



On Adaptive Resampling Procedures for Sequential Monte Carlo Methods

Pierre del Moral, Arnaud Doucet, Ajay Jasra

► To cite this version:

Pierre del Moral, Arnaud Doucet, Ajay Jasra. On Adaptive Resampling Procedures for Sequential Monte Carlo Methods. [Research Report] RR-6700, 2008, pp.46. inria-00332436v3

HAL Id: inria-00332436

<https://hal.inria.fr/inria-00332436v3>

Submitted on 21 Oct 2008 (v3), last revised 22 Oct 2008 (v4)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

On Adaptive Resampling Procedures for Sequential Monte Carlo Methods

Pierre Del Moral — Arnaud Doucet — Ajay Jasra

N° 6700

October 2008

Thème NUM

A large blue rectangle occupies the lower half of the page. Overlaid on the left side of this rectangle is a large, light grey stylized letter 'R'. To the right of the 'R', the words 'Rapport de recherche' are written in a grey serif font, with 'Rapport' on the top line and 'de recherche' on the bottom line. A horizontal grey brushstroke underline is positioned below the text.

*Rapport
de recherche*

On Adaptive Resampling Procedures for Sequential Monte Carlo Methods

Pierre Del Moral*, Arnaud Doucet[†], Ajay Jasra[‡]

Thème NUM — Systèmes numériques
Équipes-Projets CQFD

Rapport de recherche n° 6700 — October 2008 — 46 pages

Abstract: Sequential Monte Carlo (SMC) methods are a general class of techniques to sample approximately from any sequence of probability distributions. These distributions are approximated by a cloud of weighted samples which are propagated over time using a combination of importance sampling and resampling steps. This article is concerned with the convergence analysis of a class of SMC methods where the times at which resampling occurs are computed on-line using criteria such as the effective sample size. This is a popular approach amongst practitioners but there are very few convergence results available for these methods. It is shown here that these SMC algorithms correspond to a particle approximation of a Feynman-Kac flow of measures on adaptive excursion spaces. By combining a non-linear distribution flow analysis to an original coupling technique, we obtain functional central limit theorems and uniform exponential concentration estimates for these algorithms. The original exponential concentration theorems presented in this study significantly improve previous concentration estimates obtained for SMC algorithms.

Key-words: Sequential Monte Carlo methods, Feynman-Kac formulae, mean field particle models, effective sample size, adaptive functional criteria.

* Centre INRIA Bordeaux et Sud-Ouest & Institut de Mathématiques de Bordeaux, Université de Bordeaux I, 351 cours de la Libération 33405 Talence cedex, France, Pierre.DelMoral@inria.fr

[†] Department of Statistics, University of British Columbia, 333-6356 Agricultural Road, Vancouver, BC, V6T 1Z2, Canada, arnaud@stat.ubc.ca

[‡] Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-Ku Tokyo 106-8569, Japan, ajay@ism.ac.jp

Méthodes de Monte Carlo séquentielles et adaptatives

Résumé : Cet article porte sur la modélisation et l'analyse de la convergence de méthodes de Monte Carlo séquentielles et adaptatives. Ces techniques de simulation stochastiques sont fondées sur l'évolution aléatoire de systèmes de particules en interaction et associées à des critères d'adaptation. Nous présentons des interprétations de ces algorithmes en termes d'approximation de type champ moyen de flots de mesures de Feynman-Kac dans des espaces d'excursions. Les instants d'interaction sont calculés en lignes selon des critères d'adaptation empiriques tels des critères de tailles de ré-échantillonnage effectives et des critères d'entropie relative de type Boltzmann. Ces stratégies d'adaptation empiriques sont d'un usage courant en pratique souvent présentées comme de simples heuristiques, sans aucun fondement mathématique rigoureux. Nous présentons de nouvelles techniques d'analyse de type semi-groupe fondées sur des couplage de processus à valeurs mesures. Cette analyse permet d'obtenir des théorèmes de la limite centrale fonctionnels et des résultats de concentration exponentielle uniformes par rapport au paramètre temporel. Notons enfin que ces travaux sont les premières études rigoureuses sur la concentration et les fluctuations de ces classes d'algorithmes de simulation adaptatifs. Par ailleurs, les théorèmes de concentration exponentielle présentés dans cette étude améliorent de récents travaux sur ce thème dans le cadre des interprétations particulières de flots de mesures de Feynman-Kac.

Mots-clés : Méthodes de Monte Carlo séquentielles et adaptatives, formules de Feynman-Kac, méthodes particulières de type champ moyen, taille d'échantillonnage effective, critères adaptatifs fonctionnels.

1 Introduction

Sequential Monte Carlo (SMC) methods are a generic class of simulation-based algorithms to sample approximately from any sequence of probability distributions. These methods are now extensively used in statistics, computer science, engineering, and physics; see [4, 9, 12] for many applications. SMC methods approximate the target probability distributions of interest by a large number of random samples termed particles which evolve over time according to a combination of importance sampling and resampling steps.

In the resampling steps, new particles are sampled with replacement from a weighted empirical measure associated to the current particles; see Section 2.1.2 for more details. These resampling steps are crucial and, without them, it is impossible to obtain uniform (in time) convergence results for SMC estimates. However, resampling too often has a negative effect: it decreases the number of distinct particles, increases the variance of SMC estimates and thus should only be applied when necessary. Consequently, in most practical implementations of SMC methods, the times at which resampling occurs are selected by monitoring a criterion which assesses the quality of the current particle approximation. Whenever this criterion is above or below a given, user-set, threshold then resampling occurs. This approach was originally proposed in [13] and, subsequently, has been widely adopted.

For this class of adaptive SMC methods, the times at which resampling occurs are computed on-line using our current SMC approximation and thus are random. However, most of the theoretical results on SMC algorithms assume resampling occurs at deterministic times; see [8] for an exception which is discussed later on. The objective of this article is to provide a thorough treatment of this problem.

Sharp convergence results are provided for this type of SMC algorithms; this is achieved through a coupling argument. Under weak assumptions, the random times at which resampling occurs converge, almost surely as the number of particles goes to infinity, toward some deterministic resampling times. We consider a reference SMC algorithm which resamples at these deterministic times; the times are not explicitly known so this algorithm cannot be implemented but all the convergence results developed in the literature hold for it. It is then shown that the difference, in probability, between this reference SMC algorithm and the adaptive algorithm is exponentially small in the number of particles. This approach allows us to transfer rather straightforwardly the convergence results of the reference SMC algorithm to the adaptive SMC algorithm. In particular, we establish multivariate and functional central limit theorems, yielding what seems to be the first results of this type for this class of algorithms. Moreover, for standard SMC methods and the adaptive SMC schemes presented here, we present new exponential concentration estimates, including time uniform estimates which significantly improve over those presented in [4]. These results allow us to calibrate the number of samples so as to achieve a given precision with a given probability, uniformly with respect to the time parameter.

1.1 Structure of the Article

The rest of the paper is organized as follows. Note that the proofs of some technical results are given in the appendix.

In Section 2, we detail some generic models (Section 2.1.2) and present the class of adaptive SMC algorithms studied here (Section 2.2). Various resampling criteria that are used by practitioners are discussed in Section 2.3. It is shown that they correspond to empirical approximations of limiting criteria defined in Section 2.4.1. This allows us to present our main coupling result in Section 2.4: the difference in probability between the reference SMC algorithm using these limiting criteria and the adaptive SMC algorithm using approximations of these limiting criteria is exponentially small in the number of particles.

In Section 3, a precise description of the sequence of distributions approximated by the reference SMC algorithm is given. We review Feynman-Kac distribution flows in Sections 3.1 and 3.2 then describe such flows on excursion spaces (Section 3.3). In Section 3.4 a general class of functional criteria which are used to define the deterministic resampling times are presented; the construction of these times is detailed in Section 3.5. We further detail the construction of these times for the criteria previously introduced in Section 2.4.1.

In Section 4, a theoretical analysis of the reference SMC algorithm is presented; this is defined formally in Section 4.1. A concentration analysis is given in Section 4.2 which significantly improves the exponential estimates for SMC approximations presented in earlier studies; e.g. those in Section 7.4.3 of [4]. These results are used to obtain a concentration result for the empirical criteria around their limiting values.

The results above are used, in Section 5, to bound the differences between the deterministic resampling times and their empirical approximations, up to an event with an exponentially small probability. This is achieved by introducing a randomized criterion. Finally we analyze the fluctuations of adaptive SMC algorithms in Section 6.

2 Adaptive SMC Algorithms and Main Results

2.1 Target distributions on path spaces

2.1.1 Some Notation

Standard notation used in this article is as follows. $\mathcal{M}(E)$, $\mathcal{P}(E)$, and $\mathcal{B}_b(E)$, respectively denote, the set of bounded and signed measures, the subset of all probability measures on some measurable space $(E, \mathcal{E})$, and the Banach space of all bounded and measurable functions f on E when equipped with the uniform norm $\|f\| = \sup_{x \in E} |f(x)|$. $\text{Osc}_1(E)$ is the convex set of $\mathcal{E}$ -measurable functions f with oscillations $\text{osc}(f) = \sup_{(x,y) \in E^2} \{|f(x) - f(y)|\} \leq 1$. $\mu(f) = \int \mu(dx) f(x)$, is the integral of a function $f \in \mathcal{B}_b(E)$, w.r.t a measure $\mu \in \mathcal{M}(E)$. In a slight abuse the notation, let $\mu(A) = \mu(1_A)$ with $A \in \mathcal{E}$ and 1_A the indicator of A . Recall that a bounded integral operator M from a measurable space $(E, \mathcal{E})$ into an auxiliary measurable space $(F, \mathcal{F})$, is an operator $f \mapsto M(f)$ from $\mathcal{B}_b(F)$ into $\mathcal{B}_b(E)$ such that the functions $M(f)(x) = \int_F M(x, dy) f(y)$ are measurable and bounded, for any $f \in \mathcal{B}_b(F)$. A bounded integral operator M from a measurable space $(E, \mathcal{E})$ into an auxiliary measurable space $(F, \mathcal{F})$

also generates a dual operator $\mu \mapsto \mu M$ from $\mathcal{M}(E)$ into $\mathcal{M}(F)$ defined by $(\mu M)(f) := \mu(M(f))$.

2.1.2 Motivation and Models

Most of the research in Monte Carlo simulation is centered around the computation of integrals, for a wide range of f

$$\begin{aligned}\gamma(f) &:= \int f(x) G(x) \mu(dx) = \mathbb{E}(f(X) G(X)) \\ \eta(f) &:= \gamma(f)/\gamma(1)\end{aligned}\tag{2.1}$$

where X is a random variable distributed according to μ on some measurable space $(E, \mathcal{E})$, G is a given non-negative potential type function on E and $f \in \mathcal{B}_b(E)$. If it is possible to draw independent copies $(X^i)_{i \geq 1}$ from μ , an approximation of the pair of measures (γ, η) can be obtained as

$$\gamma^N(\cdot) := \frac{1}{N} \sum_{i=1}^N G(X^i) \delta_{X^i}(\cdot) \quad \text{and} \quad \eta^N(\cdot) := \sum_{i=1}^N \frac{G(X^i)}{\sum_{j=1}^N G(X^j)} \delta_{X^i}(\cdot)$$

where δ_a is the Dirac measure at the point $a \in E$. In practice, these approximations can suffer from several problems. For instance, assume that $G = 1_A$ and $\mu(A) := \mathbb{P}(X \in A) \simeq 10^{-p}$, for some large parameter $p \geq 1$. In this situation, it may not be possible to find a sample X^i within the set A , so that $\gamma^N(\cdot) = 0$ and the measure $\eta^N(\cdot)$ is not well defined. To overcome this difficulty, importance sampling techniques are often used. This strategy consists of finding an auxiliary importance distribution, say μ' , such that $\mu \ll \mu'$ with Radon-Nikodym derivative $\frac{d\mu}{d\mu'}$, and to set

$$\gamma(f) = \mathbb{E}(f(X') G'(X')) = \int f(x) G'(x) \mu'(dx) \quad \text{with} \quad G'(x) := G(x) \frac{d\mu}{d\mu'}(x)$$

where $X' \sim \mu$ and μ' a distribution on E . This new model has exactly the same form as the one discussed above, so that we can use the same Monte Carlo strategy, replacing the pair of parameters (μ, G) by (μ', G') .

For importance sampling, even if we have found a good importance distribution, another significant problem often arises. Namely, the Radon-Nikodym derivative reflects the relative differences between the desired reference measure and the importance sampling one. These weight functionals are generally degenerate (also known as *weight degeneracy*). For instance, assume X represents the path of a sequence of random variables $(X'_0, \dots, X'_n)$, and the potential functions G are a product of a sequence of functions from the origin up to a given time horizon n

$$X = (X'_0, \dots, X'_n) \quad \text{and} \quad G(X) = \prod_{0 \leq p < n} G'_p(X'_p)$$

where G'_p is a potential function of the random state X'_p . As n increases, so does the relative variance of G , and accurate estimation of (2.1) is very difficult. In this situation, the pair of measures (γ, η) given below depend on the time

parameter, and they are given for every test function f_n on the path space by the following Feynman-Kac integral representation formulae

$$\eta_n(f_n) = \gamma_n(f_n)/\gamma_n(1) \quad \text{with} \quad \gamma_n(f_n) = \mathbb{E} \left(f_n(X'_0, \dots, X'_n) \prod_{0 \leq p < n} G'_p(X'_p) \right). \quad (2.2)$$

From a pure mathematical point of view, we emphasize that every importance sampling/change of measure, of the type discussed above, reduces to the same type of Feynman-Kac model.

These probabilistic models and their SMC approximations appear in numerous fields including: probability, Bayesian statistics, theoretical chemistry, quantum physics and engineering sciences. For instance, in signal processing, $(X'_n)_{n \geq 0}$ is often a Markov chain and represents a signal process and the potential functions are the likelihood functions associated with some noisy observations. The corresponding measures η_n are the conditional distributions of the random paths of the signal, given a series of observations. In particle physics, the law of an unabsorbed particle X'_n evolving in an absorbing medium E_n associated with an energy well V is given by a Feynman-Kac integral formulae. In this case, there is a Boltzmann-Gibbs potential function $G'_n(X'_n) = \exp(-\beta V(X'_n))$. For more details on these models and their connections to molecular simulation and particle filtering techniques, see, for example, chapter 12 of [4].

