar{\Gamma}(f,f)$$

as n tends to infinity.

Again using convergence theorems (but now in $\mathbb{L}^1(\mathbb{P}_{\eta_0})$) one obtains that

$$\left(M_t^2(f) - \int_0^t \bar{\Gamma}(f,f)(s,X_s)\,ds\right)_{0\leq t\leq T}$$

is a martingale. ∎

Remark 3.3:

(a) In the hypothesis $f \in \mathcal{B}_T$ of the previous lemma, we don't really need to assume that $(M_t(f))_{0\leq t\leq T}$ is càdlàg, since the equality $(M_t(f))_{0\leq t\leq T} = (M_t)_{0\leq t\leq T}$ insures this property.

(b) The result is also true if instead of assuming that $f_n \in \mathbf{A}_T \otimes \mathcal{A}$, we rather suppose that f_n, $f_n - f_m \in \mathcal{A}_T$, for $n,m \geq 0$, and that (97) is satisfied with Γ replaced by $\bar{\Gamma}$ (the other hypotheses remaining the same, for instance $\left\|\bar{L}(f_n)(t,\,\cdot\,)\right\|$ is uniformly bounded in $0 \leq t \leq T$ and $n \in \mathbb{N}$).

Example 3.4

Assume that E is $\mathbb{R}^n$ for some $n \in \mathbb{N}^*$ and $\mathcal{A}$ contains at least all C^2 functions with compact support and that the L_t, $t \geq 0$, are second order differential operators (without 0-order terms) with locally bounded coefficients (in time but uniformly in space).
Then Lemma 3.2 shows that $\mathcal{A}_T$ contains at least all $C_{\mathrm{b}}^{1,2}$ functions.

This approach is a little more flexible that the one using only stochastic calculus (to see that at least in case there is only continuous martingales (as in a Brownian filtration), $\mathcal{A}_T$ is stable by composition with C^2 functions, in the sense that if f_1, $f_2, \cdots, f_p$ belongs to $\mathcal{A}_T$ and if $F \in C^2(\mathbb{R}^p)$, then $F(f_1, f_2, \cdots, f_p)$ belongs to $\mathcal{A}_T$).
Using a classical embedding theorem the same results are also true for smooth separable manifolds.

Let us observe that if $f \in \mathcal{A}_T$, then the process given by

$$\left(f^2(t,X_t) - f^2(0,X_0) - \int_0^t \left[\bar{\Gamma}(f,f)(s,X_s) + 2f(s,X_s)\bar{L}(f)(s,X_s)\right]\, ds\right)_{0\leq t\leq T}$$

is also a martingale.

To see this claim we use the following decomposition (for $0 \leq t \leq T$)

$$\begin{aligned}
&\left(f(t,X_t) - f(0,X_0) - \int_0^t \bar{L}(f)(s,X_s)\,ds\right)^2 \\
&= \quad f^2(t,X_t) - f^2(0,X_0) + \left(\int_0^t \bar{L}(f)(s,X_s)\,ds\right)^2 \\
&\qquad -2f(t,X_t)\int_0^t \bar{L}(f)(s,X_s)\,ds + 2f(0,X_0)M_t(f) \\
&= \quad f^2(t,X_t) - f^2(0,X_0) - 2\int_0^t f(s,X_s)\bar{L}(f)(s,X_s)\,ds + 2f(0,X_0)M_t(f) \\
&\qquad -2\int_0^t \left(\int_0^s \bar{L}(f)(u,X_u)\,du\right)\, dM_s(f)
\end{aligned}$$

where the last equality comes from an integration by parts (this is also the proof of the second part of Lemma 3.1). In other words this means that if $f \in \mathcal{A}_T$, then $f^2 \in \mathcal{B}_T$, and we can take $\bar{L}(f^2) = \bar{\Gamma}(f,f) + 2f\bar{L}(f)$.

If we look at $(\mathcal{B}_T, \bar{L})$ as an (contingently multi-valued) linear operator, then $\mathbb{P}_{\eta_0}$ is also the solution to the time-space martingale problem associated with the initial condition η_0 and to $(\mathcal{B}_T, \bar{L})$.

Finally, let us mention other properties of a pregenerator L_0 which will be useful latter on. If Γ_0 and $(\mathbb{P}_x)_{x\in E}$ are respectively the "carré du champs" and solutions of the martingale problems associated with L_0, then we have for all $\varphi \in \mathcal{A}$ and $x \in E$,

$$\Gamma_0(\varphi,\varphi)(x) \quad = \quad \lim_{t\to 0_+} \frac{\mathbb{P}_x[(\varphi(X_t) - \varphi(x))^2]}{t} \tag{98}$$

It follows that if $\phi, \varphi \in \mathcal{A}$ and $f \in \mathbf{A}_T \otimes \mathcal{A}$, then we have that

$$\|\Gamma_0(\phi\varphi,\phi\varphi)\| \quad \leq \quad 2(\|\phi\|^2\, \|\Gamma_0(\varphi,\varphi)\| + \|\varphi\|^2\, \|\Gamma_0(\phi,\phi)\|)$$

and for all $x \in E$,

$$\Gamma_0\left(\int_0^T f(s,\,\cdot\,)\,ds, \int_0^T f(s,\,\cdot\,)\,ds\right)(x) \quad \leq \quad T\int_0^T \Gamma_0(f(s,\,\cdot\,), f(s,\,\cdot\,))(x)\,ds$$

3.1.2 Strong Regularity Assumptions

The hypotheses introduced below are not strictly necessary for the results presented in the following sections, but they are often the shortest way to check condition (H2), at least in abstract setting (ie not in the cases of example 3.4), so maybe this section should be skipped at a first reading. More precisely, they allow for an understanding of the latter hypothesis, by separating the roles of U and several aspects of regularity for the semigroup associated to X.

To make a clear differentiation, each of them has been associated to a hypothesis below. Thus the strong regularity assumption consists in the set of the following conditions $(\mathrm{H4})_T$, $(\mathrm{H7})_T$, $(\mathrm{H8})_T$, $(\mathrm{H9})_T$ and $(\mathrm{H10})_T$, the other ones are only intermediate steps. Furthermore the considerations presented here are more or less standard in the so-called semigroup approach to the theory of Markov processes, and they will enable us to give more tractable expressions for some norms arising in the study of exponential bounds presented in section 3.3.3.

The time inhomogeneous semigroup associated with the Markov process (91) is the family $(P_{s,t})_{0\leq s\leq t}$ of operators on $\mathcal{B}_{\mathrm{b}}(E)$ defined by

$$\forall\, 0\leq s\leq t,\, \forall\, \varphi\in\mathcal{B}_{\mathrm{b}}(E),\, \forall\, x\in E,\qquad P_{s,t}(\varphi)(x) \;=\; \mathbb{E}_{s,x}[\varphi(X_t)] \tag{99}$$

For $T>0$ which will be supposed fixed in this subsection, we consider the following set of strong hypotheses:

$(\mathbf{H3})_T$ ***There exists a constant*** $C_{1,T}\geq 1$ ***such that for all*** $0\leq t\leq T$***,***

$$C_{1,T}^{-1}\Gamma_0 \;\leq\; \Gamma_t \;\leq\; C_{1,T}\Gamma_0$$

in the sense that for all $\varphi\in\mathcal{A}$ ***and all*** $x\in E$***,***

$$C_{1,T}^{-1}\Gamma_0(\varphi,\varphi)(x)\leq\Gamma_t(\varphi,\varphi)(x)\leq C_{1,T}\Gamma_0(\varphi,\varphi)(x).$$

Let us define a norm $|\!|\!|\cdot|\!|\!|$ on $\mathcal{A}$ by

$$\forall\,\varphi\in\mathcal{A},\qquad |\!|\!|\varphi|\!|\!| \;=\; \sqrt{\|\varphi\|^2+\|\Gamma_0(\varphi,\varphi)\|}$$

$(\mathbf{H4})_T$ ***For all*** $0\leq t\leq T$***, we have*** $U_t\in\mathcal{A}$***, and the application***

$$[0,T]\ni t \;\mapsto\; U_t\in\mathcal{A}$$

is continuous with respect to $|\!|\!|\cdot|\!|\!|$***.***

$(\mathbf{H5})_T$ ***For all*** $0\leq s\leq t\leq T$***,*** $\mathcal{A}$ ***is stable by*** $P_{s,t}$***, and for*** $\varphi\in\mathcal{A}$ ***and*** $0\leq t\leq T$ ***fixed, the mappings***

$$\begin{aligned}[0,t]\ni s \;&\mapsto\; P_{s,t}(\varphi)\in\mathcal{A}\\ [t,T]\ni s \;&\mapsto\; P_{t,s}(\varphi)\in\mathcal{A}\end{aligned}$$

are continuous with respect to $|\!|\!|\cdot|\!|\!|$***.***

$(\mathbf{H6})_T$ ***For all*** $0\leq s\leq t\leq T$***,*** $\mathcal{A}$ ***is stable by*** $P_{s,t}$***, and there exists a constant*** $C_{2,T}\geq 1$ ***such that for all*** $0\leq s\leq t\leq T$ ***and all*** $\varphi\in\mathcal{A}$***,***

$$\|\Gamma_0(P_{s,t}(\varphi),P_{s,t}(\varphi))\| \;\leq\; C_{2,T}\,\|\Gamma_0(\varphi,\varphi)\|$$

(H7)$_T$ ***For all*** $0 \leq s \leq t \leq T$***,*** $\mathcal{A}$ ***is stable by*** $P_{s,t}$***, and for all*** $\varphi \in \mathcal{A}$***, there is a finite constant*** $C_{3,T}(\varphi) > 0$ ***such that for all*** $0 \leq s, t \leq T$***,***

$$\|L_s(P_{t,T}(\varphi))\| \leq C_{3,T}(\varphi)$$

(H8)$_T$ ***For all*** $0 \leq s \leq t \leq T$***,*** $\mathcal{A}$ ***is stable by*** $P_{s,t}$***, and we have in the*** $\vert\!\vert\!\vert \cdot \vert\!\vert\!\vert$ ***sense, on*** $\mathcal{A}$***,***

$$\left\{ \begin{array}{rcl} \dfrac{d}{ds} P_{s,t} & = & -L_s P_{s,t} \\ \dfrac{d}{dt} P_{s,t} & = & P_{s,t} L_t \end{array} \right.$$

(H9)$_T$ ***For all*** $\varphi \in \mathcal{A}$***,*** $[0,T] \ni t \mapsto \Gamma_t(\varphi,\varphi)$ ***is differentiable (in the sense of the norm*** $\|\cdot\|$***), and there exists a finite constant*** $C_{4,T} \geq 0$ ***such that for all*** $0 \leq t \leq T$ ***and all*** $\varphi \in \mathcal{A}$***,***

$$\left\| \frac{\partial_t \Gamma_t(\varphi,\varphi)}{\Gamma_t(\varphi,\varphi)} \right\| \leq C_{4,T}$$

(where ∂_t ***stands for*** d/dt***).***

If $\mathcal{A}$ is stable by L_t, for a given $t \geq 0$, we can define for $\phi, \varphi \in \mathcal{A}$,

$$\Gamma_{2,t}(\phi,\varphi) = \frac{1}{2}(L_t(\Gamma_t(\phi,\varphi)) - \Gamma_t(L_t(\phi),\varphi) - \Gamma_t(\phi, L_t(\varphi)))$$

We are now in position to introduce our last assumption:

(H10)$_T$ ***For all*** $0 \leq t \leq T$***,*** $\mathcal{A}$ ***is stable by*** L_t***, and there exists a constant*** $R_T \in \mathbb{R}$ ***such that for all*** $\varphi \in \mathcal{A}$ ***and*** $x \in E$***,***

$$\Gamma_{2,t}(\varphi,\varphi)(x) \geq R_T \Gamma_t(\varphi,\varphi)(x)$$

(R_T ***will then denote the best constant possible verifying this property, i.e. the largest one).***

Before studying several links between these hypotheses and (H2)$_T$ (which corresponds to (H2) for a $T > 0$ fixed), let us introduce $B(T,\mathcal{A})$ the set of $f \in \mathcal{B}_{\mathrm{b}}([0,T]\times E)$ such that for all $0 \leq t \leq T$, $f(t,\cdot) \in \mathcal{A}$ and such that

$$\vert\!\vert\!\vert f \vert\!\vert\!\vert_{[0,T]} \stackrel{\text{def.}}{=} \sup_{0 \leq t \leq T} \vert\!\vert\!\vert f(t,\cdot) \vert\!\vert\!\vert < +\infty$$

This quantity is a norm on $B(T,\mathcal{A})$, and we will note $\vert\!\vert\!\vert f \vert\!\vert\!\vert_t$ the semi-norm $\vert\!\vert\!\vert f(t,\cdot) \vert\!\vert\!\vert$, for any $0 \leq t \leq T$ and $f \in B(T,\mathcal{A})$.

In much the same way let $\|f\|_t = \|f(t,\cdot)\|$ and denote

$$\|f\|_{[0,T]} = \sup_{0 \leq t \leq T} \|f\|_t$$

for $f \in \mathcal{B}_{\mathrm{b}}([0,T] \times E)$.

Our first remark is:

Lemma 3.5 *Under (H3)$_T$, (H5)$_T$ and (H7)$_T$, for all $\varphi\in\mathcal{A}$, the mapping*

$$G_{T,\varphi}\,:\,[0,T]\times E\ni(t,x)\mapsto P_{t,T}(\varphi)(x)$$

belongs to $\mathcal{A}_T$ (the continuity in the first variable is only necessary in (H5)$_T$).

Proof: For each $n\in\mathbb{N}$ we introduce the functions $f_n\in\mathbf{A}_T\otimes\mathcal{A}$ defined by $\forall\ 0\leq t\leq T,\ \forall\ x\in E,$

$$f_n(t,x)\ =\ \left(1+k-\frac{(1+n)t}{T}\right)P_{\frac{kT}{1+n},T}(\varphi)(x)+\left(\frac{(1+n)t}{T}-k\right)P_{\frac{(1+k)T}{1+n},T}(\varphi)(x)$$

where $k=\lfloor(n+1)t/T\rfloor$. Clearly, under (H5)$_T$ we have that $G_{T,\varphi}\in B(T,\mathcal{A})$ and

$$\lim_{n\to\infty}|\!|\!|G_{T,\varphi}-f_n|\!|\!|_{[0,T]}\ =\ 0$$

By Lemma 3.2 and condition (H3)$_T$, to prove the announced result, it remains to show that

$$\sup_{n\in\mathbb{N}}\|L(f_n)\|_{[0,T]}\ <\ +\infty$$

To this end we first notice that for any $0\leq t\leq T$ and $x\in E$

$$\begin{aligned}L(f_n)(t,x)\ =\ &\frac{1+n}{T}\left(P_{\frac{(1+k)T}{1+n},T}(\varphi)(x)-P_{\frac{kT}{1+n},T}(\varphi)(x)\right)\\&+\left(1+k-\frac{(1+n)t}{T}\right)L_t[P_{\frac{kT}{1+n},T}(\varphi)](x)\\&+\left(\frac{(1+n)t}{T}-k\right)L_t[P_{\frac{(1+k)T}{1+n},T}(\varphi)](x)\end{aligned}$$

where k is defined as above. Therefore, in view of (H7)$_T$, to see the previous affirmation, we only need to check that

$$\sup_{n\in\mathbb{N},\,0\leq k\leq n}\frac{1+n}{T}\left\|P_{\frac{(1+k)T}{1+n},T}(\varphi)-P_{\frac{kT}{1+n},T}(\varphi)\right\|\ <\ +\infty$$

To see this claim it is sufficient to write, for all $x\in E$,

$$\begin{aligned}&P_{\frac{(1+k)T}{1+n},T}(\varphi)(x)-P_{\frac{kT}{1+n},T}(\varphi)(x)\\&=\mathbb{E}_{\frac{kT}{1+n},x}\left[P_{\frac{(1+k)T}{1+n},T}(\varphi)(x)-P_{\frac{(1+k)T}{1+n},T}(\varphi)(X_{\frac{(1+k)T}{1+n}})\right]\\&=-\mathbb{E}_{\frac{kT}{1+n},x}\left[\int_{\frac{kT}{1+n}}^{\frac{(k+1)T}{1+n}}L_s(P_{\frac{(1+k)T}{1+n},T}(\varphi))(X_s)\,ds\right]\end{aligned}$$

■

We consider next the collection $(Q_{s,t})_{0\leq s\leq t}$ of linear operator on $\mathcal{B}_{\rm b}(E)$ defined for any $0\leq s\leq t$, $\varphi\in\mathcal{B}_{\rm b}(E)$, $x\in E$, as follows

$$Q_{s,t}(\varphi)(x)\ =\ \mathbb{E}_{s,x}\left[\exp\left(\int_s^t U_u(X_u)\,du\right)\varphi(X_t)\right]$$

It is easily seen that $(Q_{s,t})_{0\leq s\leq t}$ is a well defined time inhomogeneous semigroup of non-negative operators on $\mathcal{B}_{\mathrm{b}}(E)$ (but in general non-Markovian, except in the trivial case where $U\equiv 0$, ie $(Q_{s,t})_{0\leq s\leq t}=(P_{s,t})_{0\leq s\leq t}$).

To see that Lemma 3.5 is also true for $(Q_{s,t})_{0\leq s\leq t}$, under the additional conditions (H4)$_T$ and (H6)$_T$, let us work out a relation between $(P_{s,t})_{0\leq s\leq t}$ and $(Q_{s,t})_{0\leq s\leq t}$:

Taking $t=T$ in (95) and integrating the above equality with respect to $\mathbb{P}_{0,x}$, we obtain (in fact for all $\varphi\in\mathcal{B}_{\mathrm{b}}(E)$),

$$Q_{0,T}(\varphi)(x) \;=\; P_{0,T}(\varphi)(x)+\int_0^T P_{0,s}(U_sQ_{s,T}(\varphi))(x)\,ds$$

More generally and in the same way one can prove that for any $0\leq t\leq T$,

$$Q_{t,T}(\varphi) \;=\; P_{t,T}(\varphi)+\int_t^T P_{t,s}(U_sQ_{s,T}(\varphi))\,ds$$

This identity leads us to consider, for $\varphi\in\mathcal{A}$ fixed, the application Z defined on $\mathcal{B}_{\mathrm{b}}([0,T]\times E)$ by

$\forall\ f\in\mathcal{B}_{\mathrm{b}}([0,T]\times E)$, $\forall\ 0\leq t\leq T$, $\forall\ x\in E$,

$$Z(f)(t,x) \;=\; P_{t,T}(\varphi)(x)+\int_t^T P_{t,s}(U_sf(s,\,\cdot\,))(x)\,ds$$

Under (H4)$_T$ there exists a constant $C_T^{(1)}>0$ such that

$\forall\ f_1,f_2\in\mathcal{B}_{\mathrm{b}}([0,T]\times E)$, $\forall\ 0\leq t\leq T$,

$$\|Z(f_1)-Z(f_2)\|_t \;\leq\; C_T^{(1)}\int_t^T \|f_1-f_2\|_s\,ds$$

Let $Z^n=Z\circ\ldots\circ Z$ (n-times) be the n-step iterate of Z. It is standard to check that

$$\forall\ f_1,f_2\in\mathcal{B}_{\mathrm{b}}([0,T]\times E),\qquad \|Z^n(f_1)-Z^n(f_2)\|_{[0,T]} \;\leq\; \frac{(C_T^{(1)}T)^n}{n\,!}\,\|f_1-f_2\|_{[0,T]}$$

This implies that $F_{T,\varphi}=Q_{\,\cdot\,,T}(\varphi)(\,\cdot\,)$ is the unique solution of the equation $Z(f)=f$, with $f\in\mathcal{B}_{\mathrm{b}}([0,T]\times E)$.

Another useful remark is that elementary calculations show that for any $f\in\mathcal{B}_{\mathrm{b}}([0,T]\times E)$

$$\left(Z(f)(t,X_t)-Z(f)(0,X_0)-\int_0^t P_{s,T}(U_sf(s,\,\cdot\,))(X_s)\,ds\right)_{0\leq t\leq T}$$

is a martingale (the càdlàg property comes from (H5)$_T$). This means that one can take

$$\forall\ 0\leq t\leq T,\ \forall\ x\in E,\qquad \bar{L}(Z(f))(t,x) \;=\; P_{t,T}(U_tf(t,\,\cdot\,))(x) \tag{100}$$

Proposition 3.6 *Under conditions (H3)$_T$, (H4)$_T$, (H5)$_T$, (H6)$_T$ and (H7)$_T$, assumption (H2)$_T$ holds, and there exists a finite constant $C_T^{(2)}>0$, such that for all $\varphi\in\mathcal{A}$,*

$$|\!|\!|\bar{F}_{T,\varphi}|\!|\!|_{[0,T]} \;\leq\; C_T^{(2)}|\!|\!|\varphi|\!|\!|$$

Proof: Let $\bar{B}(T,\mathcal{A})$ denote the Banach space which is the $\|\!|\cdot\|\!|_{[0,T]}$-completion of $\mathbf{A}_T\otimes\mathcal{A}$. Note that if $\bar{f}\in\bar{B}(T,\mathcal{A})$, then we can naturally associate to it numbers $\|\!|\bar{f}\|\!|_t$, for $0\le t\le T$, and as above $\|\!|\bar{f}\|\!| = \sup_{0\le t\le T}\|\!|\bar{f}\|\!|_t$. In the same way, as the norm $\|\!|\cdot\|\!|_{[0,T]}$ dominates $\|\cdot\|_{[0,T]}$, to each $\bar{f}\in\bar{B}(T,\mathcal{A})$ we associate a measurable bounded function $f\,:\,[0,T]\times E\to\mathbb{R}$ (we will occasionally abuse notation, saying that $f\in\bar{B}(T,\mathcal{A})$, and also note that we can associate to $\bar{f}$ another measurable bounded function which could be in an obvious way written $\Gamma_0(\bar{f},\bar{f})$). In view of the Lemma 3.2 and $(\mathrm{H3})_T$, if it wasn't for the boundedness condition on the $(L(f_n))_{n\in\mathbb{N}}$, such a function f would have a good chance to belong to $\mathcal{A}_T$. We will use this procedure here to show the belonging of $F_{T,\varphi}$ to $\mathcal{A}_T$, but rather taking into account the remark after Lemma 3.2 and (100).

From now on we consider the restriction of Z to $\mathbf{A}_T\otimes\mathcal{A}$. We slight abuse notations and still denote Z these restrictions. Using approximations techniques (as the one presented in the proof of Lemma 3.5), the remarks at the end of section 3.1.1 and the hypothesis $(\mathrm{H6})_T$, it is easily seen that $Z(f)\in\bar{B}(T,\mathcal{A})$, for $f\in\mathbf{A}_T\otimes\mathcal{A}$. More precisely, Z can be extended to $\bar{B}(T,\mathcal{A})$ and one can find a finite constant $C_T^{(3)}>0$ depending on $C_{2,T}$ and $\|\!|U\|\!|_{[0,T]}$ and such that for any $f_1,\ f_2\in\bar{B}(T,\mathcal{A})$ and $0\le t\le T$

$$\|\!|Z(f_1)-Z(f_2)\|\!|_t^2 \;\le\; C_T^{(3)}\int_t^T \|\!|f_1-f_2\|\!|_s^2\,ds$$

Then proceeding as above, it appears that for all $n\ge 1$,

$$\forall\ f_1,f_2\in\bar{B}(T,\mathcal{A}),\qquad \|\!|Z^n(f_1)-Z^n(f_2)\|\!|_{[0,T]}^2 \;\le\; \frac{(C_T^{(3)}T)^n}{n\,!}\|\!|f_1-f_2\|\!|_{[0,T]}^2$$

As a result there is a unique solution denoted by $\bar{F}_{T,\varphi}$ to the equation

$$Z(\bar{f})=\bar{f}$$

with $\bar{f}\in\bar{B}(T,\mathcal{A})$. Its corresponding bounded measurable function is clearly $F_{T,\varphi}$. Furthermore, $\bar{F}_{T,\varphi}$ is classically shown to be the limit in $\bar{B}(T,\mathcal{A})$ of $Z^n(0)$, for n large. By Lemma 3.2, to be convinced that $F_{T,\varphi}\in\mathcal{A}_T$, it remains to prove that

$$\sup_{n\in\mathbb{N}}\|L(Z^n(0))\|_{[0,T]} \;<\; +\infty$$

but the remark before the proposition shows that this is a consequence of

$$\sup_{n\in\mathbb{N}}\|Z^n(0)\|_{[0,T]} \;<\; +\infty$$

which itself follows from the considerations above the proposition.
The last part of this proposition is now clear from the previous approximations.

■

Remark 3.7: In practice it may be important to know the dependence of $C_T^{(2)}$ on U. Using the above proof we first observe that $C_T^{(3)}=C_T^{(4)}\|\!|U\|\!|_{[0,T]}^2$, where $C_T^{(4)}>0$ is a finite constant which does not depend on U.