We remark that in most of the application areas discussed above, the free evolution process X'_n and the potential functions G'_n are given by the physical problem at hand. In other instances, the choice of these parameters is more flexible. Let us assume that the target measure is defined in terms of a Boltzmann-Gibbs mass transformation of a reference ‘static’ measure λ , with respect to some potential function g_n on some measurable space E ; that is

$$\pi_n(dx) = \Psi_{g_n}(\lambda)(dx) := \frac{1}{\mathcal{Z}_n} g_n(x) \lambda(dx) \quad \text{with} \quad \mathcal{Z}_n := \lambda(g_n).$$

For instance, in combinatorial counting problems λ often represents the uniform measure on a given finite set E and $g_n = 1_{A_n}$ the indicator function a subset $A_n \subset E$. In global optimization problems, λ may represent a reference measure on a given space E and $g_n = \exp(-\beta_n V)$ a Boltzmann-Gibbs factor associated with an inverse temperature parameter β and an energy type non-negative function V . Numerous algorithms have been proposed in the Markov chain Monte Carlo community to sample from these Boltzmann-Gibbs measures; that is, we design Markov kernels M'_n with a prescribed invariant measure $\pi_n = \pi_n M'_n$. For instance, Metropolis-Hastings transitions consist of sampling a candidate, and to accept or reject it according to an appropriate acceptance ratio which ensures that the target measure π_n is reversible w.r.t these local transitions. The well-known simulated annealing algorithm consists of finding a judicious schedule β_n increasing sufficiently slowly to ensure that successive Metropolis-Hasting moves between these changes are close to equilibrium. Another strategy, consists of finding a sequence of multiplicative potential functions, so that

$$(g_{n+1} = g_n \times G'_n \quad \text{and} \quad \pi_n = \pi_n M'_n) \Rightarrow \pi_{n+1} = \Psi_{G'_n}(\pi_n) M'_{n+1}. \quad (2.3)$$

Returning to the examples discussed above, we can choose the exponential Boltzmann-Gibbs potential and respectively the indicator potential functions

defined by

$$G'_n = \exp(-(\beta_{n+1} - \beta_n)V), \quad \text{and respectively} \quad G'_n = 1_{A_{n+1}} \quad (2.4)$$

with a non-decreasing inverse temperature parameter β_n and a non-increasing sequence of subsets A_n .

Using the Markov property and the multiplicative structure of the Feynman-Kac measures (2.2), it can be proved that the measures π_n given in (2.3) coincide with the n -time marginals of the path measures defined in (2.2); that is

$$\pi_n(f) \propto \mathbb{E} \left(f(X'_n) \prod_{0 \leq p < n} G'_p(X'_p) \right) \quad \text{and} \quad \lambda(g_n) := \lambda(g_0) \times \prod_{0 \leq p < n} \pi_p(G'_{p+1}). \quad (2.5)$$

We underline that these integral representation formulae are valid for any choice of the parameters (G'_n, M'_n) satisfying the pair of admissibility conditions presented in the L.H.S of (2.3). This formula is known as Crooks identity in statistical physics [2]; see [7] for a generalization. It is a discrete-time version of the celebrated Jarzynski identity [10]; see [14] for a continuous-time SMC implementation.

2.2 Adaptive sequential Monte Carlo

SMC methods are a popular class of methods for approximately sampling random variables according to the Feynman-Kac path measures (2.2) and computing their normalizing constants. These algorithms are sometimes referred to as population Monte Carlo, branching and interacting particle systems or diffusion/quantum Monte Carlo methods. The central idea is now given.

First, N particles evolve according to the free path evolution of the reference Markov chain X'_n . In a second stage, particles with low G'_n -potential value are killed, and those with a relative larger potential are duplicated. In this sense, we define a branching particle model with a genealogical tree occupation measure that approximates the desired measures (2.2) as N tends to ∞ . The theoretical foundations of these algorithms are now rather well understood. From the pure mathematical point of view, they can be regarded as a mean field particle approximation of a non-linear Feynman-Kac flow of distributions.

In recent years, researchers have built new branching/resampling strategies. The idea is to evolve the particles up to some times that depend on criteria computed using the current particle approximation. These adaptive algorithms have several advantages over standard methods:

1. In the context of the Feynman-Kac path measures (2.2), these criteria can be used to design online strategies controlling the ‘quality’ of the current particle approximation. Popular criteria include the variance of the weights, or entropy like distances between the free evolution distribution and the weighted one.
2. In the context of Boltzmann-Gibbs measures (2.3), they allow us to define adaptive temperature schedules or adaptive subset sequences. For instance, we can decide to change the temperature parameter at some appropriate time periods. These time parameters and the next cooling temperature may depend on the performance of the exploration of the individuals in the energy landscape.

In terms of the latter point, some related investigations can be found in [11].

2.2.1 Adaptive Framework

Adaptive SMC algorithms proceed as follows. Let $((E'_n, \mathcal{E}'_n))_{n \geq 0}$ be a sequence of measurable spaces. The idea is to propagate forward a collection of N sample paths copies $(Y_n^{(N,i)} \in E'_n)$

$$\forall 1 \leq i \leq N \quad \mathcal{Y}_1^{(N,i)} := (Y_0^{(N,i)}, Y_1^{(N,i)}, \dots, Y_{t_1^N}^{(N,i)})$$

of the path valued random variables $(X'_0, \dots, X'_{t_1^N})$ up to the first time t_1^N the importance weights

$$W_{0,t_1^N}(\mathcal{Y}_1^{(N,i)}) := \prod_{0 < p \leq t_1^N} G'_p(Y_p^{(N,i)})$$

($G'_0 \equiv 1$, although this need not be the case), of the N path samples become, in some respect, degenerate; see Section 2.3 for more details. Loosely speaking, at time t_1^N the weighted occupation measure of the system

$$\tilde{\eta}_1^N(\cdot) := \sum_{i=1}^N \frac{W_{0,t_1^N}(\mathcal{Y}_1^{(N,i)})}{\sum_{j=1}^N W_{0,t_1^N}(\mathcal{Y}_1^{(N,j)})} \delta_{\mathcal{Y}_1^{(N,i)}}(\cdot)$$

is an N particle approximation measure of the updated target measure given by

$$\hat{\eta}_1(f) = \hat{\gamma}_1(f)/\hat{\gamma}_1(1) \quad \text{with} \quad \hat{\gamma}_1(f) = \mathbb{E} \left(f(X'_0, \dots, X'_{t_1^N}) \prod_{0 < p \leq t_1^N} G'_p(X'_p) \right).$$

The precise analysis of these approximations is one of the objectives of the article will be discussed later.

To prevent the weighted measure $\tilde{\eta}_1^N$ from becoming degenerate after the selected time period t_1^N , these measures are replaced by a bootstrap type empirical measure

$$\hat{\eta}_1^N(\cdot) := \frac{1}{N} \sum_{i=1}^N \delta_{\hat{\mathcal{Y}}_1^{(N,i)}}(\cdot)$$

associated with N path particles $\hat{\mathcal{Y}}_1^{(N,i)} := (\hat{Y}_{0,t_1^N}^{(N,i)}, \hat{Y}_{1,t_1^N}^{(N,i)}, \dots, \hat{Y}_{t_1^N,t_1^N}^{(N,i)})$ that are resampled; for instance randomly from the current weighted measure $\tilde{\eta}_1^N$.

In a second stage we propagate forward, from every end point $\hat{Y}_{t_1^N,t_1^N}^{(N,i)}$ of the selected paths $\hat{\mathcal{Y}}_1^{(N,i)}$, a collection of N sample path copies

$$(Y_{t_1^N+1}^{(N,i)}, Y_{t_1^N+2}^{(N,i)}, \dots, Y_{t_2^N}^{(N,i)})$$

of the path-valued random variables $(X'_{t_1^N+1}, X'_{t_1^N+2}, \dots, X'_{t_2^N})$ starting at $X'_{t_1^N} = \hat{Y}_{t_1^N,t_1^N}^{(N,i)}$, up to the first time t_2^N , the importance weights of the N path samples given by

$$W_{t_1^N,t_2^N}(\mathcal{Y}_2^{(N,i)}) := \prod_{t_1^N < p \leq t_2^N} G'_p(Y_p^{(N,i)})$$

again become degenerate. In the above expression, $\mathcal{Y}_2^{(N,i)}$ stands for the resulting path sequences of length $t_2^N + 1$ given by the formula

$$\mathcal{Y}_2^{(N,i)} := \left(\widehat{\mathcal{Y}}_1^{(N,i)}, (Y_{t_1^N+1}^{(N,i)}, Y_{t_1^N+2}^{(N,i)}, \dots, Y_{t_2^N}^{(N,i)}) \right).$$

The following weighted occupation measure of the system

$$\widehat{\eta}_2^N(\cdot) := \sum_{i=1}^N \frac{W_{t_1^N, t_2^N}(\mathcal{Y}_2^{(N,i)})}{\sum_{j=1}^N W_{t_1^N, t_2^N}(\mathcal{Y}_2^{(N,j)})} \delta_{\mathcal{Y}_2^{(N,i)}}(\cdot)$$

is an approximation of the updated target measure at time t_2^N :

$$\widehat{\eta}_2(f) = \widehat{\gamma}_2(f) / \widehat{\gamma}_2(1) \quad \text{with} \quad \widehat{\gamma}_2(f) = \mathbb{E} \left(f(X'_0, \dots, X'_{t_2^N}) \prod_{0 < p \leq t_2^N} G'_p(X'_p) \right).$$

As before, to prevent the degeneracy of the weighted measures, after the selected time period t_2^N , the weighted measure $\widehat{\eta}_2^N$ is replaced by a bootstrap type empirical measure

$$\widehat{\eta}_2^N(\cdot) := \frac{1}{N} \sum_{i=1}^N \delta_{\widehat{\mathcal{Y}}_2^{(N,i)}}(\cdot)$$

associated with N particle paths $\widehat{\mathcal{Y}}_2^{(N,i)} := (\widehat{Y}_{0, t_2^N}^{(N,i)}, \widehat{Y}_{1, t_2^N}^{(N,i)}, \dots, \widehat{Y}_{t_2^N, t_2^N}^{(N,i)})$ resampled from the current weighted measure $\widehat{\eta}_2^N$, and so on up to the desired final time horizon.

Remark 2.1 *This adaptive excursion model is quite flexible. For instance, in the context of the Boltzmann-Gibbs model (2.3), it is possible to choose the same potential function $G'_{t_n^N+1} := G'_p$ between the interaction times $t_n^N < p \leq t_{n+1}^N$. In other words, the potentials are only changed at the random interaction times t_n^N*

$$\begin{aligned} \forall t_n^N < p \leq t_{n+1}^N \quad G'_p &:= G'_{t_n^N+1} \\ &\downarrow \\ \forall t_{n+1}^N < p \leq t_{n+2}^N \quad G'_p &:= G'_{t_{n+1}^N+1} \\ &\downarrow \\ &\dots \end{aligned}$$

As an example, a temperature schedule can be used with the same decay rate between the interaction times $t_n^N < p \leq t_{n+1}^N$, and then changed only at the interaction times. Using this strategy, we have used only n different potential functions at time t_{n+1}^N :

$$G'_{t_0^N+1} \rightsquigarrow G'_{t_1^N+1} \rightsquigarrow \dots \rightsquigarrow G'_{t_n^N+1} \rightsquigarrow .$$

Without further work, the forthcoming analysis can be extended to situations where the potential functions $(G'_p)_{p > t_n^N}$, at a given level n , depend on the series of parameters $(t_0^N, \dots, t_n^N)$. More formally, it is possible to choose

$$\forall t_n^N < p \leq t_{n+1}^N \quad G'_p(x) := G''_{p, (t_0^N, \dots, t_n^N)}(x) \quad (2.6)$$

for some collection of mappings $G''_{p,(t_0,\dots,t_n)}$, with $0 = t_0 \leq \dots \leq t_n < p$. In much the same way, in the context of the Boltzmann-Gibbs model (2.3), we can use a sequence of Markov transitions $(M'_p)_{p > t_n^N}$ as well as an adaptive resampling criterion at a given level n that depends on $(t_0^N, \dots, t_n^N)$.

2.3 Some empirical criteria

A few empirical criteria have been proposed in the SMC literature to measure the degeneracy of the weighted samples; these criteria can be used to design adaptive resampling strategies. In all of these situations, the resampling times $(t_n^N)_{n \geq 0}$ are random variables that depend on the current SMC approximation.

Four criteria are now presented which are termed 1)-4), in the order that they appear. Due to their practical importance, and to better connect these strategies with existing results in the literature, we present a rather detailed discussion on the resulting adaptive resampling strategies.

2.3.1 Adaptive control of the weights

In this case, the particles explore the state-space with a series of local transitions

$$\hat{Y}_{t_n^N, t_n^N}^{(N,i)} \longrightarrow Y_{t_{n+1}^N}^{(N,i)} \longrightarrow Y_{t_{n+2}^N}^{(N,i)} \longrightarrow \dots \longrightarrow Y_s^{(N,i)} \quad (2.7)$$

up to the first time $s = t_{n+1}^N$ ($\geq t_n^N$) the proportion of particles with a weight of at least size s_n is below a prescribed a_n -threshold

$$C_{t_n^N, s}^N := \frac{1}{N} \text{Card} \left\{ 1 \leq i \leq N : \prod_{t_n^N < k \leq s} G'_k(Y_k^{(N,i)}) \geq s_n \right\} \leq a_n \quad (2.8)$$

where $0 < a_n \leq 1$. To our knowledge, this type of criteria was first discussed as an heuristic type resampling schedule in [3]. Next, we provide some comments on the choice of the adaptive parameters (s_n, a_n) . The forthcoming discussion applies to all the criteria discussed in the article.

For $[0, 1]$ -valued potential functions G'_k , it is sensible to choose $s_n \in [0, 1]$. In this situation, the random mappings

$$s \in [t_n^N, \infty) \mapsto C_{t_n^N, s}^N \in [0, 1]$$

are non-increasing functions starting at $C_{t_n^N, t_n^N}^N = 1$ (the standard convention $\prod_{\emptyset} = 1$ is used). It should be noted that when none of the sampled particles at time $s = t_n^N + 1$ have a weight greater than the prescribed threshold s_n , then $t_{n+1}^N = t_n^N$ and the above sampling strategy is restarted. As discussed in Remark 2.1, our framework is general enough that it allows us to analyze, without further work, fully adaptive algorithms. For example, we may use a sequence of threshold parameters that depend on the previous times steps

$$s_n = s'_n(t_0^N, \dots, t_n^N) \quad \text{and} \quad a_n = a'_n(t_0^N, \dots, t_n^N)$$

for some functions s' and a'_n on $\mathbb{R}_+^{n+1}$. So, in cases where $t_{n+1}^N = t_n^N$, it is possible to either choose to restart the complete run with the same set of potentials and

Markov transitions used before and/or to decide to decrease the weight size parameter s_n as well as the threshold level a_n .

The choice of these threshold parameters depends on the problem at hand. From the previous discussion, if s_n is excessively large, the algorithm may get stuck at a given time step t_n^N during several runs. In the further developments of the article, we provide a set of regularity conditions under which these empirical criteria approximate a limiting functional integration criteria. It will be shown, that the random times t_n^N converge, as N tends to ∞ , to a series of deterministic times t_n expressed in terms of these functional criteria. In this situation, for a large population size N we ensure that $t_{n+1}^N > t_n^N$ with high probability as soon as $t_{n+1} > t_n$.

To illustrate this idea, suppose that $(X'_k)_{k \geq 0}$ is a Markov chain on $E'_k = \mathbb{Z}$ starting at the origin $X'_0 = 0$. For instance, X'_k can be taken as a simple random walk on the integer lattice. If we choose the indicator potential functions and the unit weight size parameters

$$G'_k = 1_{[k, \infty)} \quad \text{and} \quad s_n = 1 \quad (2.9)$$

then we find

$$\prod_{t_n^N < k \leq s} G'_k(Y_k^{(N,i)}) \geq 1 \iff \left(\forall t_n^N < k \leq s \quad Y_k^{(N,i)} \geq k \right)$$

for every $1 \leq i \leq N$. Thus,

$$C_{t_n^N, s}^N = \frac{1}{N} \text{Card} \left\{ 1 \leq i \leq N : \forall t_n^N < k \leq s \quad Y_k^{(N,i)} \geq k \right\}. \quad (2.10)$$

When $a_n = 1/2$, we find that t_{n+1}^N is such that

$$t_{n+1}^N = \inf \left\{ s \geq t_n^N : \right.$$

$$\left. \text{Card} \left\{ 1 \leq i \leq N : Y_{t_n^N+1}^{(N,i)} \geq t_n^N + 1, Y_{t_n^N+2}^{(N,i)} \geq t_n^N + 2, \dots, Y_s^{(N,i)} \geq s \right\} \leq N/2 \right\}.$$

Roughly speaking, for the simple random walk model, this ensures that about half of the N particles are able to reach the next upper level set at every elementary time step. In this situation, the random times t_{n+1}^N are more likely to take the value $t_n^N + 1$, or to return the value t_n^N . If we take a lower threshold parameter, say $a_n = 1/100$, these random times t_{n+1}^N are more likely to take large values. Indeed, in this case, about $N/2$ particles will be able to reach the first level, then about $N/4$ the second, $N/8$ the third one, and so on up to $N/128$ the seventh. Therefore, for $a_n = 1/100$ it is expected that $t_{n+1}^N = t_n^N + 6$.

The random walk model discussed above can easily be extended to multi-splitting branching excursion type models, absorbed at some lower level set $B \subset \mathbb{Z} \cap \mathbb{R}_-$. Here, the random path processes $(Y_k^{(N,i)})_{k \geq t_n^N}$ are independent copies of a random excursion starting from a given level set t_n^N and ending at the absorbing set B . The next adaptive step t_{n+1}^N coincides with the first upper level set the N particles are able to reach with a proportion less than a_n (that is, before being absorbed by the set B). The resulting algorithm for adaptive multi-level splitting coincides with the adaptive branching algorithm for rare event simulation discussed in [1].

2.3.2 Normalizing constants

In this scenario, the particles explore the space with a series of local transitions (2.7) up to the first time $s = t_{n+1}^N$ the average of the importance weights - which is an estimate of a normalizing constant - goes below a prescribed threshold a_n

$$C_{t_n^N, s}^N := \frac{1}{N} \sum_{i=1}^N \prod_{t_n^N < k \leq s} G'_k(Y_k^{(N,i)}) \leq a_n. \quad (2.11)$$

For indicator potential functions, this criterion coincides with the one discussed above; see (2.10). Furthermore, for $(0, 1]$ -valued potential functions G'_k , it is convenient in (2.11) to set $a_n \in [0, 1]$; the random mappings

$$s \in [t_n^N, \infty) \mapsto C_{t_n^N, s}^N \in [0, 1]$$

are non-increasing functions starting at $C_{t_n^N, t_n^N}^N = 1$. Notice that $a_n = 1 \implies t_{n+1}^N = t_n^N$. For $(0, 1)$ -valued threshold parameters, we have the rather crude estimates

$$\underline{g}_k := \inf G'_k \leq \sup G'_k := \bar{g}_k$$

$$\Downarrow$$

$$\prod_{t_n^N < k \leq s} \underline{g}_k := \underline{C}_{t_n^N, s}^N \leq C_{t_n^N, s}^N \leq \bar{C}_{t_n^N, s}^N := \prod_{t_n^N < k \leq s} \bar{g}_k. \quad (2.12)$$

This pair of estimates can be improved by considering the infimum and the supremum of the potential product weight functions along the M'_k -admissible paths sequences (2.7) starting at the selected states $\hat{Y}_{t_n^N, t_n^N}^{(N,i)}$. Note that (2.12) implies

$$\underline{t}_{n+1}^N \leq t_{n+1}^N \leq \bar{t}_{n+1}^N$$

with

$$\underline{t}_{n+1}^N := \inf \left\{ s \geq t_n^N : \underline{C}_{t_n^N, s}^N \leq a_n \right\} \quad \text{and} \quad \bar{t}_{n+1}^N := \inf \left\{ s \geq t_n^N : \bar{C}_{t_n^N, s}^N \leq a_n \right\}.$$

For instance, for the exponential Boltzmann-Gibbs potential given in (2.4), we have

$$\underline{V} := \inf V \leq \sup V := \bar{V} \implies \underline{g}_k = e^{-(\beta_{k+1} - \beta_k) \bar{V}} \quad \text{and} \quad \bar{g}_k = e^{-(\beta_{k+1} - \beta_k) \underline{V}}.$$

This clearly implies that

$$\underline{C}_{t_n^N, s}^N = e^{-(\beta_{s+1} - \beta_{t_n^N+1}) \bar{V}} \quad \text{and} \quad \bar{C}_{t_n^N, s}^N = e^{-(\beta_{s+1} - \beta_{t_n^N+1}) \underline{V}}.$$

In addition, further assume that the temperature schedule $(\beta_k)_{k > t_n^N}$ after time t_n^N is associated with a fixed decay rate $\delta_n > 0$, then

$$\begin{aligned} \forall t_n^N < k \quad \Delta \beta_{k+1} = (\beta_{k+1} - \beta_k) = \delta_n &\implies (\beta_{s+1} - \beta_{t_n^N+1}) = (s - t_n^N) \delta_n \\ &\implies \begin{cases} \underline{C}_{t_n^N, s}^N &= e^{-(s - t_n^N) \delta_n \bar{V}} \\ \bar{C}_{t_n^N, s}^N &= e^{-(s - t_n^N) \delta_n \underline{V}}. \end{cases} \end{aligned}$$

From the previous discussion, it can be concluded that

$$t_n^N + \frac{1}{\delta_n \bar{V}} \log(1/a_n) := \underline{t}_{n+1}^N \leq t_{n+1}^N \leq \bar{t}_{n+1}^N = t_n^N + \frac{1}{\delta_n \bar{V}} \log(1/a_n).$$

For small values of the inverse temperature decay rates δ_n , it is ensured that the criterion (2.11) is met in at least $\frac{1}{\delta_n \bar{V}} \log(1/a_n)$ time steps. Alternatively, if we want to ensure that (2.11) is met in at least m_n time steps then $\delta_n \leq (m_n \bar{V})^{-1} \log(1/a_n)$ is needed; that is

$$\delta_n \leq (m_n \bar{V})^{-1} \log(1/a_n) \implies (t_{n+1}^N - t_n^N) \geq m_n.$$

2.3.3 Variance of the weights

In this situation, the particles explore the state-space with a series of local transitions (2.7) up to the first time $s = t_{n+1}^N$ the empirical particle variance is larger than some prescribed threshold a_n

$$\sum_{i=1}^N \left(\frac{\prod_{t_n^N < k \leq s} G'_k(Y_k^{(N,i)})}{\sum_{j=1}^N \prod_{t_n^N < k \leq s} G'_k(Y_k^{(N,j)})} \right)^2 \geq (1 + a_n)/N$$

or equivalently

$$C_{t_n^N, s}^N := \frac{1}{N} \sum_{i=1}^N \left(\frac{\prod_{t_n^N < k \leq s} G'_k(Y_k^{(N,i)})}{\frac{1}{N} \sum_{j=1}^N \prod_{t_n^N < k \leq s} G'_k(Y_k^{(N,j)})} \right)^2 - 1 \geq a_n.$$

This is equivalent to resampling when the so-called effective sample size is below a prescribed threshold. This popular strategy was first proposed in [13]. For indicator potential functions, this criterion coincides with those above; as $(G'_k)^2 = G'_k$.

2.3.4 Boltzmann's entropy

Assume that the potential functions G'_k take values in $(0, \infty)$. In this case, the particles explore the state-space with a series of local transitions (2.7) up to the first time $s = t_{n+1}^N$ such that the relative entropy of the empirical particle measures w.r.t its weighted versions are larger than some threshold a_n

$$C_{t_n^N, s}^N := -\frac{1}{N} \sum_{i=1}^N \sum_{t_n^N < k \leq s} \log G'_k(Y_k^{(N,i)}) \geq a_n. \quad (2.13)$$

For the exponential Boltzmann-Gibbs potential given in (2.4), we have

$$C_{t_n^N, s}^N := \frac{1}{N} \sum_{i=1}^N \sum_{t_n^N < k \leq s} (\beta_{k+1} - \beta_k) V(Y_k^{(N,i)}).$$

In this situation, the random mappings

$$s \in [t_n^N, \infty) \mapsto C_{t_n^N, s}^N \in [0, 1]$$

are non-decreasing functions starting at $C_{t_n, t_n}^N = 0$ (using the standard convention $\sum_{\emptyset} = 0$). Arguing as above, we have the rather crude estimates

$$\begin{aligned} \underline{V} &:= \inf V \leq \sup V := \bar{V} \\ &\Downarrow \\ \underline{C}_{t_n, s}^N &= (\beta_{s+1} - \beta_{t_n+1}) \underline{V} \leq C_{t_n, s}^N \leq \bar{C}_{t_n, s}^N := (\beta_{s+1} - \beta_{t_n+1}) \bar{V}. \end{aligned}$$

This implies that

$$t_{n+1}^N \leq t_{n+1}^N \leq \bar{t}_{n+1}^N$$

with

$$t_{n+1}^N := \inf \left\{ s \geq t_n^N : \bar{C}_{t_n, s}^N \geq a_n \right\} \quad \text{and} \quad \bar{t}_{n+1}^N := \inf \left\{ s \geq t_n^N : \underline{C}_{t_n, s}^N \geq a_n \right\}.$$

When the temperature schedule $(\beta_k)_{k > t_n^N}$ is associated with a prescribed and fixed decay rate $\delta_n > 0$, it follows that

$$\underline{C}_{t_n, s}^N = (s - t_n^N) \delta_n \underline{V} \quad \text{and} \quad \bar{C}_{t_n, s}^N = (s - t_n^N) \delta_n \bar{V}$$

and therefore

$$t_{n+1}^N := t_n^N + a_n / (\delta_n \bar{V}) \leq t_{n+1}^N \leq \bar{t}_{n+1}^N := t_n^N + a_n / (\delta_n \underline{V}).$$

In this case, to ensure that the criterion (2.13) is met in at least m_n iterations, we need to choose an inverse temperature decay rate $\delta_n \leq a_n / (m_n \bar{V})$; that is

$$\delta_n \leq a_n / (m_n \bar{V}) \implies (t_{n+1}^N - t_n^N) \geq m_n.$$

2.4 Statement of some results

The following Section provides a guide of the major definitions and results in this article; these will be repeated at the relevant points in the paper.

2.4.1 A limiting reference SMC algorithm

Let $(t_n)_{n \geq 0}$ be the deterministic sequence of time steps obtained by replacing the empirical criteria $C_{t_n, s}^N$ by their limiting values $C_{t_n, s}$, i.e. as $N \uparrow \infty$; e.g. in the examples 1) and 2)

$$t_{n+1} := \inf \{ t_n \leq s : C_{t_n, s} \leq a_n \}.$$

In all situations, set $t_0^N = t_0 = 0$.