Therefore for any $n \in \mathbb{N}^*$ and $f_1, f_2 \in \bar{B}(T, \mathcal{A})$ we have that

$$\begin{aligned}
\|\!|Z^n(f_1) - Z^n(f_2)|\!\|_{[0,T]} &\leq \sqrt{\frac{(C_T^{(4)}T)^n}{n\,!}}\,\|\!|U|\!\|_{[0,T]}^n \|\!|f_1 - f_2|\!\|_{[0,T]} \\
&\leq \frac{\sqrt{C(C_T^{(4)}T)^n}}{\lfloor n/2 \rfloor\,!}\,\|\!|U|\!\|_{[0,T]}^n \|\!|f_1 - f_2|\!\|_{[0,T]}
\end{aligned}$$

where we have use the Stirling formula to find a finite constant $C > 0$ such that for all $n \geq 1$,

$$\frac{1}{n\,!} \leq C \frac{1}{(\lfloor n/2 \rfloor\,!)^2}$$

Writing

$$\begin{aligned}
\|\!|\bar{F}_{T,\varphi}|\!\|_{[0,T]} &\leq \sum_{n \geq 0} \|\!|Z^{n+1}(0) - Z^n(0)|\!\|_{[0,T]} \\
&\leq \|\!|P_{\cdot,T}(\varphi)(\,\cdot\,)|\!\|_{[0,T]} \sum_{n \geq 0} \frac{\sqrt{C(C_T^{(4)}T)^n}}{\lfloor n/2 \rfloor\,!}\,\|\!|U|\!\|_{[0,T]}^n
\end{aligned}$$

makes it clear that there exist two finite constants $C_T^{(5)}$, $C_T^{(6)} > 0$ such that

$$C_T^{(2)} \leq C_T^{(5)} \exp(C_T^{(6)} \|\!|U|\!\|_{[0,T]})$$

In the end, we have

Proposition 3.8 *The conditions (H8)$_T$, (H9)$_T$ and (H10)$_T$ implies (H3)$_T$, (H5)$_T$ and (H6)$_T$*

For a proof of this result and more discussions on the curvature hypothesis (H10)$_T$, see [45].

Remarks 3.9:
a) We believe that the hypothesis (H7)$_T$ is not natural here, and we would have preferred to work only with closures related to functions and their carrés du champs. Nevertheless, note that if there is a finite constant $C_{5,T} > 0$ such that for all $\varphi \in \mathcal{A}$ and all $0 \leq s, t \leq T$,

$$\|L_t(\varphi) - L_s(\varphi)\| \leq C_{5,T} \|\!|\varphi|\!\|$$

(as it is the case in our applications to nonlinear filtering, cf. section 4.2), and if the right hand side of the first equation in (H8)$_T$ is $\|\cdot\|$-continuous, then (H7)$_T$ is satisfied.

b) In view of the generality of our setting, the reader may wondering why we have not only considered the time-homogeneous case, by adding the time as a coordinate of the Markov process. The corresponding generator on the state space $\mathbb{R}_+ \times E$ is L

acting on $\mathbf{A} \otimes \mathcal{A}$. But in general U does not belong to this domain, since typically in our applications to nonlinear filtering U will not be differentiable with respect to time (only a regularity of Hölder exponent less than $1/2$ can be expected). So the hypothesis $(\mathrm{H4})_T$ will not be verified in this context, giving one reason for which it seems interesting to us to separate the role of time.
Nevertheless, note that a way to get round this particular difficulty would be to complete $\mathcal{A}$ (or $\mathbf{A} \otimes \mathcal{A}$ in the homogeneous case) with respect to $|||\ |||$.

3.1.3 Asymptotic Stability

In this section we take up the study of the asymptotic stability of the limiting process $\eta = \{\eta_t\ ;\ t \geq 0\}$ which was begun in section 2.1.2 in discrete time settings. Before getting into the details we first need to make a few general observations and give some definitions. We retain notations of section 2.1.2 and we denote $\alpha(H)$ and $\beta(H)$ the contraction and Dobrushin ergodic coefficients associated with a given transition probability kernel H on E and given by (25) and (24).

Next we denote $\Phi = \{\Phi_{s,t}\ ;\ s \leq t\}$ the nonlinear semigroup in distribution space associated with the dynamics structure of η, namely

$$\forall\ s \leq t, \qquad \eta_t \;\; = \;\; \Phi_{s,t}(\eta_s) \tag{101}$$

One can check that each mapping $\Phi_{s,t}$, $s \leq t$, has the following handy form given for any $\eta \in \mathbf{M}_1(E)$ and for any bounded Borel function f on E by

$$\Phi_{s,t}(\eta)(f) = \frac{\eta(H_{s,t}(f))}{\eta(H_{s,t}(1))}$$

where

$$H_{s,t}(f)(x) \stackrel{\text{def.}}{=} \mathbb{E}_{s,x}\,(\, f(X_t)\; Z_{s,t}\,) \qquad \text{and} \qquad Z_{s,t} \stackrel{\text{def.}}{=} \exp\,(\int_s^t U_r(X_r)\, dr)$$

As in the discrete time case there is a simple trick which allows to connect the nonlinear semigroup Φ with a family of linear semigroups.

Lemma 3.10 *For any $s \leq t$ and $\mu \in \mathbf{M}_1(E)$ we have the following decomposition*

$$\Phi_{s,t}(\mu) = \Psi_{s,t}(\mu) H^{(t)}_{s,t}$$

where the mapping $\Psi_{s,t} : \mathbf{M}_1(E) \to \mathbf{M}_1(E)$ is defined by

$$\Psi_{s,t}(\mu)(f) = \frac{\mu(\ g_{s,t}\ f\ )}{\mu(\ g_{s,t}\ )} \qquad \text{where} \qquad g_{s,t} \stackrel{\text{def.}}{=} H_{s,t}1$$

and for any $t \geq 0$, $H^{(t)} = \{H^{(t)}_{s,r}\ ;\ s \leq r \leq t\}$ is a linear semigroup defined for any $f \in \mathcal{B}_{\mathrm{b}}(E)$ and $\mu \in \mathbf{M}_1(E)$ and $s \leq r \leq t$ by

$$\mu H^{(t)}_{s,r} f = \int_E \mu(dx)\ H^{(t)}_{s,r} f(x) \qquad \text{and} \qquad H^{(t)}_{s,r} f = \frac{H_{s,r}(g_{r,t}\ f)}{H_{s,r}(g_{r,t})}$$

Remark 3.11: By construction and using the Markov property of X it is easy to see that the transition kernels $H^{(t)}_{s,r}(x,dz)$, $s \leq r \leq t$ may likewise be defined for any bounded Borel function f by setting

$$\forall\, x \in E, \qquad H^{(t)}_{s,r}f(x) \;=\; \frac{\mathbb{E}_{s,x}\left(f(X_r)\,Z_{s,t}\right)}{\mathbb{E}_{s,x}\left(Z_{s,t}\right)}$$

As in discrete time settings the asymptotic stability properties of Φ can now be characterized in terms of the contraction coefficient of the linear semigroups $\{H^{(t)}\,;\, t \geq 0\}$.

Lemma 3.12 *For any $s \leq t$ we have that*

$$\beta\left(H^{(t)}_{s,t}\right) = \sup_{\mu,\nu} \|\Phi_{s,t}(\mu) - \Phi_{s,t}(\nu)\|_{\mathrm{tv}}$$

where the supremum is taken over all distributions $\mu,\nu \in \mathbf{M}_1(E)$.

On the basis of the definition of H it is now easy to establish that for any $f \in \mathcal{B}_{\mathrm{b}}(E)$, $x \in E$ and any $s \leq r \leq t$

$$H^{(t)}_{s,r}f(x) = \frac{\mathbb{E}_{s,x}\left(\,f(X_r)\,g_{r,t}(X_r)\,\mathbb{E}_{s,x}\left(Z_{s,r}\,|X_r\,\right)\right)}{\mathbb{E}_{s,x}\left(\,g_{r,t}(X_r)\,\mathbb{E}_{s,x}\left(Z_{s,r}\,|X_r\,\right)\right)}$$

Recalling the definition of the ergodic coefficient $\alpha(K)$ of a given transition probability kernel $K(x,dz)$ on E the above formula yields that for any $s \leq r \leq t$

$$\alpha\left(S^{(t)}_{s,r}\right)\;\exp-(\int_s^r \mathrm{osc}(U_\tau)\,d\tau) \leq \alpha\left(H^{(t)}_{s,r}\right) \leq \alpha\left(S^{(t)}_{s,r}\right)\;\exp(\int_s^r \mathrm{osc}(U_\tau)\,d\tau)$$

where

- For any $t \geq 0$, $\mathrm{osc}(U_t) \stackrel{\mathrm{def.}}{=} \sup\{U_t(y) - U_t(x)\,;\,(x,y) \in E^2\}$.
- For any $t \geq 0$, $S^{(t)} = \{S^{(t)}_{s,r}\,;\, s \leq r\}$ is a collection of transition probability functions on E and defined for any bounded Borel function f by

$$S^{(t)}_{s,r}f = \frac{P_{s,r}(g_{r,t}\;f)}{P_{s,r}(g_{r,t})}$$

Using the semigroup property of $H^{(t)}$, by definition of the ergodic and the contraction coefficients $\alpha(K)$ and $\beta(K)$ of a transition probability kernel K on E one proves the following result.

Proposition 3.13 *For any $v > 0$ and $s \leq r \leq t$ we have*

$$\beta\left(H^{(t)}_{s,r}\right) \leq \prod_{m \in I_v(s,r)} \left(1 - \alpha\left(S^{(t)}_{m,m+v}\right)\;\exp-(\int_m^{m+v} \mathrm{osc}(U_r)\,dr)\right)$$

where for any $v > 0$ and $s \leq r \leq t$, $I_v(s,r)$ is the subset of $\mathbb{R}_+$ defined by $I_v(s,r) \stackrel{\mathrm{def.}}{=} \{s + pv\,;\, 0 \leq p < [(r-s)/v]\}$ with $[a]$ the integer part of $a \in \mathbb{R}$.

If U is chosen so that

$$\mathrm{osc}^\star(U) \stackrel{\text{def.}}{=} \int_0^\infty \mathrm{osc}(U_t)\ dt \quad < \quad +\infty \tag{102}$$

then by definition of $S^{(t)}$ one gets the inequalities

$$\forall\ s \leq r \leq t, \qquad \alpha(P_{s,r})\ e^{-\mathrm{osc}^\star(U)} \ \leq\ \alpha(S^{(t)}_{s,r})\ \leq\ \alpha(P_{s,r})\ e^{\mathrm{osc}^\star(U)}$$

from which one concludes that

Theorem 3.14 *If the function U satisfies (102) then for any $v > 0$ the following implications hold*

$$\sum_{n\geq 1} \alpha\left(P_{(n-1)v,nv}\right) = \infty \implies \lim_{t\to\infty} \beta\left(H^{(t)}_{0,t}\right) = 0$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \alpha\left(P_{(k-1)v,kv}\right) \stackrel{\text{def.}}{=} \overline{\alpha}(v) \Longrightarrow \limsup_{t\to\infty} \frac{1}{t} \log\beta\left(H^{(t)}_{0,t}\right) \leq -\frac{\overline{\alpha}(v)}{v}\ e^{-2\mathrm{osc}^\star(U)}$$

In addition, if $\inf_{|t-s|=v} \alpha\left(P_{s,t}\right) \stackrel{\text{def.}}{=} \tilde{\alpha}(v)$ for some $v > 0$ then for any $p \geq 1$ and $T \geq p\,.\,v$ we have that

$$\sup_{t\geq 0} \sup_{\mu,\nu} \frac{1}{T} \log \|\Phi_{t,t+T}(\mu) - \Phi_{t,t+T}(\nu)\|_{\mathrm{tv}} \leq -\frac{\tilde{\alpha}(v)}{q\,.\,v}\ e^{-2\mathrm{osc}^\star(U)}$$

Next we examine an additional sufficient condition for the asymptotic stability of Φ in terms of the mixing properties of $P = \{P_{s,t}\ ;\ s \leq t\}$. Assume that the semigroup P satisfies the following condition.

$(\mathcal{P})$ ***There exists some $v > 0$ such that for any $t \geq 0$***

$$\forall\ x \in E, \qquad \epsilon_t^{1/2}(v) \ \leq\ \frac{dP_{t,t+v}(x,\bullet)}{d\mu_{t,v}} \ \leq\ \epsilon_t^{-1/2}(v)$$

for some positive constant $\epsilon_t(v) > 0$ and some reference probability measure $\mu_{t,v} \in \mathbf{M}_1(E)$.

The main simplification due to condition $(\mathcal{P})$ is the following: for any non-negative test function f we clearly have for any $0 \leq s \leq s + v \leq t$

$$\epsilon_s(v)\ \Psi_{s+v,t}(\mu)(f) \leq S^{(t)}_{s,s+v}(f) \leq \epsilon_s^{-1}(v)\ \Psi_{s+v,t}(\mu)(f)$$

This implies that $\alpha\left(S^{(t)}_{s,s+v}\right) \geq \epsilon_s(v)$ from which one can prove the following theorem.

Theorem 3.15 *Assume that the semigroup P satisfies condition $(\mathcal{P})$ for some constants $v > 0$, $\epsilon_t(v) > 0$ and some reference probability measure $\mu_{t,v} \in \mathbf{M}_1(E)$. If the function U is such that*

$$\|\mathrm{osc}(U)\| \stackrel{\text{def.}}{=} \sup_{t\geq 0} \mathrm{osc}(U_t) < \infty$$

then for any $v > 0$ the following implications hold

$$\sum_{n\geq 0} \epsilon_{nv}(v) = \infty \quad \Longrightarrow \quad \lim_{t\to\infty} \beta\left(H_{0,t}^{(t)}\right) = 0$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \epsilon_{kv}(v) \stackrel{\text{def.}}{=} \overline{\epsilon}(v) \quad \Longrightarrow \quad \limsup_{t\to\infty} \frac{1}{t} \log \beta\left(H_{0,t}^{(t)}\right) \leq -\frac{\overline{\epsilon}(v)}{v}\ e^{-v\,.\,\|\mathrm{osc}(U)\|}$$

In addition, if $\inf_{t\geq 0} \epsilon_t(v) \stackrel{\text{def.}}{=} \tilde{\epsilon}(v)$ *for some $v > 0$ then for any $p \geq 1$ and $T \geq p\,.\,v$ we have that*

$$\sup_{t\geq 0} \sup_{\mu,\nu} \frac{1}{T} \log \|\Phi_{t,t+T}(\mu) - \Phi_{t,t+T}(\nu)\|_{\mathrm{tv}} \leq -\frac{\tilde{\epsilon}(v)}{q\,.\,v}\ e^{-v\,.\,\|\mathrm{osc}(U)\|}$$

for any $p, q \geq 1$ such that $1/p + 1/q = 1$.

In time homogeneous settings several examples of semigroups P satisfying condition $(\mathcal{P})$ can be found in [9] and in [27]. For instance if X is a sufficiently regular diffusion on a compact manifold then $(\mathcal{P})$ holds with

$$\epsilon_t(v) = \epsilon(v) = A\ \exp -(B/v) \tag{103}$$

for some constants $0 < A, B < \infty$ and for the uniform Riemannian measure $\mu_{t,v} = \mu$ on the manifold. To illustrate our result let us examine the situation in which U is also time-homogeneous, that is $U_t = U$.

Corollary 3.16 *Assume that U is time-homogeneous and the semigroup of X satisfies condition $(\mathcal{P})$ with $\epsilon_t(v) = \epsilon(v)$ given by (103). Then for any $p \geq 1$ and $v > 0$ and $T \geq p\,.\,v$ we have that*

$$\sup_{\mu,\nu\in\mathbf{M}_1(E)} \|\Phi_{t,t+T}(\mu) - \Phi_{t,t+T}(\nu)\|_{\mathrm{tv}} \leq \exp -(\gamma\,.\,T)$$

with

$$\gamma \geq \frac{A}{q\ \ v}\ \exp -\left(\frac{B}{v} + v\,.\,\mathrm{osc}(U)\right) \qquad \text{and} \qquad \frac{1}{p} + \frac{1}{q} = 1$$

The best bound in term of the constants A, B and osc(U) is obtained for

$$v = v^\star \stackrel{\text{def.}}{=} \frac{2B}{1 + \sqrt{1 + 4B\,\mathrm{osc}(U)}}$$

The above asymptotic stability study can be extended in a simple way, but this is outside the scope of these notes, to study stability properties of the nonlinear filtering equation and its robust version. The interested reader is recommended to consult [43].

3.2 The Interacting Particle System Model

The purpose of this section is to design an IPS

$$(\xi_t)_{t\geq 0} = (\xi_t^1, \xi_t^2, \cdots, \xi_t^N)_{t\geq 0}$$

taking values in E^N, where $N \geq 1$ is the number of particles and such that for any $t \geq 0$, the empirical measures given by

$$\eta_t^N(\,\cdot\,) \;=\; \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi_t^i}(\,\cdot\,)$$

are good approximations of η_t, $t \geq 0$, for N large enough.

As announced in the introduction, the Markov process $(\xi_t)_{t\geq 0}$ will also be defined by a martingale problem. At this point it is convenient to give a more detailed description of its pregenerators which was already presented in (15). At time $t \geq 0$, the pregenerator $\mathcal{L}_t^{(N)}$ of the IPS will be of genetic type in the sense that it is defined as the sum of two pregenerators, namely

$$\mathcal{L}_t^{(N)} \;=\; \widetilde{\mathcal{L}}_t^{(N)} + \widehat{\mathcal{L}}_t^{(N)}$$

where

• The first pregenerator $\widetilde{\mathcal{L}}_t^{(N)}$ is called the mutation pregenerator, it denotes the pregenerator at time t coming from N-independent processes having the same evolution as X and it is given on $\mathcal{A}^{\otimes N}$ by

$$\forall\, \phi \in \mathcal{A}^{\otimes N}, \qquad \widetilde{\mathcal{L}}_t^{(N)}(\phi) \;=\; \sum_{i=1}^{N} L_t^{(i)}(\phi)$$

where $L_t^{(i)}$ denotes the action of L_t on the i-th variable x_i, that is

$$L_t^{(i)} = \mathrm{Id} \otimes \cdots \otimes \underbrace{L_t}_{i-\mathrm{th}} \otimes \cdots \otimes \mathrm{Id},$$

where Id is the identity operator.

• The second one is denoted by $\widehat{\mathcal{L}}_t^{(N)}$ and it is called the selection pregenerator. It is defined as the jump type generator defined for any $\phi \in \mathcal{A}^{\otimes N}$ and $x = (x_1, \ldots, x_N) \in E^N$ by

$$\widehat{\mathcal{L}}_t^{(N)}(\phi)(x) \;=\; \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} \left(\phi(x^{i,j}) - \phi(x)\right) U_t(x_j)$$

where for $1 \leq i, j \leq N$ and $x = (x_1, \cdots, x_N) \in E^N$, $x^{i,j}$ is the element of E^N given by

$$\forall\, 1 \leq k \leq N, \qquad x_k^{i,j} \;=\; \begin{cases} x_k & , \text{ if } k \neq i \\ x_j & , \text{ if } k = i \end{cases}$$

(this is meaningful for all functions $\phi \in \mathcal{B}_{\mathrm{b}}(E^N)$, and in fact $\widehat{\mathcal{L}}_t^{(N)}$ is a bounded generator on $\mathcal{B}_{\mathrm{b}}(E^N)$).

Heuristically the motion of the particles $(\xi_t)_{t\geq 0}$ is decomposed into the two following rules.
Between the jumps due to interaction between particles, each particle evolves independently from the others and randomly according to the time-inhomogeneous semigroup of X.
At some random times, say τ, we introduce a competitive interaction between the particles, during this stage a given particle ξ_τ^i will be replaced by a new particle ξ_τ^j, $1 \leq j \leq N$, with a probability proportional to its "adaptation" $U_\tau(\xi_\tau^j)$.

This mechanism is similar to the one of a Moran IPS, except in the form of the intensity of replacing ξ_τ^i by ξ_τ^j (which should be symmetric in these variables) and in the total jump rate which is here "proportional" to N (instead of N^2). In fact that renormalization shows that our Moran's type IPS can also be regarded as a Nanbu's type interacting particle approximating model for a simple generalized spatially homogeneous Boltzmann equation.

There is no real difficulty to construct a probability $\mathbb{P}$ (on $D([0,+\infty[,E^N))$ which is solution to the martingale problem associated with the initial distribution $\eta_0^{\otimes N}$ and to the time-inhomogeneous family of pregenerators $(\mathcal{L}_t^{(N)})_{t\geq 0}$. More precisely, we proceed in two steps: we first consider the product process on E^N corresponding to the family of pregenerators $(\widetilde{\mathcal{L}}_t^{(N)})_{t\geq 0}$ and to the initial distribution $\eta_0^{\otimes N}$. This is quite immediate and the N coordinates are independent and have the same law as X starting from η_0 (cf for instance the Theorem 10.1 p. 253 of [55]). Then, for all $t \geq 0$, $\mathcal{L}_t^{(N)}$ is just seen as a bounded perturbation of $\widetilde{\mathcal{L}}_t^{(N)}$ by $\widehat{\mathcal{L}}_t^{(N)}$. So we can apply general results about this kind of martingale problems, see the Proposition 10.2 p. 256 of [55].

From now on, $\mathbb{E}$ will designate the expectation relative to the process $\xi = \{\xi_t\ ;\ t \geq 0\}$ under $\mathbb{P}$.

For more information about how to construct and simulate this process, see section 3 of [45].

In fact, the law $\mathbb{P}$ of the particle system ξ satisfies a more extended martingale problem, since the proofs presented by Ethier and Kurtz [55] enable us to transpose the whole pregenerator $(\mathcal{B}_T, \bar{L})$ considered in the section 3.1.1:

Let us denote by $\mathcal{B}_{T,N}$ the vector sub-space of $\mathcal{B}_{\mathrm{b}}([0,T]\times E^N)$ generated by the functions $f\ :\ [0,T]\times E^N \to \mathbb{R}$ such that there exist $f_1,\cdots,f_N \in \mathcal{B}_T$ for which we have

$$\forall\ t\in[0,T],\ \forall\ x=(x_1,\cdots,x_N)\in E^N, \qquad f(t,x) \ = \prod_{1\leq i\leq N} f_i(t,x_i)$$

If such a function f is given, we define for all $(t,x)\in[0,T]\times E$,

$$\begin{aligned}&\widetilde{\mathcal{L}}^{(N)}(f)(t,x)\\ &= \sum_{1\leq i\leq N} f_1(t,x_1)\cdots f_{i-1}(t,x_{i-1})\bar{L}(f_i)(t,x_i)f_{i+1}(t,x_{i+1})\cdots f_N(t,x_N)\end{aligned}$$

and then we extend linearly this (contingently multi-valued) operator $\widetilde{\mathcal{L}}^{(N)}$ on $\mathcal{B}_{T,N}$.

We also need to consider the pregenerator $\widehat{\mathcal{L}}^{(N)}$ acting on $\mathcal{B}_{\mathrm{b}}([0,T]\times E^N)$ in the following way:

$\forall\ f\in\mathcal{B}_{\mathrm{b}}([0,T]\times E^N), \forall\ (t,x)\in[0,T]\times E,$

$$\widehat{\mathcal{L}}^{(N)}(f)(t,x) \quad=\quad \widehat{\mathcal{L}}_t^{(N)}(f(t,\cdot))(x)$$

and next we introduce on $\mathcal{B}_{T,N}$,

$$\mathcal{L}^{(N)} \quad=\quad \widetilde{\mathcal{L}}^{(N)}+\widehat{\mathcal{L}}^{(N)}$$

This pregenerator coincides naturally with $\partial_t+\mathcal{L}_t^{(N)}$ on $\mathbf{A}_T\otimes\mathcal{A}^{\otimes N}\subset\mathcal{B}_{T,N}$.

Lemma 3.17 *Under $\mathbb{P}$, for all $T>0$ and all $f\in\mathcal{B}_{T,N}$, the process $(M_t^{(N)}(f))_{0\leq t\leq T}$ defined by*

$$\forall\ 0\leq t\leq T,\qquad M_t^{(N)}(f) \quad=\quad f(t,\xi_t)-f(0,\xi_0)-\int_0^t \mathcal{L}^{(N)}(f)(s,\xi_s)\,ds$$

is a bounded martingale.