In the examples 3) and 4), the limiting times $(t_n)_{n \geq 0}$ are defined as above by replacing condition $C_{t_n, s} \leq a_n$ by the condition $C_{t_n, s} \geq a_n$. In the four situations discussed above, these limiting criteria $C_{t_n, s}$ are respectively given by

$$1) \quad \mathbb{P}_{n, \hat{\eta}_n} \left(\prod_{t_n < k \leq s} G'_k(X'_k) \geq a_n \right) \quad 2) \quad \mathbb{E}_{n, \hat{\eta}_n} \left(\prod_{t_n < k \leq s} G'_k(X'_k) \right)$$

and

$$3) \quad \frac{\mathbb{E}_{n,\hat{\eta}_n} \left(\prod_{t_n < k \leq s} G'_k(X'_k)^2 \right)}{\mathbb{E}_{n,\hat{\eta}_n} \left(\prod_{t_n < k \leq s} G'_k(X'_k) \right)} - 1 \quad 4) \quad - \sum_{t_n < k \leq s} \mathbb{E}_{n,\hat{\eta}_n} (\log G'_k(X'_k))$$

with the updated Feynman-Kac path measures $\hat{\eta}_n$ given below

$$\hat{\eta}_n(f) = \hat{\gamma}_n(f) / \hat{\gamma}_n(1) \quad \text{with} \quad \hat{\gamma}_n(f) = \mathbb{E} \left(f(X'_0, \dots, X'_{t_n}) \prod_{0 < p \leq t_n} G'_p(X'_p) \right).$$

In the above formulae, $\mathbb{E}_{n,\hat{\eta}_n}$ is the expectation w.r.t the law of a free evolution X'_k starting at time t_n at the end point $X'_{t_n} = \hat{X}'_{t_n,t_n}$ of a random path $\hat{\mathcal{X}}_n := (\hat{X}'_{0,t_n}, \hat{X}'_{1,t_n}, \dots, \hat{X}'_{t_n,t_n})$ with distribution $\hat{\eta}_n$; that is

$$\hat{X}_{t_n,t_n} \sim \hat{\eta}_n \xrightarrow{M'_{t_n+1}} X'_{t_n+1} \xrightarrow{M'_{t_n+2}} X'_{t_n+2} \xrightarrow{M'_{t_n+3}} \dots \xrightarrow{M'_s} X'_s \quad (2.14)$$

Definition 2.2 *Let*

$$\mathcal{Y}_n^{(N)} := \left(\mathcal{Y}_n^{(N,1)}, \mathcal{Y}_n^{(N,2)}, \dots, \mathcal{Y}_n^{(N,N)} \right)$$

and

$$\hat{\mathcal{Y}}_n^{(N)} := \left(\hat{\mathcal{Y}}_n^{(N,1)}, \hat{\mathcal{Y}}_n^{(N,2)}, \dots, \hat{\mathcal{Y}}_n^{(N,N)} \right)$$

denote the N particles associated to the adaptive SMC algorithm resampling at times $(t_n^N)_{n \geq 0}$ and let

$$\mathcal{X}_n^{(N)} := \left(\mathcal{X}_n^{(N,1)}, \mathcal{X}_n^{(N,2)}, \dots, \mathcal{X}_n^{(N,N)} \right)$$

and

$$\hat{\mathcal{X}}_n^{(N)} := \left(\hat{\mathcal{X}}_n^{(N,1)}, \hat{\mathcal{X}}_n^{(N,2)}, \dots, \hat{\mathcal{X}}_n^{(N,N)} \right)$$

denote the N particles associated to the reference SMC algorithm resampling at times $(t_n)_{n \geq 0}$. We also suppose that $(\mathcal{X}_n^{(N)}, \hat{\mathcal{X}}_n^{(N)})$ and $(\mathcal{Y}_n^{(N)}, \hat{\mathcal{Y}}_n^{(N)})$ coincide on every time interval $0 \leq n \leq m$, as soon as the random times $t_n^N = t_n$, for every $0 \leq n \leq m$. This condition simply corresponds to the coupling of the two processes on the events $\cap_{0 \leq n \leq m} \{t_n^N = t_n\}$.

2.4.2 An exponential coupling theorem

We give our main results, which hold under the following regularity condition:

$$(G') \quad \forall n \geq 0 \quad q'_n := \sup_{(x,y) \in (E'_n)^2} (G'_n(x)/G'_n(y)) < \infty. \quad (2.15)$$

The analysis of more general models is technically more involved and often requires some additional assumptions on the free evolution process. We refer the interested reader to [4] for a more thorough discussion on these technicalities.

The first result is a non-asymptotic exponential concentration estimate. Below, if constants are written with an argument, then they depend only on this given argument.

Theorem 2.3 *For any $n \geq 0$, $f_n \in \text{Osc}_1(E'_0 \times \cdots \times E'_{t_n})$, any $N \geq 1$ and any $0 \leq \epsilon \leq 1/2$, there exist $c_1 < \infty$, $0 < c_2(n) < \infty$ such that we have the exponential concentration estimate*

$$\mathbb{P}(|[\eta_n^N - \eta_n](f_n)| \geq \epsilon) \leq c_1 \exp\{-N \epsilon^2 / c_2(n)\}$$

for the empirical measures

$$\eta_n^N(\cdot) = \frac{1}{N} \sum_{i=1}^N \delta_{(\hat{\mathcal{X}}_{n-1}^{(N,i)}, (X_{t_{n-1}+1}^{(N,i)}, X_{t_{n-1}+2}^{(N,i)}, \dots, X_{t_n}^{(N,i)}))}(\cdot).$$

In addition, under appropriate regularity conditions on the potential transitions G'_k and on the Markov chain X'_k given in Section (4.2.1) the above estimates are valid for the marginal measures associated with the time parameters $t_{n-1} \leq p \leq t_n$, for some constant $c_2(n) = c_2$ whose values does not depend on the time parameter.

The second result is the following exponential coupling theorem.

Theorem 2.4 *Assume the threshold parameters $(a_n)_{n \geq 0}$ are sampled realizations of a collection of absolutely continuous random variables $A = (A_n)_{n \geq 0}$. Then, for almost every realization of the sequence $(a_n)_{n \geq 0}$, $(\mathcal{X}_n^{(N)}, \hat{\mathcal{X}}_n^{(N)})_{n \geq 0}$ and $(\mathcal{Y}_n^{(N)}, \hat{\mathcal{Y}}_n^{(N)})_{n \geq 0}$ are such that, for every $m \geq 0$ and any $N \geq 1$, there exist $0 < c_1(m), c_2(m) < \infty$ and almost surely $\epsilon(m, A) \equiv \epsilon(m) > 0$ such that*

$$\mathbb{P}(\exists 0 \leq n \leq m \quad (\mathcal{Y}_n^{(N)}, \hat{\mathcal{Y}}_n^{(N)}) \neq (\mathcal{X}_n^{(N)}, \hat{\mathcal{X}}_n^{(N)}) \mid A) \leq c_1(m) e^{-N \epsilon^2(m) / c_2(m)}.$$

The proofs of Theorem 2.3 and Theorem 2.4 are detailed, respectively, in Sections 4.2.3 and 5.2. Some direct consequences of the exponential coupling Theorem 2.4 are as follows. The reference SMC algorithm $(\mathcal{X}_n^{(N)}, \hat{\mathcal{X}}_n^{(N)})_{n \geq 0}$ can be regarded as a ‘standard’ SMC algorithm as it resamples at deterministic times. A variety of convergence results have already been developed for these algorithms (see for instance [4]), including non asymptotic exponential estimates and $\mathbb{L}_p$ -mean error bounds. Up to an event having an exponentially small occurrence probability, Theorem 2.4 allows us to transfer all of these estimates to the adaptive SMC algorithm that approximates these deterministic times.

3 Description of the models

For the convenience of the reader, in the first two sections 3.1, 3.2, we briefly introduce an abstract Feynman-Kac flow of distributions. The Feynman-Kac flow of distributions in excursion space associated to a prescribed sequence of resampling times is presented in Section 3.3.

The definition of the functional criteria discussed in this work is provided in Section 3.4, and the precise definition of the associated ‘limiting’ deterministic time parameters $(t_n)_{n \geq 0}$ discussed in the introduction is provided in Section 3.5.

3.1 Feynman-Kac distributions flow

We consider a sequence of measurable state spaces S_n , a probability measure η_0 on S_0 , and a sequence of Markov transitions $\mathcal{M}_n(x_{n-1}, dx_n)$ from S_{n-1} into S_n , with $n \geq 1$. Let $(\mathcal{X}_n)_{n \geq 0}$ be a Markov chain with initial distribution $\text{Law}(\mathcal{X}_0) = \eta_0$ on S_0 and elementary transitions

$$\mathbb{P}(\mathcal{X}_n \in dy \mid \mathcal{X}_{n-1} = x) = \mathcal{M}_n(x, dy).$$

Let $(\mathcal{G}_n)_{n \geq 0}$ be a sequence of non-negative and bounded potential-likelihood functions on S_n . To simplify the presentation, and to avoid unnecessary technicalities, it is supposed that $\mathcal{G}_0 = 1$ and for $n \geq 1$, $\mathcal{G}_n \in (0, 1)$ with

$$(\mathcal{G}) \quad q_n := \sup_{(x,y) \in S_n^2} (\mathcal{G}_n(x)/\mathcal{G}_n(y)) < \infty. \quad (3.1)$$

This regularity condition can be relaxed in various ways; see [4]. The Boltzmann-Gibbs transformation Ψ_n associated with a given potential function $\mathcal{G}_n$ is the mapping

$$\Psi_n : \mu \in \mathcal{P}(S_n) \mapsto \Psi_n(\mu) \in \mathcal{P}(S_n) \quad \text{with} \quad \Psi_n(\mu)(dx) := \frac{1}{\mu(\mathcal{G}_n)} \mathcal{G}_n(x) \mu(dx).$$

Notice that $\Psi_n(\mu)$ can be rewritten as a non-linear Markov transport equation

$$\Psi_n(\mu)(dy) = (\mu \mathcal{S}_{n,\mu})(dy) = \int_{S_n} \mu(dx) \mathcal{S}_{n,\mu}(x, dy)$$

with the accept-reject type selection transition

$$\mathcal{S}_{n,\mu}(x, dy) := \mathcal{G}_n(x) \delta_x(dy) + (1 - \mathcal{G}_n(x)) \Psi_n(\mu)(dy).$$

Let $(\eta_n, \hat{\eta}_n)_{n \geq 0}$ be the flow of probability measures, both starting at $\eta_0 = \hat{\eta}_0$, and defined for any $n \geq 1$ by the following recursion

$$\forall n \geq 0 \quad \eta_{n+1} = \hat{\eta}_n \mathcal{M}_{n+1} \quad \text{with} \quad \hat{\eta}_n := \Psi_n(\eta_n). \quad (3.2)$$

Notice that $(\eta_n)_{n \geq 0}$ and $(\hat{\eta}_n)_{n \geq 0}$ satisfy the nonlinear recursion

$$\eta_{n+1} = \Phi_{n+1}(\eta_n) := \Psi_n(\eta_n) \mathcal{M}_{n+1} \quad \text{and} \quad \hat{\eta}_{n+1} = \hat{\Phi}_{n+1}(\hat{\eta}_n) := \Psi_{n+1}(\hat{\eta}_n \mathcal{M}_{n+1}).$$

It is worth mentioning that the one step mapping $\hat{\Phi}_{n+1}$ can be rewritten as follows

$$\hat{\Phi}_{n+1}(\hat{\eta}_n) := \hat{\Psi}_n(\hat{\eta}_n) \hat{\mathcal{M}}_{n+1}$$

with the Boltzmann-Gibbs transformation $\hat{\Psi}_n$ associated with a given potential function $\hat{\mathcal{G}}_n = \mathcal{M}_{n+1}(\mathcal{G}_n)$ and the Markov transition $\hat{\mathcal{M}}_{n+1}$ given for any bounded measurable test function f on S_{n+1} , and any $x \in S_n$ by the formula

$$\hat{\mathcal{M}}_{n+1}(f)(x) = \frac{\mathcal{M}_{n+1}(\mathcal{G}_n f)(x)}{\mathcal{M}_{n+1}(\mathcal{G}_n)(x)}.$$

From the pure mathematical point of view, this shows that the flow $(\hat{\eta}_n)_{n \geq 0}$ has the same dynamical structure as the flow $(\eta_n)_{n \geq 0}$. Thus, w.l.o.g we shall concentrate our analysis on the flow of measures $(\eta_n)_{n \geq 0}$.

To get one step further, it can be readily checked that the flow of measures $(\eta_n)_{n \geq 0}$ satisfies the non-linear transport equation

$$\eta_{n+1} = \eta_n \mathcal{K}_{n+1, \eta_n} \quad \text{with} \quad \mathcal{K}_{n+1, \eta_n} := \mathcal{S}_{n, \eta_n} \mathcal{M}_{n+1}.$$

Also recall that the solution $(\eta_n, \hat{\eta}_n)$ of these recursive updating-prediction equations have the following functional representations

$$\eta_n(f_n) = \gamma_n(f_n) / \gamma_n(1) \quad \text{and} \quad \hat{\eta}_n(f_n) = \hat{\gamma}_n(f_n) / \hat{\gamma}_n(1) \quad (3.3)$$

with the unnormalized Feynman-Kac measures γ_n and $\hat{\gamma}_n$ defined by the formulae

$$\gamma_n(f_n) = \mathbb{E} [f_n(\mathcal{X}_n) \prod_{0 \leq k < n} \mathcal{G}_k(\mathcal{X}_k)] \quad \text{and} \quad \hat{\gamma}_n(f_n) = \gamma_n(f_n \mathcal{G}_n). \quad (3.4)$$

3.2 Feynman-Kac semi-groups

To analyze SMC methods, it is necessary to introduce the Feynman-Kac semi-group associated to the flow of measures $(\gamma_n)_{n \geq 0}$ and $(\eta_n)_{n \geq 0}$. Let us start by denoting by $\mathcal{Q}_{n+1}(x_n, dx_{n+1})$ the bounded integral operator from S_n into S_{n+1} ; this is defined by the elementary multiplicative formula

$$\mathcal{Q}_{n+1}(x_n, dx_{n+1}) = \mathcal{G}_n(x_n) \mathcal{M}_{n+1}(x_n, dx_{n+1}).$$

Let $(\mathcal{Q}_{p,n})_{0 \leq p \leq n}$ be the corresponding linear semi-group defined by

$$\mathcal{Q}_{p,n} = \mathcal{Q}_{p+1} \mathcal{Q}_{p+2} \dots \mathcal{Q}_n$$

with the convention $\mathcal{Q}_{n,n} = I$, the identity operator. Note that $\mathcal{Q}_{p,n}$ is alternatively defined by:

$$\mathcal{Q}_{p,n}(f_n)(x_p) = \mathbb{E} \left[f_n(\mathcal{X}_n) \prod_{p \leq k < n} \mathcal{G}_k(\mathcal{X}_k) \mid \mathcal{X}_p = x_p \right]. \quad (3.5)$$

Using the Markov property, it follows that

$$\gamma_n(f_n) = \mathbb{E} \left[\mathbb{E} \left[f_n(\mathcal{X}_n) \prod_{p \leq k < n} \mathcal{G}_k(\mathcal{X}_k) \mid \mathcal{X}_p \right] \prod_{0 \leq k < p} \mathcal{G}_k(\mathcal{X}_k) \right] = \gamma_p(\mathcal{Q}_{p,n}(f_n)).$$

The last assertion shows that $(\mathcal{Q}_{p,n})_{0 \leq p \leq n}$ is the semi-group associated with the unnormalized measures $(\gamma_n)_{n \geq 0}$. Denote its normalized version by

$$\mathcal{P}_{p,n}(f_n) := \frac{\mathcal{Q}_{p,n}(f_n)}{\mathcal{Q}_{p,n}(1)}. \quad (3.6)$$

Finally, denote by $(\Phi_{p,n})_{0 \leq p \leq n}$ the non-linear semi-group associated to the flow of normalized measures $(\eta_n)_{n \geq 0}$:

$$\Phi_{p,n} = \Phi_n \circ \dots \circ \Phi_{p+2} \circ \Phi_{p+1}$$

with the convention $\Phi_{n,n} = I$, the identity operator. From the previous discussion, note that $(\Phi_{p,n})_{0 \leq p \leq n}$ is alternatively defined in terms of $(\mathcal{Q}_{p,n})_{0 \leq p \leq n}$ with the following formulae:

$$\Phi_{p,n}(\eta_p)(f_n) = \frac{\gamma_p \mathcal{Q}_{p,n}(f_n)}{\gamma_p \mathcal{Q}_{p,n}(1)} = \frac{\eta_p \mathcal{Q}_{p,n}(f_n)}{\eta_p \mathcal{Q}_{p,n}(1)}. \quad (3.7)$$

3.3 Path space and excursion models

Let $(X'_n)_{n \geq 0}$ be a Markov chain taking values in some measurable state-spaces E'_n with elementary transitions

$$\mathbb{P}(X'_n \in dx'_n \mid X'_{n-1} = x'_{n-1}) := M'_n(x'_{n-1}, dx'_n)$$

and the initial distribution $\eta'_0 = \text{Law}(X'_0)$. In addition, introduce a sequence of potential functions G'_n on the state spaces E'_n . To simplify the presentation, it is assumed that the G'_n take values in $(0, 1)$.