Now let $\mathcal{A}_{T,N}$ be the set of all $f\in\mathcal{B}_{T,N}$ for which $f^2\in\mathcal{B}_{T,N}$. With some obvious notations for such functions we can define

$$\begin{aligned}\Gamma^{(N)}(f,f) &\stackrel{\text{def.}}{=} \mathcal{L}^{(N)}(f^2)-2f\mathcal{L}^{(N)}(f)\\ &= \widetilde{\Gamma}^{(N)}(f,f)+\widehat{\Gamma}^{(N)}(f,f)\end{aligned}$$

Then an easy calculation we have already met several times shows that

Lemma 3.18 *For any $f\in\mathcal{A}_{T,N}$, the increasing process associated to the martingale $(M_t^{(N)}(f))_{0\leq t\leq T}$ is given by the formula*

$$\forall\ 0\leq t\leq T,\qquad \langle M^{(N)}(f)\rangle_t \quad=\quad \int_0^t \Gamma^{(N)}(f,f)(s,\xi_s)\,ds$$

Next we will apply this result to some special functions, for which $x\in E^N$ is seen only through its empirical measure $m^{(N)}(x)=\frac{1}{N}\sum_{1\leq i\leq N}\delta_{x_i}$, and more precisely to mappings f of the following type

$$\forall\ (t,x)\in[0,T]\times E^N,\qquad f(t,x)=m^{(N)}(x)(\widetilde{f}(t,\cdot))$$

where $\widetilde{f}\in\mathcal{A}_T$, since clearly such a function f belongs to $\mathcal{A}_{T,N}$.

3.3 Asymptotic Behavior

An important role will be played here by the unnormalized Feynman-Kac stochastic flow defined in (18). In all this section, the finite horizon $T>0$ and the initial condition η_0 are fixed.

3.3.1 Weak Propagation of Chaos

Our objective is to prove the following result, which somehow gives a justification for the interacting particle system just introduced:

Theorem 3.19 *Under the assumption (H2)$_T$, there exists a finite constant $C_T > 0$, such that for all $\varphi \in \mathcal{B}_b(E)$ and all $0 \leq t \leq T$,*

$$\mathbb{E}[|\eta_t^N(\varphi) - \eta_t(\varphi)|] \;\leq\; C_T \frac{\|\varphi\|}{\sqrt{N}}$$

The basic idea behind the proof of this result is quite simple: it consists in finding a martingale indexed by the interval $[0, T]$ whose terminal value at time T is precisely $\gamma_T^N(\varphi)$, for any given $\varphi \in \mathcal{A}$. The reader may be wondering why this quantity and not directly $\eta_T^N(\varphi)$. The reason for this choice is that for the unnormalized Feynman-Kac stochastic flow, we can take advantage of the underlying linear structure of the limiting dynamical system $(\gamma_t)_{t\geq 0}$ (ie of the relation $\gamma_t(\,\cdot\,) = \gamma_s(Q_{s,t}(\,\cdot\,))$ valid for all $t \geq s \geq 0$). As we will see, there is then a straight way to find such a martingale, via the martingale problem satisfied by the law of the particle system, but naturally we have to use the semigroup $(Q_{s,t})_{0\leq s\leq t}$ considered above. All the calculations would be immediate if one can use the heuristic formula

$$\partial_s Q_{s,t} \;=\; -L_s Q_{s,t} - U_s Q_{s,t}$$

(in the usual regular cases, this is satisfied, and maybe that at a first reading one should concentrate on these situations ...). To treat the little difficulties associated to the general case, we consider this equation in the sense of the martingale problem. This point of view led us to introduce the weak regularity condition (H2), insuring that some functions constructed through the semigroup $(Q_{s,t})_{0\leq s\leq t}$ are in the domain of an extended pregenerator.

Another important aspect of the martingales we will consider is that they are quite small, in the sense that their increasing process will be of order $1/N$. Then taking into account some results about iid random variables in order to estimate the initial approximation at time 0, the expected convergence will follow easily for the unnormalized Feynman-Kac formulae (and there will be no problem with the renormalization, so we will also end up with the convergence of the empirical measure toward the normalized Feynman-Kac formulae).

Furthermore, this approach gives at once the central limit theorem and exponential bounds for the fluctuations.

So from now on, we will look for nice martingales, through calculations of the action of the pregenerators on convenient functions, and (H2)$_T$ will be assumed fulfilled.

Lemma 3.20 *For any $\varphi \in \mathcal{A}$, the process*

$$\begin{aligned} &\big(B_t^N(\varphi)\big)_{0\leq t\leq T} \\ &\stackrel{\text{def.}}{=} \left(\eta_t^N(Q_{t,T}(\varphi)) - \eta_0^N(Q_{0,T}(\varphi)) + \int_0^t \eta_s^N(U_s)\eta_s^N(Q_{s,T}(\varphi))\, ds\right)_{0\leq t\leq T} \end{aligned}$$

is a martingale and its increasing process is given by

$$\langle B^N(\varphi)\rangle_t \;=\; \frac{1}{N}\int_0^t G(s,T,\eta_s^N,\varphi)\,ds$$

where for any $0\leq s\leq T$, $m\in\mathbf{M}_1(E)$ *and* $\varphi\in\mathcal{A}$,

$$G(s,T,m,\varphi)$$

$$= m\left[\bar{\Gamma}(F_{T,\varphi},F_{T,\varphi})(t,\,\cdot\,)\right] + m\left[(Q_{s,T}(\varphi)-m[Q_{s,T}(\varphi)])^2(U_s+m[U_s])\right]$$

Proof: Applying Lemma 3.17 and Lemma 3.18 to the function

$$f\,:\,[0,T]\times E^N\ni(t,x)\mapsto m^{(N)}(x)(Q_{t,T}(\varphi)) \tag{104}$$

it appears easily that

$$\begin{aligned}\widetilde{\mathcal{L}}^{(N)}(f)(t,x) &= -m^{(N)}(x)[U_tQ_{t,T}(\varphi)]\\ \widehat{\mathcal{L}}^{(N)}(f)(t,x) &= m^{(N)}(x)[U_tQ_{t,T}(\varphi)]-m^{(N)}(x)[U_t]m^{(N)}(x)[Q_{t,T}(\varphi)]\end{aligned}$$

and the proof of the first assertion is now complete.
For the second one, calculations are a little more tedious, but in the end, we get

$$N\widetilde{\Gamma}^{(N)}(f)(t,x)=m^{(N)}(x)[\bar{\Gamma}(F_{T,\varphi},F_{T,\varphi})(t,\,\cdot\,)]$$

and

$$N\widehat{\Gamma}^{(N)}(f)(t,x)=m^{(N)}(x)\left[(Q_{s,T}(\varphi)-m^{(N)}(x)[Q_{s,T}(\varphi)])^2(U_s+m^{(N)}(x)[U_s])\right]$$

■

In view of the above lemma to find a good upper bound of $N\mathbb{E}[\langle B^N(\varphi)\rangle_T]$ it is tempting to introduce the norm on $\mathcal{A}$ defined for any $\varphi\in\mathcal{A}$ by

$$|||\varphi|||_{[0,T]} \;=\; \sqrt{\int_0^T \|F_{T,\varphi}(t,\,\cdot\,)\|^2+\left\|\bar{\Gamma}(F_{T,\varphi},F_{T,\varphi})(t,\,\cdot\,)\right\|\,dt} \tag{105}$$

(more rigorously, the infimum over all possible choices of $\bar{\Gamma}(F_{T,\varphi},F_{T,\varphi})$ of this quantities). Using the considerations of section 3.1.2 we see that this norm is clearly related to $|||\cdot|||$

Next result shows that the above choice is in fact far from being optimal.

Lemma 3.21 *There exists a finite constant* $\bar{C}_T>0$, *such that for all* $\varphi\in\mathcal{A}$,

$$\mathbb{E}[\langle B^N(\varphi)\rangle_T] \;\leq\; \bar{C}_T\frac{\|\varphi\|^2}{N}$$

Proof: Let us write that for the function f defined in (104),
$\forall\,(t,x)\in[0,T]\times E^N$,

$$\begin{aligned}N\widetilde{\Gamma}^{(N)}(f,f)(t,x) &= m^{(N)}(x)[\bar{L}(Q^2_{\cdot,T}(\varphi))(t,\,\cdot\,)-2Q_{t,T}(\varphi)(\,\cdot\,)\bar{L}(Q_{\,\cdot\,,T}(\varphi))(t,\,\cdot\,)]\\ &= m^{(N)}(x)[\bar{L}(Q^2_{\cdot,T}(\varphi))(t,\,\cdot\,)]+2m^{(N)}(x)[U_tQ^2_{t,T}(\varphi)]\end{aligned}$$

If we define

$$g\,:\,[0,T]\times E^N\ni(t,x)\mapsto m^{(N)}(x)[Q^2_{t,T}(\varphi)]$$

then

$$\widetilde{\mathcal{L}}^{(N)}(g)(t,x) \;=\; m^{(N)}(x)[\bar{L}(Q^2_{\cdot,T}(\varphi))(t,\cdot)]$$

therefore

$$\begin{aligned}N\widetilde{\Gamma}^{(N)}(f,f)(t,x) &= \widetilde{\mathcal{L}}^{(N)}(g)(t,x)+2m^{(N)}(x)[U_tQ^2_{t,T}(\varphi)]\\ &= \mathcal{L}^{(N)}(g)(t,x)-\widehat{\mathcal{L}}^{(N)}(g)(t,x)+2m^{(N)}(x)[U_tQ^2_{t,T}(\varphi)]\end{aligned}$$

and

$$\begin{aligned}&N\widetilde{\Gamma}^{(N)}(f,f)(t,x)\\ &=\mathcal{L}^{(N)}(g)(t,x)+m^{(N)}(x)[U_tQ^2_{t,T}(\varphi)]+m^{(N)}(x)[U_t]m^{(N)}(x)[Q^2_{t,T}(\varphi)]\end{aligned}$$

and finally one gets

$$\begin{aligned}&N\Gamma^{(N)}(f,f)(t,x)\\ &\quad=\;\mathcal{L}^{(N)}(g)(t,x)+2m^{(N)}(x)[U_tQ^2_{t,T}(\varphi)]\\ &\qquad-2m^{(N)}(x)[U_tQ_{t,T}(\varphi)]m^{(N)}(x)[Q_{t,T}(\varphi)]+2m^{(N)}(x)[U_t]m^{(N)}(x)[Q^2_{t,T}(\varphi)]\end{aligned}$$

This implies that

$$\begin{aligned}&N\mathbb{E}[\langle B^N(\varphi)\rangle_T]\\ &\quad=\;\mathbb{E}\left[\int_0^T\Gamma^{(N)}(f,f)(t,\xi_t)\,dt\right]\\ &\quad=\;\mathbb{E}\Big[\eta^N_T[Q^2_{T,T}(\varphi)]-\eta^N_0[Q^2_{0,T}(\varphi)]+\\ &\qquad\int_0^T 2\eta^N_t[U_tQ^2_{t,T}(\varphi)]-2\eta^N_t[U_tQ_{t,T}(\varphi)]\eta^N_t[Q_{t,T}(\varphi)]+2\eta^N_t[U_t]\eta^N_t[Q^2_{t,T}(\varphi)]\,dt\Big]\end{aligned}$$

and the desired upper bound is now clear.

■

As announced, it is preferable to first obtain a result similar to Theorem 3.19 for γ^N.

Proposition 3.22 *There exists a finite constant $\widetilde{C}_T>0$, such that for any $\varphi\in\mathcal{A}$ and $0\le t\le T$,*

$$\mathbb{E}[\left|\gamma^N_t(\varphi)-\gamma_t(\varphi)\right|]\;\le\;\widetilde{C}_T\frac{\|\varphi\|}{\sqrt{N}}\tag{106}$$

Proof: By construction, for any $\varphi\in\mathcal{A}$, we have that

$$\gamma^N_t(\varphi)\;=\;\exp\left(\int_0^t\eta^N_s(U_s)\,ds\right)\eta^N_t(\varphi)$$

By Lemma 3.20 it follows that for $0 \leq t \leq T$,

$$\gamma_t^N(Q_{t,T}(\varphi)) = \gamma_0^N(Q_{0,T}(\varphi)) + \widetilde{B}_t^N(\varphi) \tag{107}$$

where

$$\widetilde{B}_t^N(\varphi) = \int_0^t \exp\left(\int_0^s \eta_u^N(U_u)\,du\right)\,dB_s^N(\varphi)$$

is a martingale. Its increasing process is clearly given for any $0 \leq t \leq T$ by

$$\begin{aligned}
\langle \widetilde{B}^N(\varphi)\rangle_t &= \int_0^t \exp\left(2\int_0^s \eta_u^N(U_u)\,du\right)\,d\langle B^N(\varphi)\rangle_s \\
&= \frac{1}{N}\int_0^t \exp\left(2\int_0^s \eta_u^N(U_u)\,du\right) G(s,T,\eta_s^N, Q_{s,T}(\varphi))\,ds \\
&\leq \frac{\widehat{C}_T}{N}\langle B(\varphi)\rangle_T
\end{aligned}$$

for a finite constant $\widehat{C}_T > 0$ which does not depend on φ (but it depends on U, through $\|U\|_{[0,T]}$).
Recalling that $\gamma_T(\varphi) = \gamma_0(Q_{0,T}(\varphi))$ one concludes that

$$\gamma_T^N(\varphi) - \gamma_T(\varphi) = \gamma_0^N(Q_{0,T}(\varphi)) - \gamma_0(Q_{0,T}(\varphi)) + \widetilde{B}_T^N(\varphi) \tag{108}$$

from which one gets that

$$\mathbb{E}[(\gamma_T^N(\varphi) - \gamma_T(\varphi))^2] = \mathbb{E}[(\gamma_0^N(Q_{0,T}(\varphi)) - \gamma_0(Q_{0,T}(\varphi)))^2] + \mathbb{E}[(\widetilde{B}_T^N)^2(\varphi)]$$

Now the result follows, via a Cauchy-Schwarz inequality, from

$$\mathbb{E}[(\widetilde{B}_T^N)^2(\varphi)] = \mathbb{E}[\langle \widetilde{B}^N(\varphi)\rangle_T]$$

and from the classical equality for iid variables:

$$\begin{aligned}
\mathbb{E}[(\gamma_0^N(Q_{0,T}(\varphi)) - \gamma_0(Q_{0,T}(\varphi)))^2] &= \mathbb{E}[(\eta_0^N(Q_{0,T}(\varphi)) - \eta_0(Q_{0,T}(\varphi)))^2] \\
&= \frac{\eta_0[(Q_{0,T}(\varphi) - \eta_0(Q_{0,T}(\varphi)))^2]}{N} \\
&\leq \exp(2T\,\|U\|_{[0,T]})\frac{\|\varphi\|^2}{N} \qquad \blacksquare
\end{aligned}$$

Proof of Theorem 3.19: First let us prove that the upper bound of the previous proposition is true for any $\varphi \in \mathcal{B}_{\mathrm{b}}(E)$. By density of $\mathcal{A}$ in $\mathcal{C}_{\mathrm{b}}(E)$, that is at least clear for all $\varphi \in \mathcal{C}_{\mathrm{b}}(E)$. Next let $\varphi \in \mathcal{B}_{\mathrm{b}}(E)$ be fixed. We have

$$\mathbb{E}[\left|\gamma_t^N(\varphi) - \gamma_t(\varphi)\right|] = \sup_{H\in\mathbb{L}^\infty(\mathbb{P}),\|H\|_{\mathbb{L}^\infty(\mathbb{P})}\leq 1} \mathbb{E}[H(\gamma_t^N(\varphi) - \gamma_t(\varphi))]$$

Let $H \in \mathbb{L}^\infty(\mathbb{P})$ be given with $\|H\|_{\mathbb{L}^\infty(\mathbb{P})} \leq 1$. Then the following formula

$$\forall\, \widetilde{\varphi} \in \mathcal{B}_{\mathrm{b}}(E), \qquad m_H(\widetilde{\varphi}) \stackrel{\text{def.}}{=} \mathbb{E}[H(\gamma_t^N(\widetilde{\varphi}) - \gamma_t(\widetilde{\varphi}))]$$

defines a measure $m_H \in \mathbf{M}(E)$.
But a classical approximation result says that there exists a sequence $(\varphi_n)_{n\in\mathbf{N}}$ of elements of $\mathcal{C}_{\mathrm{b}}(E)$, with $\|\varphi_n\| \leq \|\varphi\|$ for all $n \in \mathbb{N}$, such that

$$\lim_{n\to\infty} m_H(\varphi_n) \;=\; m_H(\varphi)$$

Putting together these facts, (106) is satisfied for all $\varphi \in \mathcal{B}_{\mathrm{b}}(E)$.
Now it is enough to write that for all $\varphi \in \mathcal{A}$,

$$\eta_T^N(\varphi) - \eta_t(\varphi) = \frac{1}{\gamma_T(\mathbb{1})}\left[\gamma_T^N(\varphi) - \gamma_T(\varphi) + \eta_T^N(\varphi)(\gamma_T(\mathbb{1}) - \gamma_T^N(\mathbb{1}))\right] \tag{109}$$

■

Theorem 3.19 can be improved. As a first step in this direction, let us come back to the martingale $(B_t^N(\varphi))_{0\leq t\leq T}$, for a given $\varphi \in \mathcal{A}$, and study its jumps:

Lemma 3.23 *There exists a finite constant $C_T^{(7)} > 0$ such that $\mathbb{P}$-a.s., we have*

$$\sup_{0\leq t\leq T}\left|\triangle B_t^N(\varphi)\right| \;\leq\; C_T^{(7)}\frac{\|\varphi\|}{N}$$

Proof: From the definition of $(B_t^N(\varphi))_{0\leq t\leq T}$, we see that this upper bound would be clear if we could show that $\mathbb{P}$-a.s., the coordinates of the $\mathbb{R}^N$-valued process

$$\Big(\big(Q_{t,T}(\varphi)(\xi_t^i)\big)_{1\leq i\leq N}\Big)_{0\leq t\leq T}$$

never jump together.
From the construction of $\mathbb{P}$ (cf [45] for more details), it is sufficient to prove this property for the process $\widetilde{\xi}$ which is just the product of N independent copies of X (starting from the distribution η_0). So it is in fact enough to convince oneself that the jumps of the martingale $(N_t(T,\varphi))_{0\leq t\leq T} = (F_{T,\varphi}(t,X_t))_{0\leq t\leq T}$ (under $\mathbb{P}_{\eta_0}$) are totally inaccessible, but this is a consequence of the continuity of its increasing process.

■

A similar statement holds for $(\widetilde{B}_t^N(\varphi))_{0\leq t\leq T}$.

Lemma 3.24 *For any $p > 0$, there exists a finite constant $C_{T,p}^{(8)} > 0$, which does not depend on φ and such that*

$$\sup_{0\leq t\leq T}\mathbb{E}[\left|\widetilde{B}_t^N(\varphi)\right|^p]^{1/p} \;\leq\; C_{T,p}^{(8)}\frac{\|\varphi\|}{\sqrt{N}}$$

Proof: By Hölder inequality, we only need to prove this result for $p = 2q$, with $q \in \mathbb{N}^*$, and we will proceed by an induction on q. We have already shown the case $q = 1$, so let $q \in \mathbb{N}\setminus\{0,1\}$ be given. Applying the Itô's formula for the mapping $\mathbb{R} \ni x \mapsto x^{2q}$, we get (writing $\widetilde{B} \stackrel{\text{def.}}{=} \widetilde{B}^N(\varphi)$ for simplicity) for $0 \leq t \leq T$,

$$\begin{aligned}\widetilde{B}_t^{2q} = 2q\int_0^t \widetilde{B}_s^{2q-1}\,d\widetilde{B}_s + q(2q-1)\int_0^t \widetilde{B}_s^{2q-2}\,d\langle\widetilde{B}^c\rangle_s \\ + \sum_{0\leq s\leq t}\widetilde{B}_s^{2q} - \widetilde{B}_{s-}^{2q} - 2q\widetilde{B}_{s-}^{2q-1}\triangle\widetilde{B}_s\end{aligned}$$

where $\widetilde{B}^c$ is the continuous martingale part of $\widetilde{B}$.
But quite obviously, there exists a finite constant $C_q > 0$ depending only on $q \geq 2$ and such that for any $0 \leq s \leq T$,

$$\widetilde{B}_s^{2q} - \widetilde{B}_{s-}^{2q} - 2q\widetilde{B}_{s-}^{2q-1}\triangle\widetilde{B}_s \;\leq\; C_q(\widetilde{B}_s^{2q-2} + (\triangle\widetilde{B}_s)^{2q-2})\triangle B_s^2$$

Using the previous lemma and the general fact that

$$\left(\langle\widetilde{B}\rangle_t - \langle\widetilde{B}^c\rangle_t - \sum_{0\leq s\leq t} \triangle B_s^2\right)_{0\leq t\leq T}$$

is a martingale, one can find a finite constant $C_{T,p} > 0$ such that

$$\mathbb{E}[(\widetilde{B}_t)^{2q}] \;\leq\; C_{T,p}\mathbb{E}\left[\int_0^t \left(\widetilde{B}_s^{2q-2} + \left(\frac{\|\varphi\|}{N}\right)^{2q-2}\right) d\langle\widetilde{B}\rangle_s\right]$$

Now the desired result follows from the bounds on $\langle\widetilde{B}\rangle$ we have already met in the proof of Proposition 3.22 and from the induction hypothesis (after an appropriate application of Fatou's Lemma).

∎

A consequence of these preliminary results is that

Proposition 3.25 *For all $p \geq 1$, there exists a finite constant $C_{p,T} > 0$ such that for all $\varphi \in \mathcal{B}_{\mathrm{b}}(E)$ and $0 \leq t \leq T$*

$$\mathbb{E}[\left|\eta_t^N(\varphi) - \eta_t(\varphi)\right|^p]^{1/p} \;\leq\; C_{p,T}\frac{\|\varphi\|}{\sqrt{N}}$$

Proof: Using the same arguments as in the proofs of Proposition 3.22 and Theorem 3.19, this inequality is a consequence of previous lemma and standard Marcinkiewicz-Zygmund's inequality, ensuring that there is a finite constant $C_p > 0$ such that

$$\mathbb{E}[\left|\eta_0^N(Q_{0,T}(\varphi)) - \eta_0(Q_{0,T}(\varphi))\right|^p]^{1/p} \;\leq\; \frac{C_p\,\|Q_{0,T}(\varphi)\|}{\sqrt{N}}$$

∎

Using for instance the previous proposition with $p = 4$ (note that for this case, we could have rather take into account the Lemma 3.34 p. 382 of [68], instead of Lemma 3.24), we can apply the Borel-Cantelli Lemma to see that almost surely, for a $\varphi \in \mathcal{B}_{\mathrm{b}}(E)$ fixed,

$$\lim_{N\to\infty} \eta_t^N(\varphi) \;=\; \eta_t(\varphi)$$

In fact the Proposition 3.25 can be quantitatively improved for $p = 1$, if we put the absolute value outside the expectation:

Proposition 3.26 *For $T > 0$ given, there exists a constant $\bar{C}_T \geq 0$ such that for all $\varphi \in \mathcal{B}_{\mathrm{b}}(E)$ and all $N \geq 1$, we have*

$$\left|\mathbb{E}[\eta_T^N(\varphi)] - \eta_T(\varphi)\right| \;\leq\; \bar{C}_T\frac{\|\varphi\|}{N}$$

Proof: Using the equation (108) valid for all $\varphi \in \mathcal{A}$, we get

$$\mathbb{E}[\gamma_T^N(\varphi)] \;=\; \gamma_T(\varphi)$$

But clearly this equality is then true for all $\varphi \in \mathcal{B}_{\mathrm{b}}(E)$, via usual arguments (ie γ_T^N is an estimator without biais of γ_T).

Then taking into account (109), we see that

$$\begin{aligned}
\left|\mathbb{E}[\eta_T^N(\varphi)] - \eta_T(\varphi)\right| &= \left|\mathbb{E}\left[\frac{1}{\gamma_T(\mathbb{1})}\eta_T^N(\varphi)(\gamma_T^N(\mathbb{1}) - \gamma_T(\mathbb{1}))\right]\right| \\
&= \left|\mathbb{E}\left[\frac{1}{\gamma_T(\mathbb{1})}(\eta_T^N(\varphi) - \eta_T(\varphi))(\gamma_T^N(\mathbb{1}) - \gamma_T(\mathbb{1}))\right]\right| \\
&\leq \frac{1}{\gamma_T(\mathbb{1})}\sqrt{\mathbb{E}[(\eta_T^N(\varphi) - \eta_T(\varphi))^2]}\sqrt{\mathbb{E}[(\gamma_T^N(\mathbb{1}) - \gamma_T(\mathbb{1}))^2]} \\
&\leq \frac{\bar{C}_T}{N}
\end{aligned}$$

for a constant $\bar{C}_T > 0$, according to the Proposition 3.25 and its proof.

■

The proportionality to $1/N$ of the upper bound we have just obtained is related to the strong propagation of chaos. This question will not be thoroughly investigated here, but let us give some immediate remarks about it.

First we recall the definition of this property: let $\bar{X} = (\bar{X}_t)_{t\geq 0}$ be the time inhomogeneous Markovian process taking values in E, whose initial law is η_0 and whose family of pregenerators is $(\mathcal{L}_{t,\eta_t})_{t\geq 0}$, where $(\eta_t)_{t\geq 0}$ is defined by (3) and (4) (starting from the same η_0) and where the pregenerator $\mathcal{L}_{t,\eta}$, for $t \geq 0$ and $\eta \in \mathbf{M}_1(E)$, is given by (12).

In the generalized Boltzmann equations literature, $\bar{X}$ is called the nonlinear process (or sometimes the tarjet process) because at any time $t \geq 0$, the law of $\bar{X}_t$ is η_t, but at this instant its pregenerator also uses in its definition the probability η_t. So the evolution depends on the time marginal and this is the nonlinear aspect of $\bar{X}$.

For $T > 0$ fixed, let $\bar{\mathbb{P}}_{\eta_0,[0,T]}$ denote the law of $(\bar{X}_t)_{0\leq t\leq T}$. Then the strong propagation of chaos can be expressed as the existence of a constant $C_T^{(9)} \geq 0$ such that for all $N \geq 1$ and $1 \leq k \leq N$, we are assured of

$$\left\|\mathbb{P}_{\eta_0,[0,T]}^{(N,1,\cdots,k)} - \bar{\mathbb{P}}_{\eta_0,[0,T]}^{\otimes k}\right\|_{\mathrm{tv}} \;\leq\; \frac{C_T^{(9)}k^2}{N}$$

where $\|\cdot\|_{\mathrm{tv}}$ stands for the total variation norm and where $\mathbb{P}_{\eta_0,[0,T]}^{(N,1,\cdots,k)}$ denotes the law over the time interval $[0,T]$ of the k^{th} first particles $(\xi_t^1, \xi_t^2, \cdots, \xi_t^k)_{0\leq t\leq T}$ (cf Graham and Méléard [65]).

Using on one hand a generalization of the approach presented here, but for the tensorized empirical measures alluded to in the beginning of section 3.1.1, and on the other hand two straightforward coupling arguments, it is possible to prove such a behavior, except we have not yet been able to get the right dependence in k. Nevertheless, note that as a direct consequence of Proposition 3.26, we have the simpler bound

$$\left\|\mathbb{P}^{(N,1)}_{\eta_0,T} - \bar{\mathbb{P}}_{\eta_0,T}\right\|_{\mathrm{tv}} \;\leq\; \frac{\bar{C}_T}{N}$$

where $\mathbb{P}^{(N,1)}_{\eta_0,T}$ (resp. $\bar{\mathbb{P}}_{\eta_0,T}$) denote the law of ξ^1_T (resp. $\bar{X}_T$). This comes from the identity $\bar{\mathbb{P}}_{\eta_0,T} = \eta_T$ and the fact that the distribution of ξ^1_T is also $\mathbb{E}[\eta^N_T]$, by exchangeability of the particles.

Furthermore, if we assume that the semigroup Φ is exponentially asymptotically stable, in the sense (111) given below, then we obtain an uniform in time result: for some constants $\bar{C} \geq 0$ and $0 < \alpha \leq 1$,

$$\sup_{t\geq 0}\left\|\mathbb{P}^{(N,1)}_{\eta_0,t} - \bar{\mathbb{P}}_{\eta_0,t}\right\|_{\mathrm{tv}} \;\leq\; \frac{\bar{C}}{N^\alpha}$$

More generally, to get uniform upper bounds with respect to the time parameter (under additional stability assumptions) it is important to control the dependence of the constant $C_{p,T}$ arising in Proposition 3.25. A more cautious study shows that in fact there exists a universal constant $A_p > 0$ depending on the parameter p and an additional finite constant $B > 0$ (which do not depend on p) such that

$$C_{p,T} \leq A_p \, \exp\left(B\int_0^T (1 + \|U_s\|)\, ds\right) \tag{110}$$

We end this section with a uniform convergence result with respect to the time parameter.

Theorem 3.27 *Assume that the semigroup Φ associated with the dynamics structure of η and defined in (101) is asymptotically stable in the sense that*

$$\lim_{T\to\infty}\sup_{\mu,\nu\in\mathbf{M}_1(E)}\sup_{t\geq 0}\|\Phi_{t,t+T}(\mu) - \Phi_{t,t+T}(\nu)\|_{\mathrm{tv}} = 0$$

If the function U satisfies $\|U\|^\star \stackrel{\text{def.}}{=} \sup_{t\geq 0}\|U_t\| < \infty$ then for any bounded Borel function φ we have the following uniform convergence result

$$\lim_{N\to\infty}\sup_{t\geq 0}\mathbb{E}\left(|\eta^N_t\varphi - \eta_t\varphi|\right) = 0$$

In addition, assume that the semigroup Φ is exponentially asymptotically stable in the sense that there exist some positive constant $\gamma > 0$ and $T_0 \geq 0$ such that for any $\mu,\nu \in \mathbf{M}_1(E)$ and $T \geq T_0$

$$\sup_{t\geq 0}\|\Phi_{t,t+T}(\mu) - \Phi_{t,t+T}(\nu)\|_{\mathrm{tv}} \;\leq\; e^{-\gamma T} \tag{111}$$

Then for any $p \geq 1$ and for any Borel function φ, $\|\varphi\| \leq 1$, we have the following uniform $\mathbb{L}^p$ error bound

$$\sup_{t\geq 0} \mathbb{E}\left(|\eta_t^N(\varphi) - \eta_t(\varphi)|^p\right)^{\frac{1}{p}} \leq \frac{A'_p\, e^{\gamma'}}{N^{\frac{\alpha}{2}}}$$

for any $N \geq 1$ such that

$$T(N) \stackrel{\text{def}}{=} \frac{1}{2}\frac{\log N}{\gamma + \gamma'} \geq T_0$$

where A'_p is a universal constant which only depends on $p \geq 1$ and α and γ' are given by

$$\alpha = \frac{\gamma}{\gamma + \gamma'} \quad \textit{and} \quad \gamma' = B(1 + \|U\|^\star)$$

and B is the finite constant arising in (110).

Proof: To prove this theorem we follow the same line of arguments as in the proof of Theorem 2.11. Since most of the computations are similar to those made in discrete time settings the proof will be only sketched.
The only point we have to check is that for any φ, $\|\varphi\| \leq 1$, $p \geq 1$, $T \geq T_0$ and $t \geq 0$

$$\mathbb{E}\left(\left|\eta_{t+T}^N(\varphi) - \Phi_{t,t+T}(\eta_t^N)(\varphi)\right|^p\right)^{\frac{1}{p}} \leq A_p \frac{\exp(\gamma' T)}{\sqrt{N}} \quad \text{with} \quad \gamma' = B(1 + \|U\|^\star)$$

Subtracting the equalities (107) at times $t = T$ and $t = s$ (and taking into account the definition of $\widetilde{B}$ given below (107)), we get that for any $s \leq T$ and $\varphi \in \mathcal{A}$,

$$\gamma_T^N(\varphi) = \gamma_s^N(Q_{s,T}(\varphi)) + \int_s^T \exp\left(\int_0^\tau \eta_u^N(U_u)\, du\right) dB_\tau^N(\varphi)$$

Since by definition we have

$$\gamma_s^N(1) = \exp\left(\int_0^s \eta_u^N(U_u)\, du\right) \eta_s^N(1) = \exp\left(\int_0^s \eta_u^N(U_u)\, du\right)$$

it appears by construction that

$$\frac{\gamma_s^N(Q_{s,T}(\varphi))}{\gamma_s^N(1)} = \eta_s^N(Q_{s,T}(\varphi))$$

so the above decomposition yields that

$$\eta_{s,T}^N(\varphi) \stackrel{\text{def.}}{=} \frac{\gamma_T^N(\varphi)}{\gamma_s^N(1)} = \eta_s^N(Q_{s,T}(\varphi)) + \int_s^T \exp\left(\int_s^\tau \eta_u^N(U_u)\, du\right) dB_\tau^N$$

By the same reasoning as in Lemma 3.24 one can check that for any $p \geq 1$, there exists a universal constant $A_p > 0$ depending on the parameter p and an additional finite constant $B > 0$ (which do not depend on p) such that

$$\mathbb{E}\left(\left|\eta_{s,T}^N(\varphi) - \eta_s^N(Q_{s,T}(\varphi))\right|^p\right)^{\frac{1}{p}} \leq A_p \frac{e^{B(1+\|U\|^\star)(T-s)}}{\sqrt{N}}$$

Instead of (109) we now use the decomposition

$$\eta_T^N(\varphi) - \Phi_{s,T}(\eta_s^N)(\varphi)$$

$$= \frac{1}{\eta_s^N(Q_{s,T}(1))} \left((\eta_{s,T}^N(\varphi) - \eta_s^N(Q_{s,T}(\varphi))) + \eta_T^N(\varphi) \left(\eta_s^N(Q_{s,T}(1)) - \eta_{s,T}^N(1)\right) \right)$$

to prove that

$$\mathbb{E}\left(\left|\eta_T^N(\varphi) - \Phi_{s,T}(\eta_s^N)(\varphi)\right|^p\right)^{\frac{1}{p}} \leq A_p \, \frac{e^{B(1+\|U\|^\star)\,(T-s)}}{\sqrt{N}}$$

The desired upper bound is now clear by replacing the pair parameter (s,T) by $(t, t+T)$. ■

3.3.2 Central Limit Theorem

We now turn to fluctuactions associated with the weak propagation of chaos. We proceed in much the same way as in discrete time settings. Let us introduce the "normalized" semigroup $(\bar{Q}_{s,t})_{0\leq s\leq t}$ defined by

$$\forall\, 0 \leq s \leq t,\, \forall\, \varphi \in \mathcal{A}, \qquad \bar{Q}_{s,t}(\varphi) = \frac{Q_{s,t}(\varphi)}{\eta_s(Q_{s,t}(\mathbb{1}))}$$

To see that it is in fact a semigroup, we first notice that

$$\begin{aligned} \forall\, 0 \leq s \leq t \leq T, \qquad \eta_s(Q_{s,t}(\mathbb{1})) &= \mathbb{E}_{s,\eta_s}\left[\exp\left(\int_s^t U_u(X_u)\, du\right)\right] \\ &= \exp\left(\int_s^t \eta_u(U_u)\, du\right) \end{aligned}$$

where the latter equality correspond to the basic identity

$$\gamma_t(1) = \exp(\int_0^t \eta_s(U_s)\, ds)$$

proved at the end of section 1.3, but for the shifted dynamical system $(\eta_{s+t})_{t\geq 0}$ instead of $(\eta_t)_{t\geq 0}$.
Consequently the above semigroup can be rewritten as follows

$\forall\, 0 \leq s \leq t \leq T,\, \forall\, x \in E,\, \forall\, \varphi \in \mathcal{A},$

$$\bar{Q}_{s,t}(\varphi)(x) \;=\; \mathbb{E}_{s,x}\left[\varphi(X_t)\exp\left(\int_s^t U_u(X_u) - \eta_u(U_u)\, du\right)\right]$$

Therefore it is clear that $(\bar{Q}_{s,t})_{0\leq s\leq t\leq T}$ is defined as $(Q_{s,t})_{0\leq s\leq t\leq T}$, by replacing the mapping U by the mapping

$$\widetilde{U} \,:\, \mathbb{R}_+ \times E \ni (t,x) \mapsto U_t(x) - \eta_t(U_t)$$

Theorem 3.28 *For $\varphi \in \mathcal{A}$, define*

$$W_T^N(\varphi) \;=\; \sqrt{N}(\eta_T^N(\varphi) - \eta_T(\varphi))$$

and let us denote as in (61), for $0 \le t \le T$,

$$\varphi_{t,T} \;=\; \bar{Q}_{t,T}(\varphi - \eta_T(\varphi))$$

Then under (H2)$_T$, the family $(W_T^N(\varphi))_{\varphi\in\mathcal{A}}$ converges in law to a centered Gaussian field $(W_T(\varphi))_{\varphi\in\mathcal{A}}$ whose covariances are given by

$$\forall\ \phi,\varphi \in \mathcal{A}, \qquad \mathbb{E}[W_T(\phi)W_T(\varphi)] \;=\; \eta_0[\phi_{0,T}\varphi_{0,T}] + \int_0^T \bar{G}(s,T,\eta_s,\phi,\varphi)\,ds$$

where for all $0 \le s \le T$, all $m \in \mathbf{M}_1(E)$ and all $\phi,\varphi \in \mathcal{A}$,

$$\begin{aligned}&\bar{G}(s,T,m,\phi,\varphi)\\ &= m\left[\bar{\Gamma}(\phi_{\cdot,T},\varphi_{\cdot,T})(s,\cdot)\right] + m\left[(\phi_{s,T} - m[\phi_{s,T}])(\varphi_{s,T} - m[\varphi_{s,T}])(U_s + m[U_s])\right]\end{aligned}$$

First, we will only consider one function $\varphi \in \mathcal{A}$, for which we have the analogous result of Lemma 3.20, whose proof is quite identical:

Lemma 3.29 *For $\varphi \in \mathcal{A}$, the process*

$$\left(\bar{B}_t^N(\varphi)\right)_{0\le t\le T} \;\stackrel{\text{def.}}{=}\; \left(\eta_t^N(\varphi_{t,T}) + \int_0^t (\eta_s^N(U_s) - \eta_s(U_s))\eta_s^N(\varphi_{s,T})\,ds\right)_{0\le t\le T}$$

is a martingale, whose initial value is

$$\bar{B}_0^N(\varphi) \;=\; \eta_0^N(\varphi_{0,T})$$

and whose increasing process is given by

$$\langle \bar{B}^N(\varphi)\rangle_t \;=\; \int_0^t \bar{G}(s,T,\eta_s^N,\varphi,\varphi)\,ds$$

We will examine separately each term arising in the above lemma. For the first one, we have

Lemma 3.30 *Under (H2)$_T$, the random variables*

$$\sqrt{N}\int_0^T \left|(\eta_s^N(U_s) - \eta_s(U_s))\eta_s^N(\varphi_{s,T})\right|\,ds$$

converge in probability to 0 as N tends to infinity.

Proof: It is enough to show that

$$\lim_{N\to\infty} \mathbb{E}\left[\sqrt{N}\int_0^T \left|(\eta_s^N(U_s) - \eta_s(U_s))\eta_s^N(\varphi_{s,T})\right|\,ds\right] \;=\; 0 \tag{112}$$

To this end we use the following Cauchy-Schwarz upper bound of this quantity (before going to the limit for N large)

$$\sqrt{\mathbb{E}\left[\int_0^T (\eta_s^N(U_s)-\eta_s(U_s))^2\,ds\right]}\sqrt{N\mathbb{E}\left[\int_0^T (\eta_s^N(\varphi_{s,T}))^2\,ds\right]}$$

Since $\|U\|_{[0,T]}<+\infty$, using the last assertion of the previous lemma and dominated convergence, the first factor goes to zero as N tends to infinity.
For the second term, let us write that for $0\leq s\leq T$,

$$\begin{aligned}
\eta_s^N[\varphi_{s,T}] &= \eta_s^N[\bar{Q}_{s,T}(\varphi)]-\eta_s^N[\bar{Q}_{s,T}(\mathbb{1})]\eta_T(\varphi)\\
&= \eta_s^N[\bar{Q}_{s,T}(\mathbb{1})]\left(\frac{\eta_s^N[\bar{Q}_{s,T}(\varphi)]}{\eta_s^N[\bar{Q}_{s,T}(\mathbb{1})]}-\eta_T(\varphi)\right)\\
&= \eta_s^N[\bar{Q}_{s,T}(\mathbb{1})]\left(\frac{\eta_s^N[Q_{s,T}(\varphi)]}{\eta_s^N[Q_{s,T}(\mathbb{1})]}-\frac{\eta_s[Q_{s,T}(\varphi)]}{\eta_s[Q_{s,T}(\mathbb{1})]}\right)
\end{aligned}$$

This yields that

$$\begin{aligned}
\eta_s^N[\varphi_{s,T}]^2 \leq\ & 2\eta_s^N[\bar{Q}_{s,T}(\mathbb{1})]^2\left(\frac{(\eta_s^N[Q_{s,T}(\varphi)]-\eta_s[Q_{s,T}(\varphi)])^2}{\eta_s^N[Q_{s,T}(\mathbb{1})]^2}+\right.\\
&\left.\frac{\eta_s[Q_{s,T}(\varphi)]^2}{\eta_s[Q_{s,T}(\mathbb{1})]^2\eta_s^N[Q_{s,T}(\mathbb{1})]^2}\left(\eta_s^N[Q_{s,T}(\mathbb{1})]-\eta_s[Q_{s,T}(\mathbb{1})]\right)^2\right)
\end{aligned}$$

But the proof of Theorem 3.19 shows in fact that for all $\varphi\in\mathcal{A}$

$$\sup_{0\leq s\leq T} N\mathbb{E}[(\eta_s^N[Q_{s,T}(\varphi)]-\eta_s[Q_{s,T}(\varphi)])^2] \;<\; +\infty$$

from which one concludes that (112) holds.

■

The important step in this section is the following

Proposition 3.31 *Under (H2)$_T$, the martingale $(\bar{B}_t^N(\varphi))_{0\leq t\leq T}$ converges in law (for the Skorokhod topology in $D([0,+\infty[,\mathbb{R})$) for N large toward a Gaussian centered martingale $(\bar{B}_t(\varphi))_{0\leq t\leq T}$ whose increasing process is the deterministic mapping*

$$[0,T]\ni t\mapsto\int_0^t \bar{G}(s,T,\eta_s,\varphi,\varphi)\,ds$$

and whose initial value $\bar{B}_0(\varphi)$ admits

$$\sigma(\varphi)\;\stackrel{def.}{=}\;\eta_0[(\bar{Q}_{0,T}(\varphi-\eta_T(\varphi)))^2]$$

for variance.

Proof: Conditionning with respect to the σ-algebra associated with time 0, it is not so difficult to realize that it is sufficient to prove the two following lemmas.

■

Lemma 3.32 *The random variables $\sqrt{N}\eta_0^N(\varphi_{0,T})$ converge in law for N large toward a centered Gaussian law of variance $\sigma(\varphi)$.*

Proof: It is just the usual central limit theorem for the independent variables

$$(\varphi_{0,T}(\xi_{i,0}^N))_{1\leq i\leq N},$$

where the $\xi_{i,0}^N$, $1 \leq i \leq N$, have the same law η_0. We would have noticed that $\eta_0(\varphi_{0,T}) = 0$.

∎

Lemma 3.33 *Consider the process $(\check{B}_t^N(\varphi))_{0\leq t\leq T} = (\bar{B}_t^N(\varphi) - \bar{B}_0^N(\varphi))_{0\leq t\leq T}$, under the law $\widetilde{\mathbb{P}}$ which is constructed as $\mathbb{P}$, except that the initial distribution of ξ_0^N is a Dirac measure $\delta_{x_0^N}$ for some $x_0^N \in E^N$. Then $(\check{B}_t^N(\varphi))_{0\leq t\leq T}$ converge in law toward a Gaussian centered martingale $(\check{B}_t(\varphi))_{0\leq t\leq T}$ starting in 0 and whose increasing process is the deterministic mapping*

$$[0,T] \ni t \mapsto \int_0^t \bar{G}(s,T,\eta_s,\varphi,\varphi)\,ds$$

Proof: Since $(\mathrm{H2})_T$ is assumed to be satisfied for all $\eta_0 \in \mathbf{M}_1(E)$, $(\check{B}_t^N(\varphi))_{0\leq t\leq T}$ is again a martingale under $\widetilde{\mathbb{P}}$, and we also have that

$$\langle \check{B}^N(\varphi)\rangle_t \;=\; \int_0^t \bar{G}(s,T,\eta_s^N,\varphi,\varphi)\,ds$$

Thanks to Theorem 3.11 p. 432 of [68] it is now enough to prove that

$$\forall\ \epsilon > 0, \qquad \lim_{N\to+\infty} \widetilde{\mathbb{P}}\left[\sup_{0\leq s\leq T}\left|\triangle \check{B}_s^N(\varphi)\right| \geq \epsilon\right] \;=\; 0 \tag{113}$$

and

$$\forall\ 0 \leq t \leq T, \qquad \lim_{N\to+\infty}\langle \check{B}^N(\varphi)\rangle_t \;=\; \int_0^t G(s,T,\eta_s,\varphi)\,ds \tag{114}$$

where the last limit is understood in probability ($\widetilde{\mathbb{P}}$).
But (113) is proved in quite the same way as Lemma 3.23, and (114) comes from a dominated convergence theorem, using the weak propagation of chaos and the fact that for all $0 \leq s \leq T$, U_s, $\varphi_{s,T}$ and $\bar{\Gamma}(\varphi_{\cdot,T},\varphi_{\cdot,T})(s,\cdot) \in \mathcal{B}_{\mathrm{b}}(E)$.

∎

Putting together the previous calculations, we also see that the process

$$(\eta_t^N(\varphi_{t,T}))_{0\leq t\leq T}$$

converge to the same limit as the one presented in Proposition 3.31.

Now the Theorem 3.28 follows, by considering terminal values. More precisely, it remains to replace φ by $(\varphi_1,\varphi_2,\cdots,\varphi_p)$, where $\varphi_1,\varphi_2,\cdots,\varphi_p \in \mathcal{A}$, $p \geq 1$, and to use linear relations like $\bar{B}^N(\varphi_i+\varphi_j)=\bar{B}^N(\varphi_i)+\bar{B}^N(\varphi_j)$, $1\leq i\neq j\leq p$, which implies for instance that for any $1\leq i\neq j\leq p$

$$\langle \bar{B}^N(\varphi_i),\bar{B}^N(\varphi_j)\rangle = \frac{1}{4}(\langle \bar{B}^N(\varphi_i+\varphi_j)\rangle - \langle \bar{B}^N(\varphi_i-\varphi_j)\rangle)$$

The details are left to the reader.

The expression for the limit covariance can be simplified, in order to see that in fact it depends continuously on ϕ,φ with respect to the norm $\|\cdot\|$. Furthermore, it will confirm that it doesn't depend on the choice of $\bar{\Gamma}(\varphi_{\cdot,T},\varphi_{\cdot,T})$.

Lemma 3.34 *More precisely, for any $\phi,\varphi\in\mathcal{A}$ we have*

$$\mathbb{E}[W_T(\phi)W_T(\varphi)] = \eta_T[(\varphi-\eta_T(\varphi))(\phi-\eta_T(\phi))]+2\int_0^T \eta_s[\varphi_{s,T}\phi_{s,T}U_s]\,ds$$

Proof: By a symmetrization procedure, it is sufficient to consider the case $\phi=\varphi$. Then we use calculations similar to those of Lemma 3.21.

To this end we first recall that

$$\begin{aligned}&\textstyle\int_0^T \eta_s[\bar{L}(\varphi^2_{\cdot,T})(s,\cdot)]\,ds\\ &= \int_0^T \mathbb{E}_{\eta_0}\left[\bar{L}(\varphi^2_{\cdot,T})(s,X_s)\exp\left(\int_0^s U_u(X_u)-\eta_u(U_u)\,du\right)\right]\,ds\end{aligned}$$

Writing for all $0\leq s\leq T$,

$$\varphi^2_{s,T}(X_s) = \varphi^2_{0,T}(X_0)+\int_0^s \bar{L}(\varphi^2_{\cdot,T})(u,X_u)\,du+M_s^{(\varphi^2_{\cdot,T})}$$

with a certain martingale $(M_s^{(\varphi^2_{\cdot,T})})_{s\geq 0}$, we deduce that

$$\begin{aligned}&\int_0^T \mathbb{E}_{\eta_0}\left[\bar{L}(\varphi^2_{\cdot,T})(s,X_s)\exp\left(\int_0^s U_u(X_u)-\eta_u(U_u)\,du\right)\right]\,ds\\
&= \mathbb{E}_{\eta_0}\left[\varphi^2_{T,T}(X_T)\exp\left(\int_0^T U_s(X_s)-\eta_s(U_s)\,ds\right)\right]-\mathbb{E}_{\eta_0}\left[\varphi^2_{0,T}(X_0)\right]-\\
&\qquad \mathbb{E}_{\eta_0}\left[\int_0^T \varphi^2_{s,T}(X_s)(U_s(X_s)-\eta_s(U_s))\exp\left(\int_0^s U_u(X_u)-\eta_u(U_u)\,du\right)\,ds\right]\\
&= \eta_T[\varphi^2_{T,T}]-\eta_0[\varphi^2_{0,T}]-\int_0^T \eta_s[\varphi^2_{s,T}(U_s-\eta_s(U_s))]\,ds\end{aligned}$$

Therefore

$$\begin{aligned}&\int_0^T \eta_s[\bar{\Gamma}(\varphi_{\cdot,T},\varphi_{\cdot,T})(s,\cdot)]\,ds\\
&= \int_0^T \eta_s[\bar{L}(\varphi^2_{\cdot,T})(s,\cdot)-2\varphi_{s,T}\bar{L}(\varphi_{\cdot,T})(s,\cdot)]\,ds\\
&= \eta_T[\varphi^2_{T,T}]-\eta_0[\varphi^2_{0,T}]-\int_0^T \eta_s[\varphi^2_{s,T}(U_s-\eta_s(U_s))+2\varphi_{s,T}\bar{L}(\varphi_{\cdot,T})]\,ds\end{aligned}$$

But it appears from the definition of $\varphi_{s,T}$, that

$$\forall\ (s,x)\in[0,T]\times E^N,$$

$$\begin{aligned}
\bar{L}(\varphi_{\cdot,T})(s,x) &= \exp\left(-\int_s^T \eta_u(U_u)\,du\right)\bar{L}(Q_{\cdot,T}(\varphi))(s,x)+\eta_s(U_s)\varphi_{s,T}(x)\\
&= -(U_s(x)-\eta_s(U_s))\varphi_{s,T}(x)
\end{aligned}$$

so

$$\begin{aligned}
&\int_0^T \eta_s[\bar{\Gamma}(\varphi_{\cdot,T},\varphi_{\cdot,T})(s,\cdot)]\,ds\\
&= \eta_T[\varphi_{T,T}^2]-\eta_0[\varphi_{0,T}^2]+\int_0^T \eta_s[\varphi_{s,T}^2(U_s-\eta_s(U_s))]\,ds
\end{aligned}$$

and finally the lemma follows, taking into account that

$$\forall\ 0\leq s\leq T,\qquad \eta_s(\varphi_{s,T}) \;=\; \eta_T(\varphi-\eta_T(\varphi)) \;=\; 0$$

■

Remark 3.35: This is a first step which could lead to the conclusion that the Theorem 3.28 is true more generally for the family $(W_T^N(\varphi))_{\varphi\in\mathcal{B}_b(E)}$. Note also that in the trivial case where $U\equiv 0$, we find the classical covariance for independent particles. The term $2\int_0^T \eta_s[\varphi_{s,T}^2 U_s]\,ds$ gives a measurement of the noise introduced by interactions.