Associate, with a prescribed non-decreasing sequence of time parameters $(t_n)_{n \geq 0}$, the excursion-valued random variables

$$X_0 = X'_0 \quad \text{and} \quad \forall n \geq 1 \quad X_n := (X'_{t_{n-1}+1}, X'_{t_{n-1}+2}, \dots, X'_{t_n})$$

and the random path sequences

$$\mathcal{X}_n := (X_0, \dots, X_n) = (X'_0, X'_1, \dots, X'_{t_n}) \in E_{t_n} := E'_0 \times \dots \times E'_{t_n}.$$

Note that $(\mathcal{X}_n)_{n \geq 0}$ forms a Markov chain

$$\mathcal{X}_{n+1} := (\mathcal{X}_n, X_{n+1}) = \left(\mathcal{X}_n, (X'_{t_n+1}, \dots, X'_{t_{n+1}}) \right) \quad (3.8)$$

taking values in the excursion spaces $S_n := E_{t_n}$. Furthermore, choosing, in (3.4), the potential functions

$$\forall n \geq 1 \quad \mathcal{G}_n(\mathcal{X}_n) = G_n(X_n) := \prod_{t_{n-1} < k \leq t_n} G'_k(X'_k) \quad (3.9)$$

then we readily find that

$$\begin{aligned} \gamma_n(f_n) &= \mathbb{E} \left[f_n(\mathcal{X}_n) \prod_{0 < k < n} \mathcal{G}_k(\mathcal{X}_k) \right] \\ &= \mathbb{E} \left[f_n(X_0, \dots, X_n) \prod_{0 < k < n} G_k(X_k) \right] \\ &= \mathbb{E} \left[f_n(X'_0, \dots, X'_{t_n}) \prod_{0 < k \leq t_{n-1}} G'_k(X'_k) \right]. \end{aligned}$$

By definition of the potential functions $\mathcal{G}_n$ of the excursion Feynman-Kac model (3.9), it is easily proved that the condition $(\mathcal{G})$ (eq. (3.1)) is satisfied, as soon as (G') introduced in (2.15) holds true. More precisely, it holds that

$$\begin{aligned} & (G') \\ & \Downarrow \\ & (\mathcal{G}) \text{ with } q_n \leq M' - \sup \left\{ \prod_{t_{n-1} < k \leq t_n} (G'_k(x'_k)/G'_k(y'_k)) \right\} \left(\leq \prod_{t_{n-1} < k \leq t_n} q'_k \right) \end{aligned}$$

where the essential supremum $M' - \sup \{.\}$ is taken over all M'_k -admissible paths $(x'_{t_{n-1}+1}, x'_{t_{n-1}+2}, \dots, x'_{t_n})$ and $(y'_{t_{n-1}+1}, y'_{t_{n-1}+2}, \dots, y'_{t_n})$ of the underlying Markov chain X'_k . Moreover, the marginal $\bar{\gamma}_n$ of γ_n on the state space $\bar{S}_n := (E'_{t_n+1} \times \dots \times E'_{t_{n+1}})$ is again given, for any $\bar{f}_n \in \mathcal{B}_b(\bar{S}_n)$, by

$$\bar{\gamma}_n(\bar{f}_n) := \mathbb{E} \left[\bar{f}_n(X_n) \prod_{0 < k < n} G_k(X_k) \right].$$

3.4 Functional criteria

In Section 3.3, we have assumed that a non-decreasing sequence of time parameters $(t_n)_{n \geq 0}$ was available. We now introduce the functional criteria which will be used to build this sequence. To connect the empirical criteria with their limiting functional versions, the latter need to satisfy some weak regularity conditions which are given below.

Definition 3.1 *We consider a sequence of functional criteria*

$$\forall n \geq 0 \quad \forall p \leq q \quad \mathcal{H}_{p,q}^{(n)} : \mu \in \mathcal{P}(E_q) \mapsto \mathcal{H}_{p,q}^{(n)}(\mu) \in \mathbb{R}_+$$

satisfying the following Lipschitz type regularity condition

$$\left| \mathcal{H}_{p,q}^{(n)}(\mu_1) - \mathcal{H}_{p,q}^{(n)}(\mu_2) \right| \leq \int |\mu_1 - \mu_2|(h) |H_{p,q}^{(n)}(dh)| \quad (3.10)$$

for some collection of bounded measures $H_{p,q}^{(n)}$ on $\mathcal{B}_b(E_q)$ such that

$$\delta(H_{p,q}^{(n)}) := \sup_{h \in \text{Osc}_1(E_q)} \int \text{osc}(h) H_{p,q}^{(n)}(dh) < \infty.$$

To illustrate this rather abstract construction, we provide four examples of functional criteria corresponding those discussed in Section 2.4.1. All of them are defined in terms of the importance weight function

$$W_{p,q} : (x'_0, \dots, x'_q) \in E_q \mapsto W_{p,q}(x'_0, \dots, x'_q) := \prod_{p < k \leq q} G'_k(x'_k) \in (0, 1)$$

In all the situations described below, the measures $H_{t_n,s}^{(n)}$ in (3.10) have a finite support with one or two elements.

1. **Weights thresholds control:** The functional defined below

$$\mathcal{H}_{p,q}^{(n)}(\mu) = \mu(A_{p,q}^{(n)}) \quad \text{with} \quad A_{p,q}^{(n)} := \{x_q \in E_q : W_{p,q}(x_q) \geq s_n\}$$

measures the proportion of the path sequences with a weight product larger than some parameter s_n .

2. **Normalizing constants:** In some sense, the functional defined below

$$\mathcal{H}_{p,q}^{(n)}(\mu) = \mu(W_{p,q})$$

measures the degeneracy of the normalizing constants for large values of the parameter $(q - p)$.

3. **Variance of the weights:** The functional defined below

$$\mathcal{H}_{p,q}^{(n)}(\mu) = \mu \left(\left[\frac{W_{p,q}}{\mu(W_{p,q})} - 1 \right]^2 \right)$$

coincides with the variance of the weights w.r.t the reference measure μ .

4. **Boltzmann's entropy:** The functional defined below

$$\mathcal{H}_{p,q}^{(n)}(\mu) = \text{Ent}(d\mu \mid W_{p,q} d\mu) := -\mu(\log W_{p,q})$$

measures the relative entropy distance between the reference measure μ and the updated weighted measure.

3.5 Resampling time construction

We now explain, precisely, how to define the sequence of resampling times $(t_n)_{n \geq 0}$.

Definition 3.2 For any pair of integers $0 \leq p \leq n$ and any $\eta \in \mathcal{P}(E_p)$, we denote by $\mathbb{P}_{\eta,(p,n)}$ the law of a random path sequence of length n , starting with an initial distribution η on the path space E_p and performing $(n-p)$ elementary moves with the local probability transitions M'_k , with $p < k \leq n$. More precisely, $\mathbb{P}_{\eta,(p,n)}$ is the measure defined by

$$\begin{aligned} \mathbb{P}_{\eta,(p,n)}(d(x_p, x'_{p+1}, \dots, x'_n)) &:= \eta(dx_p) M'_{p+1}(x'_p, dx'_{p+1}) \dots M'_n(x'_{n-1}, dx'_n) \\ &\in \mathcal{P}(E_p \times (E'_{p+1} \times \dots \times E'_n)) = \mathcal{P}(E_n) \end{aligned}$$

where $dx_p = d(x'_0, \dots, x'_p)$ denotes an infinitesimal neighborhood of a path sequence

$$x_p = (x'_0, \dots, x'_p) \in E_p = (E'_0 \times \dots \times E'_p).$$

Given $\mathcal{H}_{p,q}^{(n)}$, with $n \geq 0$ and $0 \leq p \leq q$, we define a non-decreasing sequence of deterministic time steps $(t_n)_{n \geq 0}$ and a flow of Feynman-Kac measures $(\eta_n, \hat{\eta}_n)$ as follows:

- Set $n = 0$ and $t_n = 0$ and $\hat{\eta}_0 = \eta_0 = \eta'_0$.
- Suppose that the time step t_n is defined as well as the pair of measures $(\eta_n, \hat{\eta}_n) \in \mathcal{P}(E_{t_n})^2$ at a given rank n . To define the time step t_{n+1} and the pair of measures $(\eta_{n+1}, \hat{\eta}_{n+1}) \in \mathcal{P}(E_{t_{n+1}})^2$, we consider $\mathcal{H}_{t_n,s}^{(n)}$, with $s \geq t_n$ and a given interval $I_n \subset \mathbb{R}_+$ of the form $I_n = [a_n, \infty)$ or $I_n = (-\infty, a_n]$ such that

$$\mathcal{H}_{t_n,t_n}^{(n)}(\mathbb{P}_{\hat{\eta}_n,(t_n,t_n)}) \notin I_n.$$

The resampling time t_{n+1} is defined as the first time $s \geq t_n$ the quantity $\mathcal{H}_{t_n,s}^{(n)}(\mathbb{P}_{\hat{\eta}_n,(t_n,s)})$ hits the set I_n ; that is

$$t_{n+1} := \inf \left\{ t_n \leq s : \mathcal{H}_{t_n,s}^{(n)}(\mathbb{P}_{\hat{\eta}_n,(t_n,s)}) \in I_n \right\}$$

Given $t_{n+1}(> t_n)$, set

$$\eta_{n+1} = \mathbb{P}_{\hat{\eta}_n, (t_n, t_{n+1})} \quad \text{and} \quad \hat{\eta}_{n+1} = \Psi_{n+1}(\eta_{n+1}) \quad (3.11)$$

with the Boltzmann-Gibbs transformation Ψ_{n+1} associated with the potential function $\mathcal{G}_{n+1} = W_{t_n, t_{n+1}}$.

We end this construction with an important observation. By definition of the Markov transition $\mathcal{M}_{n+1}$ of the excursion model $\mathcal{X}_n$ defined in Section 3.3, it can be checked that

$$\eta_{n+1} = \mathbb{P}_{\hat{\eta}_n, (t_n, t_{n+1})} = \hat{\eta}_n \mathcal{M}_{n+1}. \quad (3.12)$$

This yields the recursion

$$(3.11) \quad \text{and} \quad (3.12) \implies \eta_{n+1} = \Psi_n(\eta_n) \mathcal{M}_{n+1}.$$

Hence the flow of measures η_n and $\hat{\eta}_n$ coincide with the Feynman-Kac flow of distributions defined in (3.3) with the Markov chain and potential function $(\mathcal{X}_n, \mathcal{G}_n)$ on excursion spaces defined in (3.8) and (3.9). The only difference between these models and the ones discussed in Section 3.3 comes from the fact that the sequence of time parameters $(t_n)_{n \geq 0}$ is inductively defined in terms of a prescribed sequence of functional criteria. The SMC approximation of these distributions is discussed in Section 4.

3.6 Some applications

In this Section, we examine, in detail, the inductive construction of the deterministic resampling times t_n introduced in Section 3.5 for the criteria 1)-4) in Section 2.3.

3.6.1 Adaptive control of the weights

Consider $[0, 1)$ -valued potential functions. The resampling times are built to control the degeneracy of the Feynman-Kac weights through

$$\begin{aligned} \mathcal{H}_{t_n, s}^{(n)}(\mathbb{P}_{\hat{\eta}_n, (t_n, s)}) &= \mathbb{E}_{n, \hat{\eta}_n} \left[1_{A_{t_n, s}^{(n)}}(X'_{t_n+1}, \dots, X'_s) \right] \\ &= \mathbb{P}_{n, \hat{\eta}_n} \left(\prod_{t_n < k \leq s} G'_k(X'_k) \geq s_n \right) \end{aligned}$$

where $0 < s_n < 1$ and

$$A_{t_n, s}^{(n)} := \left\{ (x'_{t_n+1}, \dots, x'_s) \in E'_{t_n+1} \times \dots \times E'_s : \prod_{t_n < k \leq s} G'_k(x'_k) \geq s_n \right\}.$$

Here $\mathbb{E}_{n, \hat{\eta}_n}$ denotes the expectation w.r.t the law $\mathbb{P}_{n, \hat{\eta}_n}$ of a free evolution X'_k starting at time t_n at the end point X'_{t_n} of a random path $(X_0, \dots, X_n)$ with distribution $\hat{\eta}_n$ on the space E_n .