3.3.3 Exponential Bounds

We can also take advantage of the martingales we have exhibited in the previous sections to obtain exponential bounds on deviations from the limit. Here we will use the strong regularity assumptions considered in section 3.1.2, because a.s. bounds on the increasing processes will be needed (and not only $\mathbb{L}^1$ estimations, as for the weak propagation of chaos).

Our starting point will be the following basic result from the general theory of martingales (cf for instance the Corollary 3.3 of [86] using calculations from the section 4.13 of [81]):

Proposition 3.36 *Let $(M_t)_{0\leq t\leq T}$ be a martingale starting from 0, ie $M_0=0$, and such that for a constant $a\geq 0$, we have a.s., $\sup_{0\leq t\leq T}|\triangle M_t|\leq a$. Then M is locally square integrable and we are assured of the bounds*

$$\forall\ G\geq 0,\ \forall\ 0\leq\epsilon\leq G/a,\qquad \mathbb{P}\left[\sup_{0\leq t\leq T}|M_t|\geq\epsilon,\ \langle M\rangle_T\geq G\right]\;\leq\; 2\exp\left(\frac{-\epsilon^2}{4G}\right)$$

Now the procedure to get exponential upper bounds for the deviations $W_T^N(\varphi)$ introduced in Theorem 3.28 is quite standard.

First we have to estimate the "characteristics" of the martingale $\widetilde{B}(\varphi)$ defined in the proof of Proposition 3.22, for $\varphi\in\mathcal{A}$.

Lemma 3.37 *For $T>0$ given, there exist two constants $C_T^{(10)}$, $C_T^{(11)}\geq 0$, such that for all $\varphi\in\mathcal{A}$, we have almost surely,*

$$\begin{aligned}\sup_{0\leq t\leq T}\left|\triangle\widetilde{B}_t^N(\varphi)\right| &\leq \frac{C_T^{(10)}\,\|\varphi\|}{N}\\ \langle\widetilde{B}^N(\varphi)\rangle_T &\leq \frac{C_T^{(11)}|||\varphi|||_{[0,T]}^2}{N}\end{aligned}$$

where the norm $|||\cdot|||_{[0,T]}$ was defined by the formula (105) given page 96.

Proof: Using the following relation valid for all $0\leq t\leq T$,

$$\widetilde{B}_t^N(\varphi) \;=\; \int_0^t \exp\left(\int_0^s \eta_u^N(U_u)\,du\right)\,dB_s^N(\varphi)$$

clearly it is enough to get these upper bounds with $\widetilde{B}^N(\varphi)$ replaced by the martingale $B^N(\varphi)$ considered in Lemma 3.20. But then the required estimations are deduced at once from Lemma 3.20 and Lemma 3.23.

■

Taking into account a basic result on iid random variables, we end up with

Proposition 3.38 *There exists a constant $C_T^{(12)}\geq 0$ such that for all $\varphi\in\mathcal{A}$,*

$$\forall\,\epsilon>0,\;\; \mathbb{P}\left[\sup_{0\leq t\leq T}\left|\gamma_t^N(Q_{t,T}(\varphi))-\gamma_T(\varphi)\right|\geq\epsilon\right]\;\leq\;4\exp\left(\frac{-NC_T^{(12)}\epsilon^2}{\|\varphi\|^2\vee|||\varphi|||_{[0,T]}^2}\right)$$

Proof: Recall that $\gamma_T(\varphi)=\gamma_0(Q_{0,T}(\varphi))$, so we can write the decomposition

$$\begin{aligned}&\gamma_t^N(Q_{t,T}(\varphi))-\gamma_T(\varphi)\\ &=\gamma_t^N(Q_{t,T}(\varphi))-\gamma_0^N(Q_{0,T}(\varphi))+\gamma_0^N(Q_{0,T}(\varphi))-\gamma_0(Q_{0,T}(\varphi))\\ &=\widetilde{B}_t^N(\varphi)+\gamma_0^N(Q_{0,T}(\varphi))-\gamma_0(Q_{0,T}(\varphi))\end{aligned}$$

It is then sufficient to prove that for all $\epsilon>0$,

$$\begin{aligned}\mathbb{P}\left[\sup_{0\leq t\leq T}\left|\widetilde{B}_t^N(\varphi)\right|\geq\epsilon/2\right] &\leq 2\exp\left(\frac{-NC_T^{(13)}\epsilon^2}{\|\varphi\|^2\vee|||\varphi|||_{[0,T]}^2}\right)\\ \mathbb{P}\left[\left|\gamma_0^N(Q_{0,T}(\varphi))-\gamma_0(Q_{0,T}(\varphi))\right|\geq\epsilon/2\right] &\leq 2\exp\left(\frac{-NC_T^{(14)}\epsilon^2}{\|\varphi\|^2}\right)\end{aligned}$$

for some constants $C_T^{(13)}$, $C_T^{(14)}\geq 0$.
For the first inequality, we note that obviously

$$\left|\widetilde{B}_T^N(\varphi)\right| \;\leq\; \exp(T\,\|U\|_{[0,T]})\,\|\varphi\|$$

so we only have to prove it for $0 < \epsilon \leq \exp(T \|U\|_{[0,T]}) \|\varphi\|$, but then, by using the estimates of the lemma above, it comes from the Proposition 3.36 applied with $M = \widetilde{B}(\varphi)$, $a = C_T^{(10)} \|\varphi\| / N$ and

$$G = [C_T^{(11)} \vee C_T^{(10)} \exp(T \|U\|_{[0,T]})](\|\varphi\|^2 \vee |||\varphi|||_{[0,T]}^2)/N.$$

The second bound is just a consequence of the Hoeffding inequality for iid random variables (cf [50]), which states that for all $\epsilon > 0$,

$$\mathbb{P}\left[\left|\gamma_0^N(Q_{0,T}(\varphi)) - \gamma_0(Q_{0,T}(\varphi))\right| \geq \epsilon/2\right] \quad \leq \quad 2\exp\left(\frac{-N\epsilon^2}{8\left\|Q_{0,T}(\varphi)\right\|^2}\right)$$

so we can take $C_T^{(14)} = \exp(-2T \|U\|_{[0,T]})/8$.

■

Now the conclusion follows easily:

Theorem 3.39 *There exists a constant $C_T^{(15)} > 0$ such that*

$$\forall\, \varphi \in \mathcal{A}, \forall\, \epsilon > 0, \qquad \mathbb{P}[\left|W_T^N(\varphi)\right| \geq \epsilon] \quad \leq \quad 4\exp\left(\frac{-C_T^{(15)}\epsilon^2}{\|\varphi\|^2 \vee |||\varphi|||_{[0,T]}^2}\right)$$

where we recall that the fluctuation are given by $W_T^N(\varphi) = \sqrt{N}(\eta_T^N(\varphi) - \eta_T(\varphi))$, for $\varphi \in \mathcal{A}$.

Proof: We deduce this result from the usual decomposition (108) and the inequality of the latter proposition, considered at time $t = T$. If we use the full strength of the uniformity over the interval $[0, T]$, then for any $\varphi \in \mathcal{A}$ and for any $\epsilon > 0$ we rather end up with

$$\mathbb{P}[\sup_{0\leq t\leq T} \left|\Phi_{t,T}(\eta_t^N)(\varphi) - \eta_T(\varphi)\right| \geq \epsilon] \quad \leq \quad 4\exp\left(\frac{-NC_T^{(15)}\epsilon^2}{\|\varphi\|^2 \vee |||\varphi|||_{[0,T]}^2}\right)$$

■

Remarks 3.40:

a) The norm $|||\cdot|||_{[0,T]}$ does not seem an easy object to manipulate, but the considerations of section 3.1.2 enables us to replace it by more convenient ones, for instance, under hypotheses $(H4)_T$, $(H7)_T$, $(H8)_T$, $(H9)_T$ and $(H10)_T$, it appears that for a constant $C_T^{(16)} > 0$ depending only on $T \geq 0$, we have

$$|||\cdot|||_{[0,T]} \quad \leq \quad C_T^{(16)} ||| \cdot |||$$

b) Theorem 3.39 is a first step in the direction of a $\mathbb{L}^p$ Glivenko-Cantelli result, since it shows that for $T \geq 0$ fixed, the process

$$(\eta_T^N(\varphi) - \eta_T(\varphi))_{\varphi\in\mathcal{A}}$$

indexed by $\mathcal{A}$ is sub-Gaussian with respect to the norm $\|\cdot\| \vee |||\cdot|||_{[0,T]}$.

So if for a class of functions $\mathcal{F} \subset \mathcal{A}$, we have enough information about the packing and covering numbers of $\mathcal{F}$ with respect to $\|\cdot\| \vee \|\!|\!| \cdot |\!|\!\|_{[0,T]}$ (or more conveniently, with respect to $|\!|\!| \cdot |\!|\!|$, under the appropriate hypotheses), then we could conclude to results similar to those presented in section 2.2.3.

c) We also notice that the Proposition 3.25 could classically be deduced from the previous theorem (cf for instance [41]), except that we end up with the norm $\|\cdot\| \vee |\!|\!| \cdot |\!|\!|_{[0,T]}$ in the rhs of the inequality given there, instead of $\|\cdot\|$. This leads to the question of whether the Theorem 3.39 would not be satisfied with that norm.

4 Applications to Non Linear Filtering

4.1 Introduction

The object of this section is to apply the results obtained in previous sections to nonlinear filtering problems. We will study continuous time as well as discrete time filtering problems. For a detailed discussion of the filtering problem the reader is referred to the pioneering paper of Stratonovich [104] and to the more rigorous studies of Shiryaev [99] and Kallianpur-Striebel [72]. More recent developments can be found in Ocone [89] and Pardoux [90].

In continuous time settings the desired conditional distributions can be regarded as a Markov process taking values in the space of all probability measures. The corresponding evolution equation is usually called the Kushner-Stratonovitch equation. The most important measure of complexity is the infinite dimensionality of the state space of this equation.

In the first section 4.2 we formulate the continuous time nonlinear filtering problem in such a way that the results of section 3 can be applied. In section 4.3 we present an alternative approach to approximate a continuous time filtering problem. This approach is based on a commonly used time discretization procedure (see for instance [82, 74, 76, 58, 90] and [92, 94, 93, 108]).
We shall see that the resulting discrete time model has the same form as in (8) and it also characterizes the evolution in time of the optimal filter for a suitably defined discrete time filtering problem.
The fundamental difference between the Moran's type IPS and the genetic type IPS associated with this additional level of discretization lies in the fact that in the Moran IPS competitive interactions occur randomly. The resulting scheme is therefore a genuine continuous time and particle approximating model of the nonlinear filtering equation.

In section 4.4 we briefly describe the discrete time filtering problem. It will be transparent from this formulation that the desired flow of distributions have the same form as the one considered in this work (see (8) section 1.3). We will also remark that the fitness functions $\{g_n\,;\,n \geq 1\}$ and therefore the constants $\{a_n\,;\,n \geq 1\}$ defined in condition $(\mathcal{G})$ depend on the observation process so that the analysis given in previous sections will also lead to quenched results.

For instance the covariance function in Donsker Theorem and the rates functions in LDP will now depend on the observation record.

One natural question that one may ask if whether the averaged version of the stability results of section 2.1.2 and the $\mathbb{L}^p$-error bounds given in section 2.2.2 hold. In many practical situations the functions $\{a_n \; ; \; n \geq 1\}$ have a rather complicated form and it is difficult to obtain an averaged version of some results such as the exponential rates given in Theorem 2.15. Nevertheless we will see in section 4.4.2 that the averaged version of the stability results given in section 2.1.2 as well as the averaged version of the $\mathbb{L}^p$-uniform bounds given in section 2.2.2 hold for a large class of nonlinear sensors.

4.2 Continuous Time Filtering Problems

4.2.1 Description of the Models

The aim of this section is to formulate some classical nonlinear filtering problems in such a way that the previous particles interpretations can be naturally applied to them. Here is the heuristic model: let a signal process $S = \{S_t \; ; \; t \geq 0\}$ be given, it is assumed to be a time-homogeneous Markov process with càdlàg paths taking values in the Polish space E. We suppose that this signal is seen through a $\mathbb{R}^d$-valued noisy observation process $Y = \{Y_t \; ; \; t \geq 0\}$ defined by

$$\forall \; t \geq 0, \qquad Y_t \;\; = \;\; \int_0^t \; h(S_s) \; ds + V_t$$

where $V = \{V_t \; ; \; t \geq 0\}$ is a d-vector standard Wiener process independent of S, and h maps somewhat smoothly the signal state space E into $\mathbb{R}^d$.

The traditional filtering problem is concerned with estimating the conditional distribution of S_t given the observational information that is available at time t, $\mathcal{Y}_{[0,t]} = \sigma(Y_s \; ; \; 0 \leq s \leq t)$, ie to evaluate for all $f \in \mathcal{C}_b(E)$,

$$\pi_t(f) \;\; \stackrel{\text{def.}}{=} \;\; \mathbb{E}\left(f(S_t) \,\middle|\, \mathcal{Y}_{[0,t]} \right)$$

More precisely, we will make the assumption that there exist an algebra $\mathcal{A} \subset \mathcal{C}_{\rm b}(E)$ and a pregenerator $L_0 \; : \; \mathcal{A} \to \mathcal{C}_{\rm b}(E)$ such that for any initial distribution $\eta_0 \in \mathbf{M}_1(E)$, there is a unique solution $\widetilde{\mathbb{P}}_{\eta_0}$ to the martingales problem (we refer to the section 3.1.1 for more details) associated with η_0 and L_0 on $\Omega_1 = D(\mathbb{R}_+, E)$, the space of all càdlàg paths from $\mathbb{R}_+$ to E, endowed with its natural σ-algebra. Then $S \stackrel{\text{def.}}{=} (S_t)_{t \geq 0}$ will denote the canonical coordinate process on Ω_1, and from now on, the initial distribution η_0 will be supposed fixed, and we will consider the Markov process S under $\widetilde{\mathbb{P}}_{\eta_0}$.

Let $h = (h_i)_{1 \leq i \leq d}$ be the map from E to $\mathbb{R}^d$ alluded to above, we make the hypothesis that for all $1 \leq i \leq d$, $h_i \in \mathcal{A}$.

Let us also introduce the canonical probability space associated with the observation process: $\Omega_2 = C(\mathbb{R}_+, \mathbb{R}^d)$, the set of all continuous functions from $\mathbb{R}_+$ to $\mathbb{R}^d$, and $Y = (Y_t)_{t \geq 0}$ is the coordinate process on Ω_2.

Let us denote $\Omega = \Omega_1 \times \Omega_2$ and $\widetilde{\mathbb{P}}$ the probability on its usual Borelian σ-field such that its marginal on Ω_1 is $\widetilde{\mathbb{P}}_{\eta_0}$ and such that

$$V = (V_t)_{t\geq 0} \quad \stackrel{\text{def.}}{=} \quad \left(Y_t - \int_0^t h(S_s)\, ds\right)_{t\geq 0}$$

is a d-vector standard Brownian motion, as previously mentionned.

In practice, this probability $\widetilde{\mathbb{P}}$ is usually constructed via Girsanov's Theorem from another reference probability measure $\widehat{\mathbb{P}}$ on Ω, under which S and Y are independent, S has law $\widetilde{\mathbb{P}}_{\eta_0}$ and Y is a d-vector standard Brownian motion. For $t \geq 0$, let

$$\mathcal{F}_t = \sigma((S_s, Y_s)\ ;\ 0 \leq s \leq t)$$

be the σ-algebra of events up to time t, the probabilities $\widetilde{\mathbb{P}}$ and $\widehat{\mathbb{P}}$ are in fact equivalent on $\mathcal{F}_t$, and their density is given by

$$\frac{d\widetilde{\mathbb{P}}}{d\widehat{\mathbb{P}}}|\mathcal{F}_t = Z_t(S,Y) \stackrel{\text{def.}}{=} \exp\left(\int_0^t h^\star(S_s)dY_s - \frac{1}{2}\int_0^t h^\star(S_s)h(S_s)ds\right)$$

where we have used standard matrix notations, for instance

$$\int_0^t h^\star(S_s)\, dY_s = \sum_{1\leq i\leq d} \int_0^t h_i(S_s)\, dY_{i,s}$$

Under our assumptions ([56, 90]) one can prove that

$$\pi_t(f) = \frac{\widehat{\mathbb{E}}\left(f(S_t)Z_t(S,Y)\,\middle|\,\mathcal{Y}_{[0,t]}\right)}{\widehat{\mathbb{E}}\left(Z_t(S,Y)\,\middle|\,\mathcal{Y}_{[0,t]}\right)} = \frac{\int_{\Omega_1} f(\theta_t)\ Z_t(\theta,Y)\ \widetilde{\mathbb{P}}_{\eta_0}(d\theta)}{\int_{\Omega_1} Z_t(\theta,Y)\ \widetilde{\mathbb{P}}_{\eta_0}(d\theta)} \tag{115}$$

Using Itô's integration by part formula, in the differential sense we have that

$$h^\star(S_s)dY_s \;=\; d(h^\star(S_s)Y_s) - Y_s^\star\, L_0(h)(S_s)ds - Y_s^\star\, dM_s^{(h)}$$

where $L(h) = (L(h_i))_{1\leq i\leq d}\ :\ E \to \mathbb{R}^d$, and where $M^{(h)} = (M^{(h_i)})_{1\leq i\leq d}$ is a d-vector square integrable continuous martingale (relative to the natural filtration $(\mathcal{F}_t)_{t\geq 0}$) with cross-variation processes given by

$$\forall\ 1 \leq i,j \leq d, \qquad \langle M^{(h_i)}, M^{(h_j)}\rangle_t \;=\; \int_0^t \Gamma_0(h_i,h_j)(S_s)\, ds$$

(as usual Γ_0 is the carré du champ associated with the pregenerator L_0). This yields the decomposition

$$\begin{aligned}
&\ln Z_t(S,Y)\\
&= h^\star(S_t)Y_t - \int_0^t Y_s^\star\, L_0(h)(S_s)\, ds - \int_0^t Y_s^\star\, dM_s^h - \frac{1}{2}\int_0^t h^\star(S_s)h(S_s)ds
\end{aligned}$$

Before going further, let us make the following remark:
As in section 3.1.1, we denote L the operator acting on $\mathbf{A}\otimes\mathcal{A}$ by formula (92), where L_t is replaced by L_0. We also consider $\mathcal{B}_\infty$ the vector space of functions $f \in \mathcal{B}_\text{b}(\mathbb{R}_+\times E)$

for which there exists a function $\bar{L}(f) \in \mathcal{B}_{\mathrm{b}}(\mathbb{R}_+ \times E)$ such that under $\widetilde{\mathbb{P}}_{\eta_0}$, the process $(M_t(f))_{0\leq t\leq T}$ defined by

$$\forall\, 0 \leq t \leq T, \qquad M_t(f) \;=\; f(t, X_t) - f(0, X_0) - \int_0^t \bar{L}(f)(s, X_s)\, ds$$

is a martingale.

We have the following stability property:

Lemma 4.1 *If $f \in \mathbf{A} \otimes \mathcal{A}$, then $\exp(f) \in \mathcal{B}_\infty$.*

Proof: Using the formula $\exp(f) = \sum_{n\geq 0} f^n/n\,!$ and an approximation technique as the one presented in Lemma 3.2, it is enough to show that for all $T > 0$ given, we have as $n, m \to \infty$ (and on $[0,T] \times E$)

$$\begin{aligned} \Gamma\left(\sum_{n\leq p\leq m} f^p/p\,!, \sum_{n\leq p\leq m} f^p/p\,!\right) &\;\rightharpoonup\; 0 \\ L\left(\sum_{n\leq p\leq m} f^p/p\,!\right) &\;\rightharpoonup\; 0 \end{aligned}$$

But these convergence results are easy to obtain, because of the general bounds

$$\begin{aligned} \sqrt{\Gamma\left(\sum_{n\leq p\leq m} f^p/p\,!, \sum_{n\leq p\leq m} f^p/p\,!\right)} &\;\leq\; \sum_{n\leq p\leq m} \sqrt{\Gamma\left(f^p/p\,!, f^p/p\,!\right)} \\ \forall\, p \geq 1, \qquad \|\Gamma(f^p, f^p)\|_{[0,T]} &\;\leq\; p^2\, \|f\|_{[0,T]}^{2p-2}\, \|\Gamma(f,f)\|_{[0,T]} \end{aligned}$$

and of the upper bound on $L(f^n)$ which can be deduced by induction from

$$\begin{aligned} \forall\, n \geq 1, \qquad & \left\|L(f^{n+1})\right\|_{[0,T]} \\ \leq\;& \|f^n\|_{[0,T]}\, \|L(f)\|_{[0,T]} + \|f\|_{[0,T]}\, \|L(f^n)\|_{[0,T]} + \|\Gamma(f^n, f)\|_{[0,T]} \\ \leq\;& \|f^n\|_{[0,T]}\, \|L(f)\|_{[0,T]} + \|f\|_{[0,T]}\, \|L(f^n)\|_{[0,T]} \\ & \qquad + \sqrt{\|\Gamma(f,f)\|_{[0,T]}}\sqrt{\|\Gamma(f^n, f^n)\|_{[0,T]}} \end{aligned}$$

This ends the proof of the lemma. ■

Let $y \in C(\mathbb{R}_+, \mathbb{R}^d)$ be given. For $t \geq 0$ fixed, the map

$$h_t \,:\, E \ni x \;\mapsto\; \sum_{1\leq i\leq d} y_{i,t} h_i(x)$$

belongs to $\mathcal{A}$, so in the same way as above, we can define a function

$$\bar{L}_0(\exp(-h_t)) \in \mathcal{B}_{\mathrm{b}}(E).$$

We easily realize that the application

$$\mathbb{R}_+ \times E \ni (t, x) \;\mapsto\; \bar{L}_0(\exp(-h_t))(x)$$

is continuous as a locally uniform limit of continuous functions.

Then we can consider

$$\begin{aligned}&\breve{Z}_t(S,y)\\&\stackrel{\text{def.}}{=} \exp\left(-\int_0^t y_s^\star\, dM_s^{(h)} - \int_0^t \exp(h_s(S_s))\bar{L}_0(\exp(-h_s))(S_s) + L_0(h_s)(S_s)\, ds\right)\end{aligned}$$

Note that if a sequence $(y_n)_{n\geq 0}$ of elements of $C(\mathbb{R}_+,\mathbb{R}^d)$ converges uniformly on compact subsets of $\mathbb{R}_+$ toward y, then uniformly for t belonging to compact subsets, $\breve{Z}_t(S,y_n)$ converges in probability towards $\breve{Z}_t(S,y)$.

The interest of this quantity is that

Lemma 4.2 *Let a function $y\in C(\mathbb{R}_+,\mathbb{R}^d)$ be fixed. Under $\widetilde{\mathbb{P}}_{\eta_0}$, the process $(\breve{Z}_t(S,y))_{t\geq 0}$ is a martingale.*

Proof: We will first consider the case $y\in C^1(\mathbb{R}_+,\mathbb{R}^d)$ (even if this situation a.s. never occur for Y).