The mappings

$$t_n \leq s \mapsto \mathbb{P}_{n, \hat{\eta}_n} \left(\prod_{t_n < k \leq s} G'_k(X'_k) \geq s_n \right)$$

are clearly non-increasing functions starting at 1 for $s = t_n$. If we choose an interval

$$I_n := (-\infty, a_n] \quad \text{for some } a_n < 1$$

then the resampling time t_{n+1} coincides with the first time s the sequence of probabilities

$$s \mapsto \mathbb{P}_{n, \hat{\eta}_n} \left(\prod_{t_n < k \leq s} G'_k(X'_k) \geq s_n \right)$$

goes below the level a_n ; that is

$$t_{n+1} := \inf \left\{ t_n \leq s : \mathbb{P}_{n, \hat{\eta}_n} \left(\prod_{t_n < k \leq s} G'_k(X'_k) \geq s_n \right) \leq a_n \right\}.$$

This strategy allows us to decompose the trajectories of the reference free motion X'_k into a sequence of excursion-processes

$$X_0 = X'_0 \quad \text{and} \quad \forall n \geq 1 \quad X_n := (X'_{t_{n-1}+1}, X'_{t_{n-1}+2}, \dots, X'_{t_n}) \quad (3.13)$$

with

$$\forall s < t_{n+1} \quad \mathbb{P}_{n, \hat{\eta}_n} \left(\prod_{t_n < k \leq s} G'_k(X'_k) \geq s_n \right) > a_n$$

and

$$s = t_{n+1} \implies \mathbb{P}_{n, \hat{\eta}_n} \left(\prod_{t_n < k \leq t_{n+1}} G'_k(X'_k) \geq s_n \right) \leq a_n.$$

We illustrate this construction with the random walk model discussed in (2.9) (Section 2.3.1). In this situation, recall that X'_k is a Markov chain on $E'_k = \mathbb{Z}$ starting at the origin $X'_0 = 0$. If the indicator potential functions, $G'_k = 1_{[k, \infty)}$, are chosen then

$$\mathbb{E}_{0, \hat{\eta}_0} \left(\prod_{0 < k \leq t_1} G'_k(X'_k) \right) = \mathbb{P}(X'_1 \geq 1, \dots, X'_k \geq k).$$

When $a_0 = 1/2$, we find that t_1 coincides with the first level k the chain X'_k can reach with a probability less than $1/2$ with $X'_0 = 0$; more formally, we have

$$t_1 = \inf \{k \geq 0 : \mathbb{P}(X'_0 \geq 0, X'_1 \geq 1, \dots, X'_k \geq k) \leq 1/2\}.$$

Similarly, when $a_n = 1/2$, t_{n+1} coincides with the first level k the chain X'_k can reach with a probability less than $1/2$ with $X'_{t_n} = x'_{t_n}$; more formally, we have

$$t_{n+1} = \inf \{k \geq t_n : \mathbb{P}_{n, \hat{\eta}_n} (X'_{t_n+1} \geq t_n + 1, \dots, X'_k \geq k) \leq 1/2\}.$$

3.6.2 Normalizing constants

The resampling times are built to control the degeneracy of the normalizing constants estimates. To be more precise, set

$$\begin{aligned} \mathcal{H}_{t_n, s}^{(n)}(\mathbb{P}_{\hat{\eta}_n, (t_n, s)}) &= \mathbb{E}_{n, \hat{\eta}_n} \left(\prod_{t_n < k \leq s} G'_k(X'_k) \right) \\ &:= \int \hat{\eta}_n(dx_{t_n}) \mathbb{E}_{t_n, x'_{t_n}} \left(\prod_{t_n < k \leq s} G'_k(X'_k) \right) \\ &= \int \hat{\eta}_n(dx_{t_n}) M'_{t_n+1}(x'_{t_n}, dx'_{t_n+1}) \dots \\ &\quad \dots \times M'_s(x'_{s-1}, dx'_s) \prod_{t_n < k \leq s} G'_k(x'_k). \end{aligned}$$

We have

$$s = t_n \implies \mathcal{H}_{t_n, t_n}^{(n)}(\mathbb{P}_{\hat{\eta}_n, (t_n, t_n)}) = 1.$$

When the potential functions take values in $[0, 1]$, one concludes that the mapping

$$s \geq t_n \mapsto \mathcal{H}_{t_n, s}^{(n)}(\mathbb{P}_{\hat{\eta}_n, (t_n, s)})$$

is clearly non-increasing starting at 1 at time $s = t_n$. If we choose an interval

$$I_n := (-\infty, a_n] \quad \text{for some } a_n < 1$$

then the resampling time t_{n+1} coincides with the first time s the quantities

$$s \mapsto \mathbb{E}_{n, \hat{\eta}_n} \left(\prod_{t_n < k \leq s} G'_k(X'_k) \right)$$

go below the level a_n ; that is, we have

$$t_{n+1} := \inf \left\{ t_n \leq s : \mathbb{E}_{n, \hat{\eta}_n} \left(\prod_{t_n < k \leq s} G'_k(X'_k) \right) \leq a_n \right\}.$$

This criterion allows us to decompose the trajectories of the reference free motion X'_k into a sequence of excursion like processes

$$X_0 = X'_0 \quad \text{and} \quad \forall n \geq 1 \quad X_n := (X'_{t_{n-1}+1}, X'_{t_{n-1}+2}, \dots, X'_{t_n}) \quad (3.14)$$

with

$$\forall s < t_{n+1} \quad \mathbb{E}_{n, \hat{\eta}_n} \left(\prod_{t_n < k \leq s} G'_k(X'_k) \right) > a_n$$

and

$$\mathbb{E}_{n, \hat{\eta}_n} \left(\prod_{t_n < k \leq t_{n+1}} G'_k(X'_k) \right) \leq a_n.$$

3.6.3 Variance of the weights

The resampling times are built to control the variance of the weights. To be more precise, set

$$\begin{aligned}\mathcal{H}_{t_n, s}^{(n)}(\mathbb{P}_{\hat{\eta}_n, (t_n, s)}) &= \mathbb{E}_{n, \hat{\eta}_n} \left(\left(\frac{\prod_{t_n < k \leq s} G'_k(X'_k)}{\mathbb{E}_{n, \hat{\eta}_n} \left(\prod_{t_n < k \leq s} G'_k(X'_k) \right)} - 1 \right)^2 \right) \\ &= \frac{\mathbb{E}_{n, \hat{\eta}_n} \left(\prod_{t_n < k \leq s} G'_k(X'_k)^2 \right)}{\mathbb{E}_{n, \hat{\eta}_n} \left(\prod_{t_n < k \leq s} G'_k(X'_k) \right)^2} - 1.\end{aligned}$$

In this case,

$$s = t_n \implies \mathcal{H}_{t_n, t_n}^{(n)}(\mathbb{P}_{\hat{\eta}_n, (t_n, t_n)}) = 0.$$

The variance mappings $s \mapsto \mathcal{H}_{t_n, s}^{(n)}(\mathbb{P}_{\hat{\eta}_n, (t_n, s)})$ are generally increasing. For instance, in the case of indicator potential functions $G'_k = 1_{A'_k}$ we have

$$\begin{aligned}\forall t_n \leq s \leq r \quad \mathcal{H}_{t_n, s}^{(n)}(\mathbb{P}_{\hat{\eta}_n, (t_n, s)}) &= \mathbb{E}_{n, \hat{\eta}_n} \left(\prod_{t_n < k \leq s} G'_k(X'_k) \right)^{-1} - 1 \\ &\leq \mathcal{H}_{t_n, r}^{(n)}(\mathbb{P}_{\hat{\eta}_n, (t_n, r)}).\end{aligned}$$

For another example in filtering problems, we refer the reader to [5]. One natural way to control these variances is to choose an interval

$$I_n := [a_n, \infty) \quad \text{for some } a_n > 0$$

then the resampling time t_{n+1} coincides with the first time s the function

$$s \mapsto \mathbb{E}_{n, \hat{\eta}_n} \left(\prod_{t_n < k \leq s} G'_k(X'_k)^2 \right) / \mathbb{E}_{n, \hat{\eta}_n} \left(\prod_{t_n < k \leq s} G'_k(X'_k) \right)^2$$

goes above the level $1 + a_n$; that is

$$\begin{aligned}t_{n+1} &:= \inf \left\{ t_n \leq s : \mathbb{E}_{n, \hat{\eta}_n} \left(\prod_{t_n < k \leq s} G'_k(X'_k)^2 \right) \right. \\ &\quad \left. \geq [1 + a_n] \mathbb{E}_{n, \hat{\eta}_n} \left(\prod_{t_n < k \leq s} G'_k(X'_k) \right)^2 \right\}.\end{aligned}$$

This construction allows us to decompose the trajectories of the reference free motion X'_k into a sequence of excursion like processes

$$X_0 = X'_0 \quad \text{and} \quad \forall n \geq 1 \quad X_n := (X'_{t_{n-1}+1}, X'_{t_{n-1}+2}, \dots, X'_{t_n}) \quad (3.15)$$

with controlled Feynman-Kac weights variances

$$\forall t_n \leq s < t_{n+1} \quad \frac{\mathbb{E}_{n, \hat{\eta}_n} \left(\prod_{t_n < k \leq s} G'_k(X'_k)^2 \right)}{\mathbb{E}_{n, \hat{\eta}_n} \left(\prod_{t_n < k \leq s} G'_k(X'_k) \right)^2} < 1 + a_n$$

and for $s = t_{n+1}$

$$\frac{\mathbb{E}_{n,\hat{\eta}_n} \left(\prod_{t_n < k \leq t_{n+1}} G'_k(X'_k)^2 \right)}{\mathbb{E}_{n,\hat{\eta}_n} \left(\prod_{t_n < k \leq t_{n+1}} G'_k(X'_k) \right)^2} = \frac{\mathbb{E}_{n,\hat{\eta}_n} (G_{n+1}(X_{n+1})^2)}{\mathbb{E}_{n,\hat{\eta}_n} (G_{n+1}(X_{n+1}))^2} \geq 1 + a_n.$$

3.6.4 Boltzmann's entropy

This criterion allows us to control the entropy like distance between the free motion trajectories and the weighted Feynman-Kac measures. To be more precise, set

$$\mathcal{H}_{t_n,s}^{(n)}(\mathbb{P}_{\hat{\eta}_n,(t_n,s)}) = \text{Ent}(\mathbb{P}_{\hat{\eta}_n,(t_n,s)} \mid \mathbb{Q}_{\hat{\eta}_n,(t_n,s)}) = - \sum_{t_n < k \leq s} \mathbb{E}_{n,\hat{\eta}_n}(\log G'_k(X'_k))$$

with the weighted measures $\mathbb{Q}_{\eta,(p,n)}$ defined by

$$\mathbb{Q}_{\eta,(p,n)}(d(x_p, x'_{p+1}, \dots, x'_n)) = \mathbb{P}_{\eta,(p,n)}(d(x_p, x'_{p+1}, \dots, x'_n)) \times \prod_{p < k \leq n} G'_k(x'_k).$$

For $(0,1)$ -valued potential functions, the mappings

$$t_n \leq s \mapsto \text{Ent}(\mathbb{P}_{\hat{\eta}_n,(t_n,s)} \mid \mathbb{Q}_{\hat{\eta}_n,(t_n,s)})$$

are clearly non-decreasing functions starting at 0 for $s = t_n$. If we choose an interval

$$I_n := [a_n, \infty) \quad \text{for some } a_n > 0$$

then the resampling time t_{n+1} coincides with the first time s the entropy distance goes above the level a_n ; that is,

$$t_{n+1} := \inf \{ t_n \leq s : \text{Ent}(\mathbb{P}_{\hat{\eta}_n,(t_n,s)} \mid \mathbb{Q}_{\hat{\eta}_n,(t_n,s)}) \geq a_n \}.$$

This construction allows us to decompose the trajectories of the reference free motion X'_k into a sequence of excursion like processes

$$X_0 = X'_0 \quad \text{and} \quad \forall n \geq 1 \quad X_n := (X'_{t_{n-1}+1}, X'_{t_{n-1}+2}, \dots, X'_{t_n}) \quad (3.16)$$

with

$$\forall s < t_{n+1} \quad \text{Ent}(\mathbb{P}_{\hat{\eta}_n,(t_n,s)} \mid \mathbb{Q}_{\hat{\eta}_n,(t_n,s)}) < a_n$$

and

$$\text{Ent}(\mathbb{P}_{\hat{\eta}_n,(t_n,s)} \mid \mathbb{Q}_{\hat{\eta}_n,(t_n,s)}) \geq a_n.$$

4 Convergence analysis of the reference SMC algorithm

In Section 4.1, we present the reference SMC algorithm which resamples at the deterministic times $(t_n)_{n \geq 0}$. The concentration analysis of this algorithm is developed in Section 4.2. In Section 4.2.1, we provide a brief discussion on the set of constants arising in our estimates and introduce assumptions ensuring a uniform control of these constants w.r.t the time horizon. In Section 4.2.2,

several estimates on the bias and $\mathbb{L}_m$ -mean errors are provided, whereas Section 4.2.3 gives an original non-asymptotic exponential concentration estimate. In practice the times $(t_n)_{n \geq 0}$ are unknown and are approximated, on-line, using the current SMC approximation. In Section 4.3, we use the concentration results obtained in Section 4.2 to analyze the convergence of these empirical approximations towards their limiting functional values.