Under this additional assumption, the mapping

$$\bar{h}\,:\,\mathbb{R}_+\times E\ni(t,x)\;\mapsto\;h_t(x)$$

clearly belongs to $\mathbf{A}\otimes\mathcal{A}$, and it appears that if $M^{(\bar{h})}$ is the martingale such that for all $t\geq 0$,

$$\bar{h}(t,S_t) \;=\; \bar{h}(0,S_0) + \int_0^t L(\bar{h})(s,S_s)\,ds + M_t^{(\bar{h})}$$

then in fact it is given by

$$\forall\, t\geq 0,\qquad M_t^{(\bar{h})} \;=\; \int_0^t y_s^*\,dM_s^{(h)}$$

Furthermore, from Lemma 4.1, for any $t\geq 0$ we can write that

$$\exp(-\bar{h}(t,S_t)) \;=\; \exp(-\bar{h}(0,S_0)) + \int_0^t \bar{L}(\exp(-\bar{h}))(s,S_s)\,ds + M_t^{(\exp(-\bar{h}))}$$

for a certain martingale $M^{(\exp(-\bar{h}))}$.
On the other hand it appears without difficulty that for any $(s,x)\in\mathbb{R}_+\times E$

$$\exp(h_s(x))\bar{L}_0(\exp(-h_s))(x) + L_0(h_s)(x)$$

$$= \exp(\bar{h}(s,x))\bar{L}(\exp(-\bar{h}))(s,x) + L(\bar{h})(s,x)$$

(ie the time derivatives cancel). As a result, for any $t\geq 0$ we have that

$$\breve{Z}_t(S,y) = \exp\left(-\bar{h}(t,S_t) + \bar{h}(0,S_0) + \int_0^t \exp(\bar{h}(s,S_s))\bar{L}(\exp(-\bar{h}))(s,S_s)\,ds\right)$$

This may also be written in differential form

$$\begin{aligned}
&d\breve{Z}_t(S,y)\\
&= \breve{Z}_{t-}(S,y)\Big(\exp(\bar{h}(t-,S_{t-}))\bar{L}(\exp(-\bar{h}))(t-,S_{t-})\,dt\\
&+\exp(\bar{h}(t-,S_{t-}))\,dM_t^{(\exp(-\bar{h}))} - \exp(\bar{h}(t-,S_{t-}))\bar{L}(\exp(-\bar{h}))(t-,S_{t-})\,dt\Big)\\
&= \breve{Z}_{t-}(S,y)\exp(\bar{h}(t-,S_{t-}))\,dM_t^{(\exp(-\bar{h}))}
\end{aligned}$$

It follows that $(\breve{Z}_t(S,y))_{t\geq 0}$ is a martingale. Indeed it is more precisely the Doléans-Dade exponential of the martingale

$$\left(\int_0^t \exp(\bar{h}(s-,S_{s-}))\,dM_s^{(\exp(-\bar{h}))}\right)_{t\geq 0}$$

Thus, we see that for all $0\leq s\leq t$ and any random variables H_s which are measurable with respect to $\sigma(S_u\,;\,0\leq u\leq s)$, we have

$$\widetilde{\mathbb{E}}_{\eta_0}[H_s(\breve{Z}_t(S,y)-\breve{Z}_s(S,y))] \;=\; 0$$

We end the proof by noting that the left hand side is continuous with respect to $y\in C(\mathbb{R}_+,\mathbb{R}^d)$, if this set is endowed with the uniform convergence on compact subsets of $\mathbb{R}_+$.

■

Let us write, for $y\in C(\mathbb{R}_+,\mathbb{R}^d)$,

$$\forall\, t\geq 0,\qquad \ln Z_t(S,y) \;=\; h^\star(S_t)y_t + \int_0^t\, V(S_s,y_s)\;ds + \ln\breve{Z}_t(S,y)$$

where for all $(x,y)\in E\times\mathbb{R}^d$,

$$V(x,y) \;=\; \exp(y^* h(x))\bar{L}_0[\exp(-y^* h(\,\cdot\,))](x) - \frac{1}{2}h^\star(x)h(x)$$

Together with (115) this decomposition implies that

$$\pi_t(f) = \frac{\displaystyle\int_{\Omega_1} f(\theta_t)\; e^{h^\star(\theta_t)Y_t + \int_0^t\, V(\theta_s,Y_s)\;ds}\;\mathbb{P}_{\eta_0}^{[Y]}(d\theta)}{\displaystyle\int_{\Omega_1} e^{h^\star(\theta_t)Y_t + \int_0^t\, V(\theta_s,Y_s)\;ds}\;\mathbb{P}_{\eta_0}^{[Y]}(d\theta)}$$

where, for any $y\in\mathcal{C}(\mathbb{R}_+,\mathbb{R}^d)$, $\mathbb{P}_{\eta_0}^{[y]}$ is the probability measure on Ω_1 defined by its restrictions to $\mathcal{F}_t^{(1)}=\sigma(S_s,\;0\leq s\leq t)$:

$$\frac{d\mathbb{P}_{\eta_0}^{[y]}}{d\widetilde{\mathbb{P}}_{\eta_0}}|\mathcal{F}_t^{(1)} = \breve{Z}_t(S,y)$$

Here is a more tractable caracterisation of $\mathbb{P}_{\eta_0}^{[y]}$:

Proposition 4.3 *Fix a mapping $y \in C(\mathbb{R}_+, \mathbb{R}^d)$, and consider for $t \geq 0$ an operator L_t given on $\mathcal{A}$ by*

$$\forall\, \varphi \in \mathcal{A}, \qquad L_t(\varphi) \;=\; L_0(\varphi) + \exp(h_t)\bar{\Gamma}_0(\exp(-h_t), \varphi)$$

Then $(L_t)_{t\geq 0}$ is a measurable family of pregenerators, and $\mathbb{P}_{\eta_0}^{[y]}$ is the unique solution to the martingale problem associated with the initial condition η_0 and to this family.

We would have noticed there is no real difficulty in defining $\bar{\Gamma}(\exp(-h_t), \varphi)$ for $\varphi \in \mathcal{A}$.
Proof: We shall verify that for all $\varphi \in \mathcal{A}$, all $0 \leq s \leq t$ and all random variable H_s which is $\sigma(S_u\,;\, 0 \leq u \leq s)$-measurable, we have

$$\mathbb{E}_{\eta_0}^{[y]}\left[H_s\left(\varphi(S_t) - \varphi(S_s) - \int_s^t L_u(\varphi)(X_u)\,du\right)\right] \;=\; 0$$

i.e.

$$\widetilde{\mathbb{E}}_{\eta_0}\left[\breve{Z}_t(S,y)H_s\,(N_t - N_s)\right] \;=\; 0$$

where for all $t \geq 0$,

$$N_t \;=\; \varphi(S_t) - \varphi(S_0) - \int_0^t L_u(\varphi)(X_u)\,du$$

So as in Lemma 4.2, by continuity, we can assume that $y \in C^1(\mathbb{R}_+, \mathbb{R}^d)$.
With this assumption enforced, we have

$$\begin{aligned}
&d(\breve{Z}_t(S,y)N_t)\\
&= \breve{Z}_{t-}(S,y)((L_0(\varphi) - L_t(\varphi))dt + dM_t^{(\varphi)}) + N_{t-}d\breve{Z}_t(S,y) + d\langle M^{(\varphi)}, \breve{Z}(S,y)\rangle_t\\
&= \breve{Z}_{t-}(S,y)[(L_0(\varphi) - L_t(\varphi))dt + \exp(\bar{h}(t,S_t))d\langle M^{(\varphi)}, M^{(\exp(-\bar{h}))}\rangle_t]\\
&\qquad + \breve{Z}_{t-}(S,y)dM_t^{(\varphi)} + N_{t-}d\breve{Z}_t(S,y)\\
&= \breve{Z}_{t-}(S,y)[L_0(\varphi) - L_t(\varphi) + \exp(\bar{h}(t,S_t))\bar{\Gamma}(\exp(-\bar{h})), \varphi)(t,S_t)]\,dt\\
&\qquad + \breve{Z}_{t-}(S,y)dM_t^{(\varphi)} + N_{t-}d\breve{Z}_t(S,y)
\end{aligned}$$

Since

$$\widetilde{\mathbb{E}}_{\eta_0}\left[\breve{Z}_t(S,y)H_s\,(N_t - N_s)\right] \;=\; \widetilde{\mathbb{E}}_{\eta_0}\left[H_s\left(\breve{Z}_t(S,y)N_t - \breve{Z}_s(S,y)N_s\right)\right]$$

we need to check that $(\breve{Z}_t(S,y)N_t)_{t\geq 0}$ is a martingale. To this end it is enough to see that for any $(t,x) \in \mathbb{R}_+ \times E$

$$L_0(\varphi)(x) - L_t(\varphi)(x) + \exp(\bar{h}(t,x))\bar{\Gamma}(\exp(-\bar{h}), \varphi)(t,x) \;=\; 0$$

Next, since for any $(t,x) \in \mathbb{R}_+ \times E$

$$\exp(\bar{h}(t,x))\bar{\Gamma}(\exp(-\bar{h}),\varphi)(t,x) \;=\; \exp(h_t(x))\bar{\Gamma}_0(\exp(-h_t),\varphi)(x)$$

L_t is well defined.

The fact that, for any $t \geq 0$ L_t is a pregenerator comes from the previous considerations, taking $y \in C(\mathbb{R}_+, \mathbb{R}^d)$ defined by

$$\forall\, s \geq 0, \qquad y_s \;=\; y_t$$

And the uniqueness property comes from the one of $\widetilde{\mathbb{P}}_{\eta_0}$, since if $\breve{\mathbb{P}}$ is a solution to the martingale problem associated to $(L_t)_{t\geq 0}$, then it can be shown that the probability $\check{\mathbb{P}}$ defined on Ω_1 by

$$\forall\, t \geq 0, \qquad \frac{d\check{\mathbb{P}}}{d\breve{\mathbb{P}}}|\mathcal{F}_t^{(1)} \;=\; \frac{1}{\breve{Z}_t(S,y)}$$

is solution to the time-homogeneous martingale problem associated to L_0 (all initial conditions being η_0).

∎

The above formulation of the optimal filter can be regarded as a path-wise filter

$$\begin{array}{rccc} \pi_t : & \mathcal{C}([0,T]) & \longrightarrow & \mathbf{M}_1(E) \\ & y & \mapsto & \pi_{y,t} \end{array}$$

where the probability $\pi_{y,t}$ is given by

$$\forall\, f \in \mathcal{B}_{\mathrm{b}}(E), \qquad \pi_{y,t}(f) \;=\; \frac{\displaystyle\int_{\Omega_1} f(\theta_t)\; e^{h^\star(\theta_t)y_t + \int_0^t\, V(\theta_s,y_s)\, ds}\; \mathbb{P}_{\eta_0}^{[y]}(d\theta)}{\displaystyle\int_{\Omega_1} e^{h^\star(\theta_t)y_t + \int_0^t\, V(\theta_s,y_s)\, ds}\; \mathbb{P}_{\eta_0}^{[y]}(d\theta)}$$

This gives a description of the optimal filter in terms of Feynman-Kac formulae as those presented in (3) and (4) in the introduction. Namely,

$$\forall\, f \in \mathcal{B}_{\mathrm{b}}(E), \qquad \pi_{y,t}(f) \;=\; \frac{\displaystyle\int_E f(x)\; e^{h^\star(x)y_t}\; \eta_{y,t}(dx)}{\displaystyle\int_E e^{h^\star(x)y_t}\; \eta_{y,t}(dx)}$$

where

$$\forall\, f \in \mathcal{B}_{\mathrm{b}}(E), \qquad \eta_{y,t}(f) \;\stackrel{\text{def.}}{=}\; \frac{\gamma_{y,t}(f)}{\gamma_{y,t}(1)}$$

and

$$\forall\, f \in \mathcal{B}_{\mathrm{b}}(E), \qquad \gamma_{y,t}(f) \;=\; \mathbb{E}_{\eta_0}^{[y]}\left[f(X_t)\; e^{\int_0^t\, V(X_s,y_s)\, ds}\right]$$

In contrast to (115) we notice that the previous formulations do not involve stochastic integrations and therefore it is well defined for all observation paths and not only on a set of probability measure 1. This formulation is necessary to study the robustness of the optimal filter (that is the continuity of the filter with respect to the observation process), and is also essential to construct robust approximations of the optimal filter, as our interacting particles scheme.

Remarks 4.4:
(a) Note that the condition (H1) is automatically verified for family of pregenerators $(L_t)_{t\geq 0}$ constructed as in Proposition 4.3.

(b) The change of probability presented in that proposition is rather well known in case of diffusions (cf for instance [90] and [96]), ie when the trajectories of S are continuous.
Then we can suppose that $\mathcal{A}$ is stable by composition with C^∞ functions and we have
$\forall\ F \in C^\infty(\mathbb{R}), \forall\ \phi, \varphi \in \mathcal{A}$,

$$\begin{aligned} L_0(F(\varphi)) &= F'(\varphi)L_0(\varphi) + \frac{F''(\varphi)}{2}\Gamma_0(\varphi,\varphi) \\ \Gamma_0(F(\varphi),\phi) &= F'(\varphi)\Gamma_0(\varphi,\phi) \end{aligned}$$

So in the previous expressions, we can replace

$$\begin{aligned} \exp(h_t)\Gamma_0(\exp(-h_t))(x) \quad &\text{by} \quad -\Gamma_0(h_t,\varphi) \\ \exp(y^* h(x))\bar{L}_0(\exp(-y^* h(\,\cdot\,)))(x) \quad &\text{by} \quad \frac{1}{2}\, y^\star\, \Gamma_0(h,h)(x)y - y^\star\, L_0(h)(x) \end{aligned}$$

where $\Gamma_0(h,h)(x)$ denote the matrix $(\Gamma_0(h_i,h_j)(x))_{1\leq i,j\leq d}$.

4.2.2 Examples

Here are some classical examples that can be handled in our framework. The map h and the function $y \in C(\mathbb{R}_+,\mathbb{R}^d)$ will be given as before.

• Bounded Generators

The simplest example a pregenerator L_0 is that of a bounded generator. Namely, let $L_0 : E\times\mathcal{E} \to \mathbb{R}$ be a signed kernel such that

- for any $x \in E$, $L_0(x, .\cap (E\setminus\{x\}) \in \mathbf{M}(E)$ and $L_0(x,E) = 0$
- for any $A \in \mathcal{E}$, $E \ni x \mapsto L_0(x,A) \in \mathbb{R}$ is a measurable function
- there exists a constant $0 \leq M < \infty$ such that

$$\forall\ x\in E, \qquad L_0(x, E\setminus\{x\}) \ \leq\ M$$

Then for any function $f \in \mathcal{B}_\mathrm{b}(E)$, we define

$$\forall\ x\in E, \qquad L_0(f)(x) \ = \ \int f(y)\, L_0(x,dy)$$

We can take here $\mathcal{A} = \mathcal{B}_\mathrm{b}(E)$. We calculate that the carré du champ is given for any $\phi,\varphi \in \mathcal{A}$ and $x \in E$ by

$$\Gamma_0(\phi,\varphi)(x) \ = \ \int L_0(x,dy)\,(\phi(y)-\phi(x))(\varphi(y)-\varphi(x))$$

so it appears that for any $t \geq 0$, $\varphi \in \mathcal{A}$ and $x \in E$

$$L_t(\varphi)(x) = \int L_0(x, dy) \exp(h_t(x) - h_t(y)) (\varphi(y) - \varphi(x))$$

They are again jump generators, and the rate of transition from x to y at time $t \geq 0$ has just been multiplied by $\exp(h_t(x) - h_t(y))$.
For this kind of generators, all our hypotheses are trivially satisfied.

• Riemannian Diffusions

Let E be a compact Riemannian manifold. As usual, $\langle \cdot , \cdot \rangle$, $\nabla \cdot$ and $\triangle \cdot$ will denote the scalar product, the gradient and the Laplacian associated with this structure. Let $\mathcal{A}$ be the algebra of smooth functions, i.e. $\mathcal{A} = C^\infty(E)$. Suppose that we are given a vector field b, we denote

$$\begin{aligned} L_0 : \mathcal{A} &\rightarrow \mathcal{A} \\ \varphi &\mapsto \frac{\triangle \varphi}{2} + \langle b, \nabla \varphi \rangle \end{aligned}$$

It is immediate to realize that in this example the carré du champ does not depend on b and satisfy

$$\forall\, f, g \in \mathcal{A}, \qquad \Gamma_0(f, f) = \langle \nabla f, \nabla g \rangle$$

(by the way, this equality gave the name "carré du champs").

The existence and uniqueness assumption for the associated martingale problem is well known to be fulfilled. We calculate that for $t \geq 0$, L_t is obtained from L_0 by a change of drift:

$$\forall\, \varphi \in \mathcal{A}, \qquad L_t(\varphi) = \frac{\triangle \varphi}{2} + \langle b - \nabla h_t, \nabla \varphi \rangle$$

This example is also a typical one where all the assumptions of section 3.1.2 are verified.

• Euclidean Diffusions

Except for the compactness of the state space, these processes are similar to those of the previous example.

So here $E = \mathbb{R}^n$, $n \geq 1$, and let for $x \in E$,

$$\sigma(x) = (\sigma^{i,j}(x))_{1 \leq i,j \leq n} \quad \text{and} \quad b(x) = (b^i(x))_{1 \leq i \leq n}$$

be respectively a symmetric nonnegative definite matrix and a n-vector. We suppose they are uniformly Lipschitz in their dependence on $x \in \mathbb{R}^n$.

Then denoting $a = \sigma^2$, let us consider on $\mathcal{A} \stackrel{\text{def.}}{=} C_{\mathrm{b}}^2(\mathbb{R}^n)$ the pregenerator

$$\forall\, \varphi \in \mathcal{A}, \forall\, x \in \mathbb{R}^n, \quad L_0(\varphi)(x) = \sum_{1 \leq i,j \leq n} \frac{a^{i,j}}{2}(x) \partial_{i,j} \varphi(x) + \sum_{1 \leq i \leq n} b^i(x) \partial_i \varphi(x)$$

It is a classical result that the associated martingale problems are well-posed (for further details about this problem, see [105]), and more precisely, for all $x \in \mathbb{R}^n$, $\widetilde{\mathbb{P}}_{\delta_x}$ is the law of the (unique strong) solution of the stochastic differential equation

$$\begin{cases} S_0 &= x \\ dS_t &= \sigma(S_t)\,dB_t + b(S_t)\,dt \qquad ; t \geq 0 \end{cases}$$

where $(B_t)_{t\geq 0}$ is a standard n-vector Brownian motion.

Here the carré du champ is given by

$$\forall\, \phi, \varphi \in \mathcal{A}, \qquad \Gamma_0(\phi,\varphi) \;=\; \sum_{1\leq i,j\leq n} a^{i,j}(x)\partial_i\phi(x)\partial_j\varphi(x)$$

so we find that

$\forall\, t \geq 0, \forall\, \varphi \in \mathcal{A}, \forall\, x \in \mathbb{R}^n$,

$$L_t(\varphi)(x) \;=\; \sum_{1\leq i,j\leq n} \frac{a^{i,j}}{2}(x)\partial_{i,j}\varphi(x) + \sum_{1\leq i\leq n}(b^i(x) - \sum_{1\leq j\leq n} a^{i,j}(x)\partial_j h_t(x))\partial_i\varphi(x)$$

Let us make the hypothesis that a is uniformly elliptic: there exists a constant $\epsilon > 0$ such that for all $x \in \mathbb{R}^n$,

$$\forall\, z = (z_i)_{1\leq i\leq n} \in \mathbb{R}^n, \qquad \sum_{1\leq i,j\leq n} a^{i,j}(t,x)z_iz_j \;\geq\; \epsilon \sum_{1\leq i\leq n} z_i^2$$

Under the extra assumption that a and b are C_{b}^{∞}, we see that all the requirement of section 3.1.2 are met (rather taking there $\mathcal{A} = C_{\mathrm{b}}^{\infty}(\mathbb{R}^n)$). But considering the parabolic equation satisfied by $F_{T,\varphi}$, it appears that (H2) will be verified under much less regularity for the coefficients a and b.

4.3 Time Discretization of Continuous Time Filtering Problems

In this section we discuss a time discretization approximating model for the non linear filtering problem associated to the previous Euclidean diffusion signal. To clarify the presentation all processes considered in this section will be indexed on the compact interval $[0,1] \subset \mathbb{R}_+$.

The basic model for the continuous time filtering problem considered here consists in an $\mathbb{R}^p \times \mathbb{R}^q$-valued Markov process $\{(X_t, Y_t) \;:\; t \in [0,1]\}$, strong solution on a probability space (Ω, F, P) of the Itô's type stochastic differential equations

$$\begin{cases} dX_t &= a(X_t)dt + b(X_t)d\beta_t \\ dY_t &= h(X_t)dt + dV_t \end{cases}$$

where

1. $a : \mathbb{R}^p \to \mathbb{R}^p$, $b : \mathbb{R}^p \to \mathbb{R}^{p\times m}$ and $h : \mathbb{R}^p \to \mathbb{R}^q$ are bounded and Lipschitz continuous functions.

2. $\{(\beta_t, V_t) \; : \; t \in [0,1]\}$ is a $(\mathbb{R}^m \times \mathbb{R}^q)$-valued standard Brownian motion.

3. $Y_0 = 0$ and X_0 is a random variable independent of $\{(\beta_t, V_t) \; : \; t \in [0,1]\}$ with law ν so that $E(|X_0|^2) < \infty$.

The classical filtering problem is to find the conditional distribution of the signal X at time t with respect to the observations Y up to time t, that is

$$\forall \; f \in \mathcal{B}_b(\mathbb{R}^p), \qquad \pi_t f = \mathbb{E}(f(X_t)|\mathcal{Y}_{[0,t]}) \tag{116}$$

equation where $\mathcal{Y}_{[0,t]}$ is the filtration generated by the observations Y up to time t.

The first step in this direction consists in obtaining a more tractable description of the conditional expectations (116).
Introducing $Z_t > 0$ such that

$$\forall t \in [0,1], \qquad \log Z_t = \int_0^t h^\star(X_s)\, dY_s - \frac{1}{2} \int_0^t |h(X_s)|^2 \, ds$$

it is well known that the original probability measure P is equivalent to a so called *reference probability measure* P_0 given by

$$P = Z_1 \, P_0.$$

In addition, under P_0, $\{(\beta_t, Y_t) \; : \; t \in [0,1]\}$ is a $(\mathbb{R}^m \times \mathbb{R}^q)$-valued standard Brownian motion and, X_0 is a random variable with law ν, independent of (β, Y).
The following well known result gives a functional integral representation for the conditional expectations (116), which is known as the Kallianpur-Striebel formula:

$$\forall \; f \in \mathcal{B}_b(\mathbb{R}^p), \, \forall \; t \in [0,1], \quad \pi_t f \;\; = \;\; \frac{\mathbb{E}_0(f(X_t) \; Z_t|\mathcal{Y}_{[0,t]})}{\mathbb{E}_0(Z_t|\mathcal{Y}_{[0,t]})} = \frac{\mathbb{E}_0^Y(f(X_t) \; Z_t|\mathcal{Y}_{[0,t]})}{\mathbb{E}_0^Y(Z_t|\mathcal{Y}_{[0,t]})} \tag{117}$$

We use $\mathbb{E}_0^Y(.)$ to denote the integration with respect to the Brownian paths $\{\beta_t \; : \; t \in [0,1]\}$ and the variable X_0.

In this section a program for the numerical solving of (117) by using a discrete time IPS scheme is embarked on. As announced in the introduction such IPS approach is obtained by first approximating the original model by a discrete time and measure valued process. The treatment that follows is standard in nonlinear filtering literature and it is essentially contained in [75, 76] and [92].

The former discrete time approximating model of (117) is obtained by first introducing a time discretization scheme of the basic model. To this end we introduce a sequence a meshes $\{(t_0, \ldots, t_M) \; : \; M \geq 1\}$ given by

$$t_n = \frac{n}{M} \qquad n \in \{0, \ldots, M\}.$$

To obtain a computationally feasible solution we will also use the following natural assumptions:

- For any $M \geq 1$ there exists a transition probability kernel $P^{(M)}$ such that

$$\sup_{t\in[0,1]} \mathbb{E}\left(|X_{t_n} - X_{t_n}^{(M)}|^2\right) \leq \frac{K}{M}, \qquad K < \infty$$

  where $\{X_{t_n}^{(M)} \ : \ n = 0, \ldots, M\}$ is the time homogeneous Markov chain with transition probability kernel $P^{(M)}$ and such that $X_0^{(M)} = X_0$.

- We can exactly simulate random variables according to the law $P^{(M)}(x, .)$ for any $x \in \mathbb{R}^p$.