4.1 A reference SMC algorithm

The SMC interpretation of the evolution equation (3.2) is the Markov chain

$$\mathcal{X}_n^{(N)} = \left(\mathcal{X}_n^{(N,1)}, \mathcal{X}_n^{(N,2)}, \dots, \mathcal{X}_n^{(N,N)} \right) \in S_n^N$$

with elementary transitions

$$\mathbb{P} \left(\mathcal{X}_{n+1}^{(N)} \in dx_{n+1} \mid \mathcal{X}_n^{(N)} \right) = \prod_{i=1}^N \mathcal{K}_{n+1, \eta_n^N}(\mathcal{X}_n^{(N,i)}, dx_{n+1}^i) \quad (4.1)$$

with

$$\mathcal{K}_{n+1, \eta_n^N} = \mathcal{S}_{n, \eta_n^N} \mathcal{M}_{n+1} \quad \text{and} \quad \eta_n^N(\cdot) := \frac{1}{N} \sum_{j=1}^N \delta_{\mathcal{X}_n^{(N,j)}}(\cdot). \quad (4.2)$$

Loosely speaking this integral decomposition shows that this SMC algorithm has the same updating/prediction nature as the one of the ‘limiting’ Feynman-Kac model. More precisely, the deterministic two-step updating/prediction transitions in distribution spaces

$$\eta_n \xrightarrow{\mathcal{S}_{n, \eta_n}} \hat{\eta}_n = \eta_n \mathcal{S}_{n, \eta_n} = \Psi_n(\eta_n) \xrightarrow{\mathcal{M}_{n+1}} \eta_{n+1} = \hat{\eta}_n \mathcal{M}_{n+1} \quad (4.3)$$

have been replaced by a two-step resampling/mutation transition in a product space

$$\mathcal{X}_n^{(N)} \in E_{t_n}^N \xrightarrow{\text{resampling}} \hat{\mathcal{X}}_n^{(N)} \in E_{t_n}^N \xrightarrow{\text{mutation}} \mathcal{X}_{n+1}^{(N)} \in E_{t_{n+1}}^N \quad (4.4)$$

where it is recalled that the state-space at the $(n+1)$ -th iteration is the path space given by

$$E_{t_{n+1}} := \left(E'_0 \times \dots \times E'_{t_{n+1}} \right) = E_{t_n} \times \left(E'_{t_{n+1}} \times \dots \times E'_{t_{n+1}} \right).$$

In our context, the SMC algorithm keeps track of all the paths of the sampled particles and the corresponding ancestral lines are denoted as follows

$$\begin{aligned} \hat{\mathcal{X}}_n^{(N,i)} &= \left(\hat{X}_{0,t_n}^{(N,i)}, \hat{X}_{1,t_n}^{(N,i)}, \dots, \hat{X}_{t_n,t_n}^{(N,i)} \right) \\ \mathcal{X}_n^{(N,i)} &= \left(X_{0,t_n}^{(N,i)}, X_{1,t_n}^{(N,i)}, \dots, X_{t_n,t_n}^{(N,i)} \right) \in E_{t_n} := \left(E'_0 \times \dots \times E'_{t_n} \right). \end{aligned}$$

By definition of the reference Markov model $\mathcal{X}_n$ given in (3.8), every path particle $\mathcal{X}_{n+1}^{(N,i)} \in E_{t_{n+1}}$ keeps track of the selected excursion $\hat{\mathcal{X}}_n^{(N,i)} \in E_{t_n}$ and it

evolves its terminal state $\widehat{X}_{t_n, t_n}^{(N, i)}$ with $(t_{n+1} - t_n)$ elementary random moves with the Markov transitions M'_k . More formally, we have that

$$\begin{aligned} \mathcal{X}_{n+1}^{(N, i)} &= \left(\underbrace{\left(X_{0, t_{n+1}}^{(N, i)}, X_{1, t_{n+1}}^{(N, i)}, \dots, X_{t_n, t_{n+1}}^{(N, i)} \right)}_{\widehat{\mathcal{X}}_n^{(N, i)}}, \underbrace{\left(X_{t_n+1, t_{n+1}}^{(N, i)}, \dots, X_{t_{n+1}, t_{n+1}}^{(N, i)} \right)}_{\left(X_{t_n+1}^{(N, i)}, \dots, X_{t_{n+1}}^{(N, i)} \right)} \right) \\ &= \left(\widehat{\mathcal{X}}_n^{(N, i)}, \left(X_{t_n+1}^{(N, i)}, \dots, X_{t_{n+1}}^{(N, i)} \right) \right) \end{aligned}$$

with the $(t_{n+1} - t_n)$ M'_k -exploration transitions given by the following synthetic diagram:

$$\widehat{X}_{t_n, t_n}^{(N, i)} \xrightarrow{M'_{t_n+1}} X_{t_n+1}^{(N, i)} \xrightarrow{M'_{t_n+2}} X_{t_n+2}^{(N, i)} \xrightarrow{M'_{t_n+3}} \dots \xrightarrow{M'_{t_{n+1}}} X_{t_{n+1}}^{(N, i)}.$$

From this discussion, it is worth mentioning a further convention that the particle empirical measures

$$\eta_{n+1}^N(\cdot) = \frac{1}{N} \sum_{i=1}^N \delta_{\left(\widehat{\mathcal{X}}_n^{(N, i)}, \left(X_{t_n+1}^{(N, i)}, X_{t_n+2}^{(N, i)}, \dots, X_{t_{n+1}}^{(N, i)} \right) \right)}(\cdot)$$

are the terminal values at time $s = t_{n+1}$ of the flow of random measures

$$t_n \leq s \leq t_{n+1} \mapsto \mathbb{P}_{\widehat{\eta}_n^N, (t_n, s)}^N(\cdot) = \frac{1}{N} \sum_{i=1}^N \delta_{\left(\widehat{\mathcal{X}}_n^{(N, i)}, \left(X_{t_n+1}^{(N, i)}, X_{t_n+2}^{(N, i)}, \dots, X_s^{(N, i)} \right) \right)}(\cdot). \quad (4.5)$$

4.2 Concentration analysis

4.2.1 Introduction

This Section is concerned with the concentration analysis of the empirical measures η_n^N associated with (4.2) around their limiting values η_n defined in (3.3). Our concentration estimates are expressed in terms of the following constants

$$q_{p, n} = \sup_{(x, y) \in S_p^2} \frac{\mathcal{Q}_{p, n}(1)(x)}{\mathcal{Q}_{p, n}(1)(y)} \quad \text{and} \quad \beta(\mathcal{P}_{p, n}) := \sup_{f \in \text{Osc}_1(S_n)} \text{osc}(\mathcal{P}_{p, n}(f))$$

where $\mathcal{Q}_{p, n}$ is the linear semi-group (3.5) and $\mathcal{P}_{p, n}$ its normalized version (3.6). Note that these parameters can be expressed in terms of the mixing properties of the Markov transitions $\mathcal{M}_n$. It is clearly out of the scope of the present article to derive these estimates; we refer the interested reader to chapter 4 of [4]. Nevertheless, under appropriate mixing type properties we can prove that the series $\sum_{p=0}^n q_{p, n}^\alpha \beta(\mathcal{P}_{p, n})$ is uniformly bounded with respect to the final time horizon n , for any parameter $\alpha \geq 0$. Most of the results presented in this Section are expressed in terms of these series, including bias estimates, $\mathbb{L}_m$ -mean error bounds, as well as exponential concentration estimates. As a result, these non-asymptotic results can be converted into uniform in time convergence results. To get a flavor of these uniform estimates, let us assume that the Markov transitions $\mathcal{M}_k$ satisfy the following regularity property

$(\mathcal{M})_m$ There exists an $m \in \mathbb{Z}_+$ and a sequence $(\delta_p)_{p \geq 0} \in (0, 1)^\mathbb{N}$ such that

$$\forall p \geq 0 \quad \forall (x, y) \in S_p^2 \quad \mathcal{M}_{p,p+m}(x, \cdot) \geq \delta_p \mathcal{M}_{p,p+m}(y, \cdot).$$

We also introduce the following quantities:

$$\forall k \leq l \quad r_{k,l} := \sup \prod_{k \leq p < l} \frac{\mathcal{G}_p(x_p)}{\mathcal{G}_p(y_p)} \left(\leq \prod_{k \leq p < l} q_p \right)$$

with the collection of constants $(q_n)_{n \geq 1}$ introduced in (3.1). In the above displayed formula, the supremum is taken over all admissible pair of paths with elementary transitions $\mathcal{M}_p$.

Under the condition $(\mathcal{M})_m$ we have for any $n \geq m \geq 1$, and $p \geq 1$,

$$q_{p,p+n} \leq \delta_p^{-1} r_{p,p+m} \quad \text{and} \quad \beta(\mathcal{P}_{p,p+n}) \leq \prod_{k=0}^{\lfloor n/m \rfloor - 1} \left(1 - \delta_p^2 r_{p+km+1, p+(k+1)m}^{-1} \right). \quad (4.6)$$

The proof of these estimates relies on technical semi-group techniques; see chapter 4 of [4] for details. Several contraction inequalities can be deduced from these results. To understand this more closely, assume that $(\mathcal{M})_m$ is satisfied with $m = 1$, $\delta = \wedge_n \delta_n > 0$ and $q = \vee_{n \geq 1} q_n$. In this case,

$$q_{p,p+n} \leq \delta^{-1} q \quad \text{and} \quad \beta(P_{p,p+n}) \leq (1 - \delta^2)^n$$

$$\Downarrow$$

$$\forall \alpha \geq 0 \quad \sum_{p=0}^n q_{p,n}^\alpha \beta(\mathcal{P}_{p,n}) \leq q^\alpha / \delta^{(2+\alpha)}.$$

More generally, assume $(\mathcal{M})_m$ is satisfied for some $m \geq 1$ and that the parameters δ_p and $r_{k,l}$ are such that

$$\wedge \delta_p := \delta > 0 \quad \vee_p r_{p,p+m} := \bar{r} < \infty \quad \text{and} \quad \vee_p r_{p+1,p+m} := \underline{r} < \infty. \quad (4.7)$$

In this situation,

$$q_{p,p+n} \leq \delta^{-1} \bar{r} \quad \text{and} \quad \beta(P_{p,p+n}) \leq (1 - \delta^2 \underline{r}^{-1})^{\lfloor n/m \rfloor}$$

and therefore

$$\forall \alpha \geq 0 \quad \sum_{p=0}^n q_{p,n}^\alpha \beta(\mathcal{P}_{p,n}) \leq m \underline{r} \bar{r}^\alpha / \delta^{(2+\alpha)}. \quad (4.8)$$

It is well known that the mixing type condition $(\mathcal{M})_m$ is satisfied for any aperiodic and irreducible Markov chains on finite spaces, as well as for bi-Laplace exponential transitions associated with a bounded drift function. We also mention that this mixing condition is of course never met for the complete historical process $\mathcal{X}_n = (X_p')_{0 \leq p \leq t_n}$ on $S_n = \prod_{0 \leq p \leq t_n} E_p'$ discussed in Section 3.3.

Nevertheless, under appropriate conditions on the Markov transitions M'_k , it is satisfied for the time marginal model associated with the excursion valued Markov chain model on $\prod_{t_{n-1} < p \leq t_n} E_p'$. For instance, if

$$\forall k \geq 1 \quad \forall (x, y) \in (E_p')^2 \quad M'_k(x, \cdot) \geq \delta' M'_k(y, \cdot)$$

for some $\delta' > 0$, then condition $(\mathcal{M})_m$ is met with $m = 1$ and $\delta_p = \delta'$.

4.2.2 Some $\mathbb{L}_m$ -mean error bounds

At this point, it is convenient to observe that the local sampling errors induced by the mean field particle model are expressed in terms of the collection of local random field models defined below.

Definition 4.1 For any $n \geq 0$ and any $N \geq 1$, let V_n^N be the collection of random fields defined by the following stochastic perturbation formulae

$$\eta_n^N = \eta_{n-1}^N \mathcal{K}_{n, \eta_{n-1}^N} + \frac{1}{\sqrt{N}} V_n^N \left(\iff V_n^N := \sqrt{N} \left[\eta_n^N - \eta_{n-1}^N \mathcal{K}_{n, \eta_{n-1}^N} \right] \right). \quad (4.9)$$

For $n = 0$, the conventions $K_{0, \eta_{-1}^N}(x, dy) = \eta_0(dy)$ and $\eta_{-1}^N \mathcal{K}_{0, \eta_{-1}^N} = \eta_0$ are adopted.

In order to quantify high order $\mathbb{L}_m$ -mean errors we need the following Kintchine type inequality for martingales with symmetric and independent increments. This result is more or less well known. For the convenience of the reader, a proof is provided in the appendix.

Lemma 4.2 (Kintchine's inequality) Let $M_n := \sum_{0 \leq p \leq n} \Delta_p$ be a real-valued martingale with symmetric and independent increments $(\Delta_n)_{n \geq 0}$. For any integer $m \geq 1$, and any $n \geq 0$, we have

$$\mathbb{E}(|M_n|^m)^{\frac{1}{m}} \leq b(m) \mathbb{E}([M]_n^{m'/2})^{1/m'} \quad \text{with} \quad [M]_n := \sum_{0 \leq p \leq n} \Delta_p^2 \quad (4.10)$$

where m' stands for the smallest even integer $m' \geq m$ and $(b(m))_{m \geq 1}$ is the collection of constants given below:

$$b(2m)^{2m} := (2m)_m 2^{-m} \quad \text{and} \quad b(2m+1)^{2m+1} := \frac{(2m+1)_{(m+1)}}{\sqrt{m+1/2}} 2^{-(m+1/2)}. \quad (4.11)$$

Proposition 4.3 For any $N \geq 1$, $m \geq 1$, $n \geq 0$ and any test function $f_n \in \mathcal{B}_b(S_n)$ we have the almost sure estimate

$$\mathbb{E} \left(|V_n^N(f_n)|^m \middle| \mathcal{F}_{n-1}^{(N)} \right)^{\frac{1}{m}} \leq b(m) \text{osc}(f_n). \quad (4.12)$$

Proof:

By construction, we have

$$V_n^N(f_n) = \sum_{i=1}^N \Delta_{n,i}^{(N)}(f_n)$$

with

$$\Delta_{n,i}^{(N)}(f_n) := \frac{1}{\sqrt{N}} \left[f_n(\mathcal{X}_n^{(N,i)}) - \mathcal{K}_{n, \eta_{n-1}^N}(f_n)(\mathcal{X}_{n-1}^{(N,i)}) \right].$$

Given $\mathcal{X}_{n-1}^{(N)}$, we let $(\mathcal{Y}_n^{(N,i)})_{1 \leq i \leq N}$ be an independent copy of $(\mathcal{X}_n^{(N,i)})_{1 \leq i \leq N}$. We readily check that

$$\Delta_{n,i}^{(N)}(f_n) := \mathbb{E} \left(\frac{1}{\sqrt{N}} \left[f_n(\mathcal{X}_n^{(N,i)}) - f_n(\mathcal{Y}_n^{(N,i)}) \right] \middle| \mathcal{X}_{n-1}^{(N)} \right)$$

This yields the formula

$$V_n^N(f_n) = \mathbb{E} \left(M_{n,N}^{(N)}(f_n) \mid \mathcal{X}_{n-1}^{(N)} \right)$$

where $M_{n,N}^{(N)}(f_n)$ is the terminal value of the martingale sequence defined below

$$i \in \{1, \dots, N\} \mapsto M_{n,i}^{(N)}(f_n) := \sum_{j=1}^i \frac{1}{\sqrt{N}} \left[f_n(\mathcal{X}_n^{(N,j)}) - f_n(\mathcal{Y}_n^{(N,j)}) \right].$$

Using Kintchine's inequality the end of the proof of (4.12) is clear. $\blacksquare$

The proof of the following lemma is rather technical and is provided in the appendix.