It is worth noting that an example of an approximating Markov chain

$$\{X_{t_n}^{(M)} \ : \ n = 0, \ldots, M\}$$

satisfying these assumptions is given by the classical Euler scheme

$$X_{t_n}^{(M)} = X_{t_{n-1}}^{(M)} + a(X_{t_{n-1}}^{(M)})(t_n - t_{n-1}) + c(X_{t_{n-1}}^{(M)})\,(\beta_{t_n} - \beta_{t_{n-1}}) \qquad n = 1, \ldots, M \quad (118)$$

with $X_0^{(M)} = X_0$. This scheme is the crudest of the discretization scheme that can be used in our settings. Other time discretization schemes for diffusive signals are described in full detail in [91] and [108]. In view of (117) the optimal filters $\{\pi_{t_n} \ : \ n = 0, \ldots, M\}$ can be written as

$$\pi_{t_n} f = \frac{\mathbb{E}_0^Y\left(Z_{t_{n-1}}\ (H_{t_n} f)(X_{t_{n-1}})\right)}{\mathbb{E}_0^Y\left(Z_{t_{n-1}}\ \ (H_{t_n} 1)(X_{t_{n-1}})\right)}$$

where H_{t_n} is the finite transition measure on $\mathbb{R}^p$ given by

$$\begin{aligned} H_{t_n} f(x) &\stackrel{\text{def}}{=} \int H_{t_n}(x, dz)\ f(z) = E_0^Y\left(f(X_{t_n})\ g_{t_n}(X, Y)/X_{t_{n-1}} = x\right) \\ \log g_{t_n}(X, Y) &= \int_{t_{n-1}}^{t_n} h^\star(X_s)\, dY_s - \frac{1}{2}\int_{t_{n-1}}^{t_n} |h(X_s)|^2\, ds. \end{aligned} \qquad (119)$$

If, for any transition measure H and any probability measure π on $\mathbb{R}^p$ we denote by πH the finite measure so that for any bounded continuous function $f \in \mathcal{B}_b(\mathbb{R}^p)$,

$$\pi H(f) = \int \pi(dx)\ (Hf)(x),$$

then, given the observations, the dynamics structure of the conditional distributions $\{\pi_{t_n} \ : \ n = 0, \ldots, M\}$ is defined by the recursion

$$\pi_{t_n}(f) = \frac{\pi_{t_{n-1}} H_{t_n}(f)}{\pi_{t_{n-1}} H_{t_n}(1)}, \qquad n = 1, \ldots, M \qquad \text{with} \quad \pi_{t_0} = \nu$$

To approximate the stochastic integrals (119) it is convenient to note that, in a sense to be given,

$$\log g_{t_n}(X,Y) \underset{\Delta t_n \sim 0}{\sim} h^\star(X_{t_{n-1}})\,\Delta Y_{t_n} - \frac{1}{2}|h(X_{t_{n-1}})|^2\,\Delta t_n$$

with $\Delta t_n = t_n - t_{n-1}$. In this connection, a first step to obtain a computationally feasible solution consists in replacing H_{t_n} by the approximating multiplication operator

$$(H^M_{t_n} f)(x) = g^M_{t_n}(\Delta Y_{t_n}, x)\,(P^{(M)} f)(x)$$

where $g^M_{t_n}(\Delta Y_{t_n}, .) : \mathbb{R}^p \to \mathbb{R}_+$ is the positive and continuous function given by

$$g^M_{t_n}(\Delta Y_{t_n}, x) = \exp\left(h^\star(x)\,\Delta Y_{t_n} - \frac{1}{2M}|h(x)|^2\right) \quad \text{with} \quad \Delta Y_{t_n} = Y_{t_n} - Y_{t_{n-1}}$$

Remark 4.5: The choice of the approximating function $g^M_{t_n}$ given above is not unique. We can also use the functions $\tilde{g}^M_{t_n}$ given by

$$\tilde{g}^M_{t_n}(\Delta Y_{t_n}, x) = 1 + h^\star(x)\Delta Y_{t_n} + \frac{1}{2}|h(x)|^2\left(|\Delta Y_{t_n}|^2 - \frac{1}{M}\right)$$

The function h being bounded we can choose M large enough so that

$$\forall\, x \in \mathbb{R}^p, \qquad \|h\| < \sqrt{M} \qquad \text{and} \qquad g^M_{t_n}(\Delta Y_{t_n}, x) > 0$$

From now on we denote by $\{\pi^M_{t_n} \;:\; n = 0, \ldots, M\}$ the solution of the resulting approximating discrete time model

$$\left\{\begin{array}{lcl} \pi^M_{t_n} & = & \Phi^M_n(\Delta Y_{t_n}, \pi^M_{t_{n-1}}), \qquad n = 1, \ldots, M \\ \pi^M_0 & = & \nu \end{array}\right. \tag{120}$$

where $\Phi^M_n(y,\pi) = \Psi^M_n(y,\pi) P^{(M)}$ and

$$\forall\, f \in \mathcal{B}_{\mathrm{b}}(E), \qquad \Psi^M_{t_n}(y,\pi) f = \frac{\displaystyle\int f(x)\, g^M_{t_n}(y,x)\,\pi(dx)}{\displaystyle\int g^M_{t_n}(y,x)\,\pi(dx)}$$

for any $y \in \mathbb{R}^q$ and $\pi \in \mathbf{M}_1(\mathbb{R}^p)$.

Elementary manipulations show that the solution of the latter system is also given by the formula

$$\pi^M_{t_n} f = \frac{\displaystyle\int f(x_n) \prod_{m=1}^n g^M_{t_m}(\Delta Y_{t_m}, x_{m-1}) \prod_{m=1}^n P^{(M)}(x_{m-1}, dx_m)\,\nu(dx_0)}{\displaystyle\int \prod_{m=1}^n g^M_{t_m}(\Delta Y_{t_m}, x_{m-1}) \prod_{m=1}^n P^{(M)}(x_{m-1}, dx_m)\,\nu(dx_0)} \tag{121}$$

This gives a description of a discrete time approximating model for the optimal filter in terms of Feynman-Kac formulae for measure valued systems as those presented in section 1.1 and section 1.3.1.

The error bound caused by the discretization of the time interval $[0,1]$ and the approximation of the signal semigroup is well understood (see for instance Proposition 5.2 p. 31 [75], Theorem 2 in [92], Theorem 4.2 in [76] and also Theorem 4.1 in [82]). More precisely if for any $n = 0,\ldots,M-1$ and $t \in [t_n, t_{n+1})$ we denote by $\pi_t^M \stackrel{\text{def.}}{=} \pi_{t_n}^M$ we have the well known result.

Theorem 4.6 ([76]) *Let f be a bounded test function on $\mathbb{R}^p$ satisfying the Lipschitz condition*

$$|f(x) - f(z)| \leq k(f)\, |x - z|.$$

Then

$$\sup_{t\in[0,1)} \mathbb{E}\left(|\pi_t f - \pi_t^M f|\right) \leq \frac{\mathrm{C}}{\sqrt{M}} \left(\|f\| + k(f)\right) \tag{122}$$

where C is some finite constant.

We shall see in section 4.4 that the discrete time approximating model (120) can be regarded as the optimal filter of a suitably defined discrete time filtering problem. In most of the applications we have in mind the whole path of observation process $\{Y_t\ ;\ t \in [0,1]\}$ is not completely known.
Instead of that the acquisition of the observation data is made at regularly spaced times. In this specific situation the approximating model (120) and the sampled observation record $\{\Delta Y_{t_n}\ ;\ n = 1,\ldots,M\}$ give a natural framework for formulating this filtering problem and for applying the BIPS approaches developed in previous sections.

By $\{\xi_{t_n}\ ;\ n \geq 0\}$ we denote the N interacting particle scheme associated with the limiting system (120) and defined as in (13) by replacing the functions $\{\Phi_n\ ;\ n \geq 1\}$ by the functions $\{\Phi_n^{(M)}(\Delta Y_{t_n}, .)\ ;\ n \geq 1\}$.
The results of section 2.2 can be used to study the convergence of the random measures

$$\pi_{t_n}^{M,N} \stackrel{\text{def.}}{=} \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_{t_n}^i}$$

to the flow of distributions $\{\pi_{t_n}^M\ ;\ n \geq 0\}$ as $N \to \infty$.

An immediate question is to know how the discrete time N-particle scheme and the M-discretization time scheme combine? This study is still in progress. The only known result in this direction has been obtained in [25].

For any $n = 0,\ldots,M-1$ and $t \in [t_n, t_{n+1})$ we denote by $\pi_t^{M,N}$, the empirical measures associated with the system ξ_{t_n}, namely

$$\pi_t^{M,N} = \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_{t_n}^i}.$$

Theorem 4.7 *For any bounded Lipschitz test function f such that*

$$|f(x) - f(z)| \leq k(f)\ |x - z|$$

we have that

$$\sup_{t\in[0,1)} \mathbb{E}\left(|\pi_t f - \pi_t^{M,N} f|\right) \leq \frac{C_1}{\sqrt{M}}\ (\|f\| + k(f)) + C_2\ \sqrt{\frac{M}{N}}\ \|f\| \tag{123}$$

where C_1 is the finite constant appeared in Theorem 4.6 and $C_2 = 2\sqrt{2}e^{12\|h\|^2}$. In addition, if $p = q = 1$ and a, b, f, h are four times continuously differentiable with bounded derivatives then we have

$$\sup_{t\in[0,1)} \mathbb{E}\left(|\pi_t f - \pi_t^{M,N} f|\right) \leq \mathrm{Cte}\left(\frac{1}{M} + \sqrt{\frac{M}{N}}\right). \tag{124}$$

4.4 Discrete Time Filtering Problems

4.4.1 Description of the Models

The discrete time filtering problem consists in a signal process $X = (X_n\ ;\ n \geq 0)$ taking values in a Polish space E and an "observation" process $Y = (Y_n\ ;\ n \geq 1)$ taking values in $\mathbb{R}^d$ for some $d \geq 1$. We assume that the transition probability kernels $\{K_n\ ;\ n \geq 1\}$ are Feller and the initial value X_0 of the signal is an E-valued random variable with law $\eta_0 \in \mathbf{M}_1(E)$. The observation process has the form

$$\forall n \geq 1\,, \qquad Y_n = h_n(X_{n-1}) + V_n$$

where $h_n : E \to \mathbb{R}^d$ are bounded continuous and $(V_n\ ;\ n \geq 0)$ are independent random variables with positive continuous density $(\varphi_n\ ;\ n \geq 0)$ with respect to Lebesgue measure on $\mathbb{R}^d$. It is furthermore assumed that the observation noise $(V_n\ ;\ n \geq 0)$ and the signal $(X_n\ ;\ n \geq 0)$ are independent.

The filtering problem can be summarized as to find the conditional distributions

$$\forall\ f \in \mathcal{C}_b(E), \forall\ n \geq 1, \qquad \eta_n(f) \ = \ \mathbb{E}(f(X_n)/Y_1, \ldots, Y_n)$$

A version of η_n is given by a Feynman-Kac formula as the one presented in (1), namely

$$\eta_n(f) = \frac{\displaystyle\int f(x_n)\ \prod_{m=1}^{n} \varphi_m\left(Y_m - h_m(x_{m-1})\right)\ \prod_{m=1}^{n} K_m(x_{m-1}, dx_m)\ \eta_0(dx_0)}{\displaystyle\int \prod_{m=1}^{n} \varphi_m\left(Y_m - h_m(x_{m-1})\right)\ \prod_{m=1}^{n} K_m(x_{m-1}, dx_m)\ \eta_0(dx_0)} \tag{125}$$

It is transparent from this formulation that the discrete time approximating model (120) given in section 4.3 can be regarded as the optimal filter associated with a discrete time nonlinear filtering problem.

Given the observations $\{y_n\ ;\ n \geq 1\}$ the flow of distributions $\{\eta_n\ ;\ n \geq 0\}$ is again solution of a $\mathbf{M}_1(E)$-valued dynamical system of the form (8), that is

$$\forall\ n \geq 1, \forall\ \eta_0 \in \mathbf{M}_1(E), \qquad \eta_n = \Phi_n(y_n, \eta_{n-1}) \tag{126}$$

where for any $y \in \mathbb{R}^d$, $\Phi_n(y, .) : \mathbf{M}_1(E) \to \mathbf{M}_1(E)$ is the continuous function given by

$$\forall\ \eta \in \mathbf{M}_1(E), \qquad \Phi_n(y, \eta) \stackrel{\text{def.}}{=} \Psi_n(y_n, \eta)\ K_n$$

and $\Psi_n(y, .) : \mathbf{M}_1(E) \to \mathbf{M}_1(E)$ is the continuous function given by

$$\forall\ \eta \in \mathbf{M}_1(E), \forall\ f \in \mathcal{C}_b(E), \qquad \Psi_n(y, \eta)(f) = \frac{\int f(x)\ \varphi_n(y - h_n(x))\ \eta(dx)}{\int \varphi_n(y - h_n(z))\ \eta(dz)}$$

In this formulation the flow of distributions $\{\eta_n\ ;\ n \geq 0\}$ is parameterized by a given observation record $\{y_n\ :\ n \geq 1\}$ and it is solution of the measure valued dynamical system having the form (8) so that the IPS and BIPS approaches introduced in section 1.3.2 and section 2.3 can be applied.

4.4.2 Averaged Results

Our next objective is to present averaged versions of stability results given in section 2.1.2 and the averaged version of Theorem 2.11.

The only difficulty in directly applying the results of the end of section 2.1.2 stems from the fact that in our setting the fitness functions are random in the observation parameter. Instead of $(\mathcal{G})$ we will use the following assumption

$(\mathcal{G}')$ ***For any time*** $n \geq 1$***, there exist a positive function***

$$a_n : \mathbb{R}^d \to [1, \infty)$$

and a nondecreasing function $\theta : \mathbb{R} \to \mathbb{R}$ ***such that***

$$\forall\ x \in E, \forall\ y \in \mathbb{R}^d, \qquad \frac{1}{a_n(y)} \leq \frac{\varphi_n(y - h_n(x))}{\varphi_n(y)} \leq a_n(y) \tag{127}$$

and

$$|\log a_n(y+u) - \log a_n(y)| \leq \theta(\|u\|)$$

Theorem 4.8 *Assume that* $(\mathcal{G}')$ *holds with* $\sup_{n\geq 1} \|h_n\| < \infty$*. For any* $\mu \in \mathbf{M}_1(E)$ *we write* $\{\eta_n^\mu\ ;\ n \geq 0\}$ *the solution of the nonlinear filtering equation (126) starting at* μ*.*
If $(\mathcal{K})_2$ *holds then for any* $\mu, \nu \in \mathbf{M}_1(E)$ *we have the following implication*

$$\sup_{n\geq 1} \mathbb{E}(\log a_n(V_n)) < \infty \Longrightarrow \lim_{n\to\infty} \mathbb{E}(\|\eta_n^\mu - \eta_n^\nu\|_{\text{tv}}) = 0$$

If $(\mathcal{K})_3$ *holds then we also have for any* $\mu, \nu \in \mathbf{M}_1(E)$

$$\sum_{n\geq 1} \mathbb{E}(a_n^{-2}(V_n)) = \infty \quad \Longrightarrow \quad \lim_{n\to\infty} \mathbb{E}(\|\eta_n^\mu - \eta_n^\nu\|_{\text{tv}}) = 0$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{p=1}^{n} \mathbb{E}(a_p^{-2}(V_p)) > 0 \quad \Longrightarrow \quad \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}(\|\eta_n^\mu - \eta_n^\nu\|_{\text{tv}}) < 0$$

The averaged version of Theorem 2.11 can be stated as follows

Theorem 4.9 *Let $\mathcal{F}$ be a countable collection of functions f such that $\|f\| \leq 1$ and satisfying the entropy condition $I(\mathcal{F}) < \infty$. Assume that $(\mathcal{G})'$ holds and the following conditions are met*

$$\sup_{n\geq 1} \log \mathbb{E}(a_n^2(V_n))^{1/2} \stackrel{\text{def}}{=} L < \infty, \qquad \sup_{n\geq 1} \|h_n\| \stackrel{\text{def}}{=} M < \infty$$

Assume moreover that the nonlinear filtering equation (126) is asymptotically stable in the sense that,

$$\lim_{T\to\infty} \sup_{\mu,\nu\in\mathbf{M}_1(E)} \sup_{p\geq 0} \mathbb{E}\left(\|\Phi_{p,p+T}(\mu) - \Phi_{p,p+T}(\nu)\|_{\mathcal{F}} \,|Y_1,\ldots,Y_p\right) = 0$$

then we have the following uniform convergence with respect to time

$$\lim_{N\to\infty} \sup_{n\geq 0} \mathbb{E}\left(\left\|\eta_n^N - \eta_n\right\|_{\mathcal{F}}\right) = 0 \tag{128}$$

In addition, if the evolution equation (126) is exponentially asymptotically stable in the sense that there exists some positive constant $\gamma > 0$ such that for any $\mu, \nu \in \mathbf{M}_1(E)$ and $T \geq 0$

$$\sup_{p\geq 0} \mathbb{E}\left(\|\Phi_{p,p+T}(\mu) - \Phi_{p,p+T}(\nu)\|_{\mathcal{F}} \,|Y_1,\ldots,Y_p\right) \leq e^{-\gamma T}$$

then we have for any $p \geq 1$, the uniform $\mathbb{L}^p$-error bound given by

$$\sup_{n\geq 0} \mathbb{E}\left(\left\|\eta_n^N - \eta_n\right\|_{\mathcal{F}}^p\right)^{\frac{1}{p}} \leq \frac{C_p\ e^{\gamma'}}{N^{\frac{\alpha}{2}}}\ \ I(\mathcal{F})$$

where C_p is a universal constant which only depends on $p \geq 1$ and α and γ' are given by

$$\alpha = \frac{\gamma}{\gamma + \gamma'} \quad \text{and} \quad \gamma' = 1 + 2\,(L + \theta(M))$$

Remark 4.10: We now present a class of discrete time nonlinear filtering problems for which the BIPS approaches developed in this work do not apply.
Let us assume that the pair process (X, Y) takes values in $\mathbb{R}^p \times \mathbb{R}^{p'}$ and evolves according to the following Itô's differential equations

$$\left\{\begin{array}{rcl} dX_t &=& a(X_t, Y_t)\, dt + b(X_t, Y_t)\, dW_t \\ dY_t &=& a'(X_t, Y_t)\, dt + b'(X_t, Y_t)\, dW_t' \qquad Y_0 = 0 \end{array}\right.$$

where a, b, a', b' are known functions suitably defined and $(W, W'))$ is a $p{+}p'$-dimensional Wiener process. Suppose moreover that the acquisition of the observations is only made at times $n \in \mathbb{N}$ and we want to compute for all reasonable functions $f : \mathbb{R}^p \to \mathbb{R}$

$$\pi_n f = \mathbb{E}\left(f(X_n)\,|Y_1,\ldots,Y_n\right)$$

This problem is clearly a discrete time nonlinear filtering problem but the BIPS approaches developed in previous sections do not apply.

A novel BIPS strategy has been proposed in [41] to solve this problem. In contrast to the latter this new particle scheme consists in N-pair particles and the mutation transition is related to the continuous semigroup of the pair process $\{(X_t, Y_t)\ ;\ t \geq 0\}$.

Exponential rates of convergence and L^1-mean error estimates are given in [41] and central limit theorems for the particle density profiles are presented in [40, 39] but many questions such as large deviations, fluctuations on path space as well as uniform convergence results with respect to time remain unsolved.

5 Appendix and Index of Notations

Since we have tried to use similar notations for discrete and continuous time, we hope that the following separations of the indexes for both settings will be convenient for the reader.

Index of Symbols (discrete time)

Index of Conditions for Asymptotic Theorems (discrete time)

Index of Symbols (continuous time)

Index of Conditions for Asymptotic Theorems (continuous time)

Examples

Despite our best efforts, the variety of approaches to study asymptotic stability properties and limit theorems for the IPS approximating models inevitably require a set of specific assumptions which might be confusing at a first reading. We therefore gather together in this section a short discussion on the conditions we have defined and a succinct series of implications. As a guide to their usage we also analyze these assumptions in two academic examples, namely the Gaussian and bi-exponential transitions $\{K_n \,;\, n \geq 1\}$.

It is also worth observing immediatly that the limit theorems for the particle density profiles as the number of particles tends to infinity only depend on a boundedness condition $(\mathcal{G})$ defined on page 21. Furthermore in nonlinear filtering settings the fitness functions $\{g_n \,;\, n \geq 1\}$ also depend on the observation process and the appropriate condition corresponding to $(\mathcal{G})$ is the assumption $(\mathcal{G})'$ defined on page 129. It is clear that the situation becomes more involved when dispensing with the assumption $(\mathcal{G})$ or $(\mathcal{G})'$. It turns out that several results can be proved without these conditions, see for instance Proposition 2.9 and its proof on page 35. Furthermore these boundedness conditions are not really restrictive and they are commonly used in nonlinear filtering literature. We conclude this section with a short collection of fitness functions and observation processes satisfying the former conditions.

Section 2.1.2 is concerned with the asymptotic stability properties of the limiting measure valued systems $\{\eta_n \,;\, n \geq 0\}$ and $\{\widehat{\eta}_n \,;\, n \geq 0\}$. Under appropriate mixing conditions on the transition probability kernels $\{K_n \,;\, n \geq 1\}$ it is proven that the resulting limiting systems forget exponentially fast their initial conditions. The mixing type conditions we employed are $(\mathcal{K})_\varepsilon$, $(\mathcal{K})_1$, $(\mathcal{K})_2$, $(\mathcal{K})_3$ and $(\mathcal{KG})$. Recalling their description given respectively on pages 20, 24, 26 and 27 it is easy to establish the following implications

$$\begin{array}{ccccc} & & (\mathcal{K})_3 & & \\ & & \Downarrow & & \\ (\mathcal{K})_\varepsilon & \Rightarrow \quad (\mathcal{K})_1 & \Rightarrow \quad (\mathcal{K})_2 & \qquad & (\mathcal{K})_2 + (\mathcal{G}) \Rightarrow (\mathcal{KG}) \end{array}$$

As we said previously, the limit theorems for the IPS approximating schemes presented in section 2.2 only require the assumption $(\mathcal{G})$ on the fitness functions $\{g_n \,;\, n \geq 1\}$ except for the central limit theorem and large deviations principles on path space (see for instance section 2.2.1 as well as section 2.2.4 and section 2.2.5). The weakest and preliminary condition $(\mathcal{K})_0$ employed in the study the convergence of the empirical measures on path space can be regarded as a mixing type condition. Recalling its description on page 32 we clearly have

$$(\mathcal{K})_2 \Longrightarrow (\mathcal{K})_0$$

Condition $(\mathcal{K})_0$ is also a natural and appropriate condition for using a reference product measure and express the law of the IPS as a simple mean field Gibbs measure on path space.

The only additional restriction we place in the study of the central limit theorem on path space is the exponential moment condition $(\mathcal{TCL})$ given on page 48. As we shall see in the further development this condition holds for many typical examples of signal transitions. It is also worth noting that

$$(\mathcal{K})_1 \Longrightarrow (\mathcal{TCL}) \Longrightarrow (\mathcal{K})_0$$

In section 2.2.5 we analyze large deviations for the IPS approximating models for general and abstract functions $\{\Phi_n \,;\, n \geq 1\}$. We have presented a number of simple criterion only involving these one step functions. When applied to Feynman-Kac type systems we have seen that the large deviation principles on path space only require the assumptions $(\mathcal{K})_1'$ and $(\mathcal{K})_1''$ defined on page 54 and page 56. It is also easy to see that

$$(\mathcal{K})_1' \Longrightarrow (\mathcal{K})_1'' \Longrightarrow (\mathcal{K})_0$$

In view of the previous remarks the fluctuations and deviations on path space rely on the existence of a reference probability measure satisfying $(\mathcal{K})_0$. One way to remove this assumption is to study the particle density profiles.

The analysis of fluctuations of the particle density profiles is based on the dynamics structure of the limiting system given in section 2.1.1 and on limit theorems for stochastic processes. These results only depend on the boundedness condition $(\mathcal{G})$.

The same condition is used in proving large deviations for the particle density profiles. More precisely we have employed condition $(\mathcal{G})$ to check an exponential tightness condition $(\mathcal{ET})$ (cf. page 57)

$$(\mathcal{G}) \Longrightarrow (\mathcal{ET})$$

Gaussian Transitions

Let us now investigate the chain of assumptions $(\mathcal{K})_1$, $(\mathcal{K})_2$, $(\mathcal{K})_3$, $(\mathcal{K})_1'$, $(\mathcal{K})_1''$, and $(\mathcal{TCL})$ through the following Gaussian example.

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 nonnegative matrix and $b_n : \mathbb{R}^m \to \mathbb{R}^m$ is a bounded continuous function. It is not difficult to check that $(\mathcal{K})_1'$, 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

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

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

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

From the previous observation it is also not difficult to check that condition $(\mathcal{TCL})$ is also satisfied.

Let us discuss conditions $(\mathcal{K})_1$ and $(\mathcal{K})_3$ in time homogeneous settings (that is $K_n = K$) and when $E = \mathbb{R}$ and $Q_n = 1$ and $b_n = b$.
If $b : \mathbb{R} \to \mathbb{R}$ is only a bounded function, then $(\mathcal{K})_1$ and $(\mathcal{K})_2$ do not hold. For instance let us suppose that $b : \mathbb{R} \to \mathbb{R}$ is a bounded $\mathbf{B}(E)$-measurable function such that $b(0) = 0$ and $b(1) = -1$. Then, hypothesis $(\mathcal{K})_1$ is not satisfied. Suppose K satisfies $(\mathcal{K})_1$. Clearly there exists an absolutely continuous probability measure with density p such that

$$\forall\, x, z \in \mathbb{R}, \qquad c^{-1}\, p(z) \;\le\; e^{-\frac{1}{2}(z-b(x))^2} \;\le\; c\, p(z)$$

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

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

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

Now we examine condition $(\mathcal{K})_3$. First we note that for any $|z| \le M$ where $M \ge 0$ is chosen so that $\|b\| \le M$ we have

$$\forall\, x \in E, \qquad \epsilon \;\le\; \frac{dK(x,.)}{d\lambda}(z) \;\le\; \frac{1}{\epsilon}$$

where

$$\lambda(dz) = \frac{1}{\sqrt{2\pi}}\; \exp -\frac{z^2}{2}\; dz \qquad \text{and} \qquad \log \epsilon = -2M^2$$

In other words and roughly speaking $(\mathcal{K})_1$ is satisfied on the compact set

$$\{z \in \mathbb{R}\,;\; |z| \le M\}.$$

Let us assume that the drift function b satisfies

$$\forall\, |x| \ge M, \qquad b(x) \;=\; b(\text{sign}(x) M)$$

In this situation it is not difficult to check that $(\mathcal{K})_3$ holds with

$$A = [-M, M] \qquad B_1 = (-\infty, -M) \quad B_2 = (M, +\infty)$$

and

$$\begin{aligned}\lambda_1 &= \delta_{-M}K, & \lambda_2 = \delta_{-M}K\\ \gamma_1(dz) &= \frac{1}{\sqrt{2\pi}}\,\exp -\frac{1}{2}(z-M)^2\,dz, & \gamma_2(dz) = \frac{1}{\sqrt{2\pi}}\,\exp -\frac{1}{2}(z+M)^2 dz.\end{aligned}$$

Let us examine a Gaussian situation where $(\mathcal{K})_1''$ is not met. Again we suppose that $E = \mathbb{R}$ 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 $\epsilon_n : \mathbb{R} \to \mathbb{R}$ is a continuous function such that

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

It is not difficult to see that K_n is Feller. On the other hand, let us assume that K_n satisfies $(\mathcal{K})_1''$ for some function φ. Since $K_n(x,.)$ is absolutely continuous with respect to Lebesgue measure for any $x \in \mathbb{R}$, the probability measure λ_n described in $(\mathcal{K})_1''$ is absolutely continuous with respect to Lebesgue measure. Therefore, there exists a probability density p_n such that

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

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

Our study is not restricted to nonlinear filtering problem with Gaussian transitions K_n or with observations corrupted by Gaussian perturbations. We now present another kind of densities that can be handled in our framework.

Bi-exponential Transitions

Suppose $E = \mathbb{R}$ and K_n, $n \ge 1$, are given by

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

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

$$X_n = b_n(X_{n-1}) + W_n \qquad n \ge 1$$

where $(W_n)_{n\ge 1}$ is a sequence of real valued and independent random variables with bilateral exponential densities. Note that K_n may be written

$$K_n(x,dz) = \frac{1}{2}\,\alpha_n \exp\left(\alpha_n(|z| - |z - b_n(x)|)\right)\,\lambda_n(dz)$$

with

$$\lambda_n(dz) = \frac{1}{2}\,\alpha_n \exp\left(-\alpha_n|z|\right)\,dz$$

It follows that $(\mathcal{K})_1''$ holds since $|\log \frac{K_n(x,.)}{d\lambda_n}(z)|$ has Lipschitz norm $2\alpha_n + \alpha_n\|b_n\|$.

It is also clear that $(\mathcal{K})_1$ and $(\mathcal{TCL})$ hold since we have in this situation

$$\exp\left(-\alpha_n\|b_n\|\right) \le \frac{K_n(x,.)}{d\lambda_n}(z) \le \exp\left(\alpha_n\|b_n\|\right)$$

Conditions $(\mathcal{G})$ and $(\mathcal{G})'$

Next we examine condition $(\mathcal{G})'$.
As a typical example of nonlinear filtering problem assume the functions $h_n : E \to \mathbb{R}^d$, $n \geq 1$, are bounded continuous and the densities φ_n given by

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

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

$$\forall \, n \geq 1, \qquad Y_n \;=\; h_n(X_{n-1}) + V_n \tag{129}$$

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 that $(\mathcal{G})'$ holds with

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

where $\|R_n^{-1}\|$ is the spectral radius of R_n^{-1}. In addition we have

$$|\log a_n(y+u) - \log a_n(y)| \leq L_n \, |u| \qquad \text{with} \quad L_n = \|R_n^{-1}\| \, \|h_n\|$$

It is therefore not difficult to check that the assumptions of Theorem 4.9 is satisfied when

$$\sup_{n\geq 1}(\|h_n\|, \|R_n^{-1}\|) < \infty$$

In this situation it is also clear that the conditions of Theorem 4.8 and Theorem 4.9 are met. To see this claim it suffices to note that Jensen's inequality yields that

$$\log E(a_n^{-2}(V_n)) \geq -\|R_n^{-1}\| \, \|h_n\|^2 - 2\|R_n^{-1}\| \, \|h_n\| \, E(|V_n|)$$

and we also have

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

Our result is not restricted to Gaussian noise sources. For instance, let us assume that $d = 1$ and φ_n is a bilateral exponential density

$$\varphi_n(v) = \frac{\alpha_n}{2} \, \exp -(\alpha_n |v|) \qquad \alpha_n > 0$$

In this case one gets that $(\mathcal{G})'$ holds with

$$\log a_n(y) = \alpha_n \, \|h_n\|$$

which is independent of the observation parameter y. One concludes easily that the conditions of Theorem 4.8 and Theorem 4.9 are satisfied as soon as

$$\sup_{n\geq 0} \{\alpha_n, \|h_n\|\} < \infty$$

On the other hand if $\sum_{n\geq 0}\alpha_n\|h_n\| < \infty$ then condition $(\mathcal{KG})$ is satisfied and Theorem 2.7 can be used to study the asymptotic stability of the nonlinear filtering equation for any strongly ergodic signal process (cf. remark 2.8).
We end this section with an example of Cauchy noise sources. Suppose that $d = 1$ and φ_n is the density given by

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

In this situation one can check that

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

Thus, $(\mathcal{G})'$ holds with

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

References

[1] J. Abela, D. Abramson, A. De Silval, M. Krishnamoorthy, and G. Mills. Computing optimal schedules for landing aircraft. Technical report, Department of Computer Systems Eng. R.M.I.T., Melbourne, May 1993.

[2] J.-M. Alliot, D. Delahaye, J.-L. Farges, and M. Schoenauer. Genetic algorithms for automatic regrouping of air traffic control sectors. In J. R. McDonnell, R. G. Reynolds, and D. B. Fogel, editors, *Proceedings of the 4th Annual Conference on Evolutionary Programming*, pages 657–672. MIT Press, March 1995.

[3] G. Ben Arous and M. Brunaud. Methode de laplace : étude variationnelle des fluctuations de diffusions de type “champ moyen”. *Stochastics*, 31-32:79–144, 1990.

[4] R. Atar and O. Zeitouni. Exponential stability for nonlinear filtering. *Annales de l'Institut Henri Poincaré*, 33(6):697–725, 1997.

[5] R. Atar and O. Zeitouni. Lyapunov exponents for finite state space nonlinear filtering. *Society for Industrial and Applied Mathematics. Journal on Control and Optimization*, 35(1):36–55, January 1997.

[6] J. Baker. Adaptative selection methods for genetic algorithms. In J. Grefenstette, editor, *Proc. International Conf. on Genetic Algorithms and their Applications*. L. Erlbaum Associates, 1985.

[7] J. Baker. Reducing bias and inefficiency in the selection algorithm. In J. Grefenstette, editor, *Proc. of the Second International Conf. on Genetic Algorithms and their Applications*. L. Erlbaum Associates, 1987.

[8] A. Bakirtzis, S. Kazarlis, and V. Petridis. A genetic algorithm solution to the economic dispatch problem. http://www.dai.ed.ac.uk/groups/evalg /eag_local_copies_of_papers.body.html.

[9] D. Bakry. L'hypercontractivité et son utilisation en théorie des semigroupes. In P. Bernard, editor, *Lectures on Probability Theory. Ecole d'Eté de Probabilités de Saint-Flour XXII-1992*, Lecture Notes in Mathematics 1581. Springer-Verlag, 1994.

[10] P. Barbe and P. Bertail. *The Weighted Bootstrap.* Lecture Notes in Statistics 98. Springer-Verlag, 1995.

[11] E. Beadle and P. Djuric. A fast weighted Bayesian bootstrap filter for nonlinear model state estimation. *Institute of Electrical and Electronics Engineers. Transactions on Aerospace and Electronic Systems*, AES-33:338–343, January 1997.

[12] B.E. Benes. Exact finite-dimensional filters for certain diffusions with nonlinear drift. *Stochastics*, 5:65–92, 1981.

[13] E. Bolthausen. Laplace approximation for sums of independent random vectors i. *Probability Theory and Related Fields*, 72:305–318, 1986.

[14] C.L. Bridges and D.E. Goldberg. An analysis of reproduction and crossover in a binary-coded genetic algorithm. In J.J. Grefenstette, editor, *Proc. of the Second International Conf. on Genetic Algorithms and their Applications.* L. Erlbaum Associates, 1987.

[15] R.S. Bucy. *Lectures on discrete time filtering, Signal Processing and Digital Filtering.* Springer-Verlag, 1994.

[16] A. Budhiraja and D. Ocone. Exponential stability of discrete time filters for bounded observation noise. *Systems and Control Letters*, 30:185–193, 1997.

[17] Z. Cai, F. Le Gland, and H. Zhang. An adaptative local grid refinement method for nonlinear filtering. Technical report, INRIA, October 1995.

[18] H. Carvalho, P. Del Moral, A. Monin, and G. Salut. Optimal nonlinear filtering in gps/ins integration. *Institute of Electrical and Electronics Engineers. Transactions on Aerospace and Electronic Systems*, 33(3):835–850, July 1997.

[19] R. Cerf. *Une théorie asymptotique des algorithmes génétiques.* Thèse de doctorat, Université Montpellier II, March 1994.

[20] M. Chaleyat-Maurel and D. Michel. Des résultats de non existence de filtres de dimension finie. *Comptes Rendus de l'Académie des Sciences de Paris. Série I. Mathématique*, 296, 1983.

[21] D. Crisan, J. Gaines, and T.J. Lyons. A particle approximation of the solution of the kushner-stratonovitch equation. *Society for Industrial and Applied Mathematics. Journal on Applied Mathematics*, 58(5):1568–1590, 1998.

[22] D. Crisan and M. Grunwald. Large deviation comparison of branching algorithms versus resampling algorithm. Preprint, 1998.

[23] D. Crisan and T.J. Lyons. Nonlinear filtering and measure valued processes. *Probability Theory and Related Fields*, 109:217–244, 1997.

[24] D. Crisan, P. Del Moral, and T.J. Lyons. Interacting particle systems approximations of the Kushner-Stratonovitch equation. *Advances in Applied Probability*, 31(3), September 1999.

[25] D. Crisan, P. Del Moral, and T.J. Lyons. Non linear filtering using branching and interacting particle systems. *Markov Processes and Related Fields*, 5(3):293–319, 1999.

[26] G. Da Prato, M. Furhman, and P. Malliavin. Asymptotic ergodicity for the Zakai filtering equation. *Comptes Rendus de l'Académie des Sciences de Paris. Série I. Mathématique*, 321(5):613–616, 1995.

[27] E.B. Davies. *Heat Kernels and Spectral Theory.* Cambridge University Press, 1989.

[28] M. Davis. New approach to filtering for nonlinear systems. *Institute of Electrical and Electronics Engineers. Proceedings*, 128(5):166–172, 1981. Part D.

[29] D. Dawson. Measure-valued Markov processes. In P.L. Hennequin, editor, *Lectures on Probability Theory. Ecole d'Eté de Probabilités de Saint-Flour XXI-1991*, Lecture Notes in Mathematics 1541. Springer-Verlag, 1993.

[30] P. Del Moral. Non-linear filtering: interacting particle resolution. *Markov Processes and Related Fields*, 2(4):555–581, 1996.

[31] P. Del Moral. Filtrage non linéaire par systèmes de particules en interaction. *Comptes Rendus de l'Académie des Sciences de Paris. Série I. Mathématique*, 325:653–658, 1997.

[32] P. Del Moral. Maslov optimization theory: optimality versus randomness. In V.N. Kolokoltsov and V.P. Maslov, editors, *Idempotency Analysis and its Applications*, Mathematics and its Applications 401, pages 243–302. Kluwer Academic Publishers, Dordrecht/Boston/London, 1997.

[33] 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.

[34] P. Del Moral. A uniform convergence theorem for the numerical solving of non linear filtering problems. *Journal of Applied Probability*, 35:873–884, 1998.

[35] 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.

[36] P. Del Moral and A. Guionnet. On the stability of measure valued processes. Applications to non linear filtering and interacting particle systems. Publications du Laboratoire de Statistique et Probabilités, no 03-98, Université Paul Sabatier, 1998.

[37] P. Del Moral and A. Guionnet. A central limit theorem for non linear filtering using interacting particle systems. *The Annals of Applied Probability*, 9(2):275–297, 1999.

[38] P. Del Moral and A. Guionnet. On the stability of measure valued processes with applications to filtering. *Comptes Rendus de l'Académie des Sciences de Paris. Série I. Mathématique*, 329:429–434, 1999.

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

[40] P. Del Moral and J. Jacod. The monte-carlo method for filtering with discrete time observations. central limit theorems. Publications du Laboratoire de Probabilités, no 515, 1999.

[41] P. Del Moral, J. Jacod, and P. Protter. The Monte Carlo method for filtering with discrete time observations. Publications du Laboratoire de Probabilités, no 453, June 1998.

[42] P. Del Moral and M. Ledoux. Convergence of empirical processes for interacting particle systems with applications to nonlinear filtering. to appear in Journal of Theoretical Probability, January 2000.

[43] P. Del Moral and L. Miclo. Asymptotic stability of non linear semigroup of Feynman-Kac type. Préprint, publications du Laboratoire de Statistique et Probabilités, no 04-99, 1999.

[44] P. Del Moral and L. Miclo. On the convergence and the applications of the generalized simulated annealing. *SIAM Journal on Control and Optimization*, 37(4):1222–1250, 1999.

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

[46] P. Del Moral, J.C. Noyer, and G. Salut. Résolution particulaire et traitement non-linéaire du signal : application radar/sonar. In *Traitement du signal*, September 1995.

[47] P. Del Moral, G. Rigal, and G. Salut. Estimation et commande optimale non linéaire. Technical Report 2, LAAS/CNRS, March 1992. Contract D.R.E.T.-DIGILOG.

[48] B. Delyon and O. Zeitouni. Liapunov exponents for filtering problems. In M.H.A. Davis and R.J. Elliot, editors, *Applied Stochastic Analysis*, pages 511–521. Springer-Verlag, 1991.

[49] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Application.* Jones and Bartlett, 1993.

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

[51] J.-D. Deuschel and D.W. Stroock. *Large Deviations.* Pure and applied mathematics 137. Academic Press, 1989.

[52] P. Diaconis and B. Efron. Méthodes de calculs statistiques intensifs sur ordinateurs. *Pour la Science*, 1983. translation of the American Scientist.

[53] R.L. Dobrushin. Central limit theorem for nonstationnary Markov chains, i,ii. *Theory of Probability and its Applications*, 1(1 and 4):66–80 and 330–385, 1956.

[54] E.B. Dynkin and A. Mandelbaum. Symmetric statistics, Poisson processes and multiple Wiener integrals. *The Annals of Statistics*, 11:739–745, 1983.

[55] S. Ethier and T. Kurtz. *Markov Processes, Characterization and Convergence.* Wiley series in probability and mathematical statistics. John Wiley and Sons, New York, 1986.

[56] M. Fujisaki, G. Kallianpur, and H. Kunita. Stochastic differential equations for the non linear filtering problem. *Osaka J. Math.*, 1:19–40, 1972.

[57] F. Le Gland. Monte-Carlo methods in nonlinear filtering. In *Proceedings of the 23rd IEEE Conference on Decision and Control*, pages 31–32, Las Vegas, December 1984.

[58] F. Le Gland. High order time discretization of nonlinear filtering equations. In *28th IEEE CDC*, pages 2601–2606, Tampa, 1989.

[59] F. Le Gland, C. Musso, and N. Oudjane. An analysis of regularized interacting particle methods for nonlinear filtering. In *Proceedings of the 3rd IEEE European Workshop on Computer-Intensive Methods in Control and Signal Processing*, Prague, September 1998.

[60] D.E. Goldberg. Genetic algorithms and rule learning in dynamic control systems. In *Proceedings of the First International Conference on Genetic Algorithms*, pages 8–15, Hillsdale, NJ, 1985. L. Erlbaum Associates.

[61] D.E. Goldberg. Simple genetic algorithms and the minimal deceptive problem. In L. Davis, editor, *Genetic Algorithms and Simulated Annealing.* Pitman, 1987.

[62] D.E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison-Wesley, Reading, MA., 1989.

[63] D.E. Goldberg and P. Segrest. Finite Markov chain analysis of genetic algorithms. In J.J. Grefenstette, editor, *Proc. of the 2nd Int. Conf. on Genetic Algorithms.* L. Erlbaum Associates, 1987.

[64] N.J. Gordon, D.J. Salmon, and A.F.M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F*, 140:107–113, 1993.

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

[66] A. Guionnet. About precise Laplace's method; applications to fluctuations for mean field interacting particles. Preprint, 1997.

[67] J.H. Holland. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, 1975.

[68] J. Jacod and A.N. Shiryaev. *Limit Theorems for Stochastic Processes*. A Series of Comprehensive Studies in Mathematics 288. Springer-Verlag, 1987.

[69] J.M. Johnson and Y. Rahmat-Samii. Genetic algorithms in electromagnetics. In *IEEE Antennas and Propagation Society International Symposium Digest*, volume 2, pages 1480–1483, 1996.

[70] F. Jouve, L. Kallel, and M. Schoenauer. Mechanical inclusions identification by evolutionary computation. *European Journal of Finite Elements*, 5(5-6):619–648, 1996.

[71] F. Jouve, L. Kallel, and M. Schoenauer. Identification of mechanical inclusions. In D. Dagsgupta and Z. Michalewicz, editors, *Evolutionary Computation in Engeneering*, pages 477–494. Springer Verlag, 1997.

[72] G. Kallianpur and C. Striebel. Stochastic differential equations occuring in the estimation of continuous parameter stochastic processes. Tech. Rep. 103, Department of statistics, Univ. of Minnesota, September 1967.

[73] G. Kitagawa. Monte-Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal on Computational and Graphical Statistics*, 5(1):1–25, 1996.

[74] H. Korezlioglu. Computation of filters by sampling and quantization. Technical Report 208, Center for Stochastic Processes, University of North Carolina, 1987.

[75] H. Korezlioglu and G. Maziotto. Modelization and filtering of discrete systems and discrete approximation of continuous systems. In *Modélisation et Optimisation des Systèmes, VI Conférence INRIA*, Nice, 1983.

[76] H. Korezlioglu and W.J. Runggaldier. Filtering for nonlinear systems driven by nonwhite noises: an approximating scheme. *Stochastics and Stochastics Reports*, 44(1-2):65–102, 1993.

[77] H. Kunita. Asymptotic behavior of nonlinear filtering errors of Markov processes. *Journal of Multivariate Analysis*, 1(4):365–393, 1971.

[78] H. Kunita. Ergodic properties nonlinear filtering processes. In K.C. Alexander and J.C. Watkins, editors, *Spatial Stochastic Processes*. Birkhaüser Boston, Boston, MA, 1991.

[79] S. Kusuoda and Y . Tamura. Gibbs measures for mean field potentials. *J. Fac. Sci. Univ. Tokyo, Sect. IA, Math*, 31, 1984.

[80] J.W. Kwiatkowski. Algorithms for index tracking. Technical report, Department of Business Studies, The University of Edinburgh, 1991.

[81] R.S. Liptser and A.N. Shiryayev. *Theory of Martingales.* Dordrecht: Kluwer Academic Publishers, 1989.

[82] G.B. Di Masi, M. Pratelli, and W.G. Runggaldier. An approximation for the nonlinear filtering problem with error bounds. *Stochastics*, 14(4):247–271, 1985.

[83] 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.

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

[85] C. Musso and N. Oudjane. Regularized particle schemes applied to the tracking problem. In *International Radar Symposium, Munich, Proceedings*, September 1998.

[86] Y. Nishiyama. Some central limit theorems for l^∞-valued semimartingales and their applications. *Probability Theory and Related Fields*, 108:459–494, 1997.

[87] A. Nix and M.D. Vose. Modelling genetic algorithms with Markov chains. *Annals of Mathematics and Artificial Intelligence*, 5:79–88, 1991.

[88] D. Ocone and E Pardoux. Asymptotic stability of the optimal filter with respect to its initial condition. *Society for Industrial and Applied Mathematics. Journal on Control and Optimization*, 34:226–243, 1996.

[89] D.L. Ocone. *Topics in nonlinear filtering theory.* Phd thesis, MIT, Cambridge, 1980.

[90] E. Pardoux. Filtrage non linéaire et équations aux dérivés partielles stochastiques associées. In P.L. Hennequin, editor, *Ecole d'Eté de Probabilités de Saint-Flour XIX-1989*, Lecture Notes in Mathematics 1464. Springer-Verlag, 1991.

[91] E. Pardoux and D. Talay. Approximation and simulation of solutions of stochastic differential equations. *Acta Applicandae Mathematicae*, 3(1):23–47, 1985.

[92] J. Picard. Approximation of nonlinear filtering problems and order of convergence. In *Filtering and Control of Random Processes*, Lecture Notes Control and Inf. Sc. 61. Springer, 1984.

[93] J. Picard. An estimate of the error in time discretization of nonlinear filtering problems. In C.I. Byrnes and A. Lindquist, editors, *Proceedings of the 7th MTNS — Theory and Applications of nonlinear Control Systems, Stockholm, 1985*, pages 401–412, Amsterdam, 1986. North–Holland Pub.

[94] J. Picard. Nonlinear filtering of one-dimensional diffusions in the case of a high signal-to-noise ratio. *Society for Industrial and Applied Mathematics. Journal on Applied Mathematics*, 16:1098–1125, 1986.

[95] S.T. Rachev. *Probability Metrics and the Stability of Stochastic Models.* Wiley, New York, 1991.

[96] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion.* Springer-Verlag, 1991.

[97] J. Shapcott. Index tracking: genetic algorithms for investment portfolio selection. Technical Report SS92-24, EPCC, September 1992.

[98] T. Shiga and H. Tanaka. Central limit theorem for a system of markovian particles with mean field interaction. *Zeitschrift für Wahrscheinlichkeitstheorie verwandte Gebiete*, 69, 1985. 439–459.

[99] A.N. Shiryaev. On stochastic equations in the theory of conditional Markov processes. *Theor. Prob. Appl.*, 11:179–184, 1966.

[100] A.N. Shiryaev. *Probability.* Number 95 in Graduate Texts in Mathematics. Springer-Verlag, New-York, second edition, 1996.

[101] B. Simon. *Trace ideals and their applications.* London Mathematical Society Lecture Notes Series 35. Cambridge University Press, 1977.

[102] L. Stettner. On invariant measures of filtering processes. In K. Helmes and N. Kohlmann, editors, *Stochastic Differential Systems, Proc. 4th Bad Honnef Conf.*, Lecture Notes in Control and Inform. Sci., pages 279–292, 1989.

[103] L. Stettner. Invariant measures of pair state/approximate filtering process. In *Colloq. Math. LXII*, pages 347–352, 1991.

[104] R.L. Stratonovich. Conditional Markov processes. *Theor. Prob. Appl.*, 5:156–178, 1960.

[105] D. Stroock and S. R. S. Varadhan. *Multidimensional Diffusion Processes.* Springer-Verlag, 1979.

[106] D. W. Stroock. *An Introduction to the Theory of Large Deviations.* Universitext. Springer-Verlag, New-York, 1984.

[107] M. Talagrand. Sharper bounds for Gaussian and empirical processes. *The Annals of Probability*, 22:28–76, 1994.

[108] D. Talay. Efficient numerical schemes for the approximation of expectations of functionals of the solution of s.d.e. and applications. In *Filtering and Control of random processes (Paris, 1983)*, Lecture Notes in Control and Inform. Sci. 61, pages 294–313, Berlin-New York, 1984. Springer.

[109] H. Tanaka. Limit theorems for certain diffusion processes. In *Proceedings of the Taniguchi Symp., Katata, 1982*, pages 469–488, Tokyo, 1984. Kinokuniya.

[110] D. Treyer, D.S. Weile, and E. Michielsen. The application of novel genetic algorithms to electromagnetic problems. In *Applied Computational Electromagnetics, Symposium Digest*, volume 2, pages 1382–1386, Monterey, CA, March 1997.

[111] M. D. Vose. *The Simple Genetic Algorithm, Foundations and Theory.* The MIT Press Books, August 1999.

[112] D. Williams. *Probability with Martingales.* Cambridge Mathematical Textbooks, 1992.

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