Lemma 4.4 *For any $0 \leq p \leq n$, any $\eta, \mu \in \mathcal{P}(S_p)$ and any $f_n \in \text{Osc}_1(S_n)$, we have the first order decomposition for the non-linear semi-group $\Phi_{p,n}$ defined in (3.7):*

$$[\Phi_{p,n}(\mu) - \Phi_{p,n}(\eta)](f_n) = 2 q_{p,n} \beta(\mathcal{P}_{p,n}) [\mu - \eta](\mathcal{U}_{p,n,\eta}(f_n)) + \mathcal{R}_{p,n}(\mu, \eta)(f_n)$$

with a remainder term

$$|\mathcal{R}_{p,n}(\mu, \eta)(f_n)| \leq 4 q_{p,n}^3 \beta(\mathcal{P}_{p,n}) \|\mu - \eta\|(\mathcal{V}_{p,n,\eta}(f)) \times \|\mu - \eta\|(\mathcal{W}_{p,n,\eta}(f_n))$$

and a collection of functions $\mathcal{U}_{p,n,\eta}(f), \mathcal{V}_{p,n,\eta}(f), \mathcal{W}_{p,n,\eta}(f)$ in $\text{Osc}_1(S_p)$, whose values only depends on the parameters (p, n, η) .

We now present a bias estimate and some $\mathbb{L}_m$ bounds of independent interest.

Theorem 4.5 *For any $n \geq 0$, $f_n \in \text{Osc}_1(S_n)$ and any $N \geq 1$,*

$$N \left| \mathbb{E}(\eta_n^N(f_n)) - \eta_n(f_n) \right| \leq \sigma_{1,n} \quad \text{with} \quad \sigma_{1,n} := 4 \sum_{p=0}^n q_{p,n}^3 \beta(\mathcal{P}_{p,n}).$$

In addition, for any $m \geq 1$ we have

$$\sqrt{N} \mathbb{E} \left(\left| [\eta_n^N - \eta_n](f_n) \right|^m \right)^{\frac{1}{m}} \leq \frac{1}{\sqrt{N}} b(2m)^2 \sigma_{1,n} + b(m) \sigma_{2,n}$$

with

$$\sigma_{2,n} := 2 \sum_{p=0}^n q_{p,n} \beta(\mathcal{P}_{p,n}).$$

Proof:

Using Lemma 4.4, we have the telescoping sum decomposition

$$\begin{aligned} W_n^N &:= \sqrt{N} [\eta_n^N - \eta_n] \\ &= \sqrt{N} \sum_{p=0}^n [\Phi_{p,n}(\eta_p^N) - \Phi_{p,n}(\Phi_p(\eta_{p-1}^N))] = \mathcal{I}_n^N + \mathcal{J}_n^N \end{aligned}$$

with the pair of random measures $(\mathcal{I}_n^N, \mathcal{J}_n^N)$ given for any $f_n \in \text{Osc}_1(S_n)$ by

$$\begin{aligned}\mathcal{I}_n^N(f_n) &:= 2 \sum_{p=0}^n q_{p,n} \beta(\mathcal{P}_{p,n}) V_p^N \left(\mathcal{U}_{p,n, \Phi_p(\eta_{p-1}^N)}(f_n) \right) \\ \mathcal{J}_n^N(f_n) &:= \sqrt{N} \sum_{p=0}^n \mathcal{R}_{p,n}(\eta_p^N, \Phi_p(\eta_{p-1}^N))(f_n).\end{aligned}$$

Now, observe that

$$\mathbb{E}(W_n^N(f_n)) = \mathbb{E}(\mathcal{J}_n^N(f_n)).$$

Using Proposition 4.3, for any $f_n \in \text{Osc}_1(S_n)$ it can be checked that

$$\mathbb{E} \left(|\mathcal{I}_n^N(f_n)|^m \right)^{\frac{1}{m}} \leq b(m) \sigma_{2,n}.$$

In much the same way, we find that

$$\sqrt{N} \mathbb{E} \left(|\mathcal{J}_n^N(f_n)|^m \right)^{\frac{1}{m}} \leq b(2m)^2 \sigma_{1,n}. \quad (4.13)$$

The end of the proof is now clear. $\blacksquare$

4.2.3 A concentration theorem

The following concentration theorem is the main result of this Section. It improves the exponential estimates presented in Section 7.4.3 in [4], by removing a multiplicative correction factor of order $\sqrt{N}$ in front of the exponential decay.

Theorem 4.6 *For any $n \geq 0$, $f_n \in \text{Osc}_1(S_n)$, $N \geq 1$, and any $0 \leq \epsilon \leq 1/2$,*

$$\mathbb{P} \left(|[\eta_n^N - \eta_n](f_n)| \geq \epsilon \right) \leq 6 \exp \left(-\frac{N \epsilon^2}{8 \sigma_{1,n}} \right) \quad (4.14)$$

where the constant $\sigma_{1,n}$ is defined in Theorem 4.5.

In addition, suppose $(\mathcal{M})_m$ is satisfied for some $m \geq 1$ and condition (4.7) holds true for some $\delta > 0$ and some finite constants $(\underline{r}, \bar{r})$. In this situation, for any value of the time parameter n , for any $f_n \in \text{Osc}_1(S_n)$, $N \geq 1$ and for any $\rho \in (0, 1)$ the probability that

$$|[\eta_n^N - \eta_n](f_n)| \leq \frac{4\bar{r}}{\delta^2} \sqrt{\frac{2m\bar{r}}{N\delta} \log \left(\frac{6}{\rho} \right)}$$

is greater than $(1 - \rho)$.

Proof:

We use the same notation as in the proof of Theorem 4.5. Recall that $b(2m)^{2m} = \mathbb{E}(X^{2m})$ for every centered Gaussian random variable with $\mathbb{E}(X^2) = 1$ and

$$\forall s \in [0, 1/2) \quad \mathbb{E}(\exp \{sX^2\}) = \sum_{m \geq 0} \frac{s^m}{m!} b(2m)^{2m} = \frac{1}{\sqrt{1-2s}}.$$

Using (4.13), for any $f_n \in \text{Osc}_1(S_n)$ and $0 \leq s < 1/(2\sigma_{1,n})$ it follows that

$$\mathbb{E} \left(\exp \left\{ s \sqrt{N} \mathcal{J}_n^N(f_n) \right\} \right) \leq \sum_{m \geq 0} \frac{(s\sigma_{1,n})^m}{m!} b(2m)^{2m} = \frac{1}{\sqrt{1-2s\sigma_{1,n}}}. \quad (4.15)$$

To simplify the presentation, set

$$f_{p,n}^N := \mathcal{U}_{p,n,\Phi_p(\eta_{p-1}^N)}(f_n) \quad \text{and} \quad \alpha_{p,n} := 2 q_{p,n} \beta(\mathcal{P}_{p,n})$$

where $\mathcal{U}_{p,n,\eta}(\cdot)$ was introduced in Lemma 4.4. In this notation, it follows

$$V_p^N(f_{p,n}^N) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \left(f_{p,n}^N(\mathcal{X}_p^{(N,i)}) - \mathcal{K}_{p,\eta_{p-1}^N}(f_{p,n}^N)(\mathcal{X}_{p-1}^{(N,i)}) \right).$$

Recalling that $\mathbb{E}(e^{tX}) \leq e^{t^2 c^2/2}$, for every real-valued and centered random variable X with $|X| \leq c$, we prove that

$$\begin{aligned} & \mathbb{E} \left(\exp \left\{ t \alpha_{p,n} V_p^N(f_{p,n}^N) \right\} \mid \mathcal{X}_{p-1}^{(N)} \right) \\ &= \prod_{i=1}^N \int_{S_p} \mathcal{K}_{p,\eta_{p-1}^N}(\mathcal{X}_{p-1}^{(N,i)}, dx) e^{\frac{t \alpha_{p,n}}{\sqrt{N}} \left(f_{p,n}^N(x) - \mathcal{K}_{p,\eta_{p-1}^N}(f_{p,n}^N)(\mathcal{X}_{p-1}^{(N,i)}) \right)} \\ &\leq \exp \left(\frac{t^2 \alpha_{p,n}^2}{2} \right). \end{aligned}$$

Iterating the argument, we find that

$$\mathbb{E} \left(e^{t \mathcal{I}_n^N(f_n)} \right) = \mathbb{E} \left(\exp \left\{ t \sum_{p=0}^n \alpha_{p,n} V_p^N(f_{p,n}^N) \right\} \right) \leq \exp \left(\frac{t^2 \sigma_n^2}{2} \right) \quad (4.16)$$

with

$$\sigma_n^2 := 4 \sum_{p=0}^n q_{p,n}^2 \beta(\mathcal{P}_{p,n})^2.$$

From these Laplace upper bounds, the proof of the exponential estimates now follows standard arguments. Indeed, for any $0 \leq s < 1/(2\sigma_{1,n})$ and any $\epsilon > 0$, by (4.15) we have

$$\mathbb{P} \left(\sqrt{N} \mathcal{J}_n^N(f_n) \geq \epsilon \right) \leq \frac{1}{\sqrt{1-2s\sigma_{1,n}}} \exp \{-\epsilon s\}.$$

Replacing ϵ by ϵN

$$\mathbb{P} \left(\mathcal{J}_n^N(f_n)/\sqrt{N} \geq \epsilon \right) \leq \frac{1}{\sqrt{1-2s\sigma_{1,n}}} \exp \{-\epsilon s N\}.$$

Now, if choosing $s = 3/(8\sigma_{1,n})$, then

$$\mathbb{P} \left(\mathcal{J}_n^N(f_n)/\sqrt{N} \geq \epsilon \right) \leq 2 \exp \{-\epsilon N/(3\sigma_{1,n})\}.$$

To estimate the probability tails of $\mathcal{I}_n^N(f_n)$, we use (4.16) and the fact that $\epsilon > 0$ and $t \geq 0$

$$\mathbb{P}(\mathcal{I}_n^N(f_n) \geq \epsilon) \leq \exp \left\{ - \left(\epsilon t - \frac{t^2}{2} \sigma_n^2 \right) \right\}.$$

Now choosing $t = \epsilon/\sigma_n^2$ and replacing ϵ by $\sqrt{N}\epsilon$ then we obtain

$$\forall \epsilon > 0 \quad \mathbb{P}(\mathcal{I}_n^N(f)/\sqrt{N} \geq \epsilon) \leq \exp \left(- \frac{N\epsilon^2}{2\sigma_n^2} \right).$$

Using the decomposition

$$[\eta_n^N - \eta_n] = \mathcal{I}_n^N/\sqrt{N} + \mathcal{J}_n^N/\sqrt{N}$$

we find that for any parameter $\alpha \in [0, 1]$

$$\mathbb{P}([\eta_n^N - \eta_n](f_n) \geq \epsilon) \leq \mathbb{P}(\mathcal{I}_n^N(f_n)/\sqrt{N} \geq \alpha\epsilon) + \mathbb{P}(\mathcal{J}_n^N(f_n)/\sqrt{N} \geq (1-\alpha)\epsilon).$$

From previous calculations,

$$\mathbb{P}([\eta_n^N - \eta_n](f_n) \geq \epsilon) \leq \exp \left(- \frac{N\epsilon^2\alpha^2}{2\sigma_n^2} \right) + 2 \exp \left(- \frac{N\epsilon(1-\alpha)}{3\sigma_{1,n}} \right). \quad (4.17)$$

Now choose $\alpha = (1-\epsilon)(\geq 1/2)$, then $\alpha^2 \geq 1/4$ and

$$\begin{aligned} \mathbb{P}([\eta_n^N - \eta_n](f_n) \geq \epsilon) &\leq \exp \left(- \frac{N\epsilon^2}{8\sigma_n^2} \right) + 2 \exp \left(- \frac{N\epsilon^2}{3\sigma_{1,n}} \right) \\ &\leq 3 \exp \left(- \frac{N\epsilon^2}{8(\sigma_{1,n} \vee \sigma_n^2)} \right). \end{aligned}$$

It remains to observe that

$$q_{p,n} \geq 1 \quad \text{and} \quad \beta(\mathcal{P}_{p,n}) \leq 1 \implies \sigma_n^2 \leq \sigma_{1,n}$$

and

$$\begin{aligned} |[\eta_n^N - \eta_n](f_n)| \geq \epsilon &\iff [\eta_n^N - \eta_n](f_n) \geq \epsilon \quad \text{or} \quad [\eta_n^N - \eta_n](f_n) \leq -\epsilon \\ &\iff [\eta_n^N - \eta_n](f_n) \geq \epsilon \quad \text{or} \quad [\eta_n^N - \eta_n](-f_n) \geq \epsilon \end{aligned}$$

so that

$$\mathbb{P}(|[\eta_n^N - \eta_n](f_n)| \geq \epsilon) \leq \mathbb{P}([\eta_n^N - \eta_n](f_n) \geq \epsilon) + \mathbb{P}([\eta_n^N - \eta_n](-f_n) \geq \epsilon).$$

The end of the proof of (4.14) is now easily completed. We further assume that the mixing condition $(\mathcal{M})_m$ is satisfied for some $m \geq 1$ and condition (4.7) holds true for some $\delta > 0$ and some finite constants $(\underline{r}, \bar{r})$. In this situation, by (4.8) the following uniform concentration estimate holds

$$\sup_{n \geq 0} \mathbb{P}(|[\eta_n^N - \eta_n](f_n)| \geq \epsilon) \leq 6 \exp \left(- \frac{N\epsilon^2\delta^5}{32m\underline{r}\bar{r}^3} \right).$$

The proof of the theorem is concluded by choosing $\epsilon := \frac{1}{\sqrt{N}} \frac{4\bar{r}}{\delta^2} \sqrt{\frac{2m\underline{r}\bar{r}}{\delta} \log \frac{6}{\rho}}$.

■

Remark 4.7 *Returning to the end of the proof of Theorem 4.6, it can be observed that the exponential concentration estimates can be improved, marginally, by choosing in (4.17) the parameter $\alpha = \alpha_n(\epsilon) \in [0, 1]$ such that $a_n(\epsilon) \alpha^2 = b_n (1 - \alpha)$, with $a_n(\epsilon) := \frac{\epsilon}{2\sigma_n^2}$ and $b_n = \frac{1}{3\sigma_{1,n}}$. Elementary manipulations yield*

$$\begin{aligned} \alpha_n(\epsilon) &= \frac{b_n}{2a_n(\epsilon)} \left(\sqrt{1 + \frac{4a_n(\epsilon)}{b_n}} - 1 \right) \\ &= \frac{\sigma_n^2}{3\sigma_{1,n}} \frac{1}{\epsilon} \left(\sqrt{1 + \frac{6\sigma_{1,n}}{\sigma_n^2} \epsilon} - 1 \right) (\xrightarrow{\epsilon \downarrow 0} 1) \end{aligned}$$

and therefore

$$\forall \epsilon \geq 0 \quad \mathbb{P}([\eta_n^N - \eta_n](f) \geq \epsilon) \leq 3 \exp \left(-N \frac{\epsilon^2}{2\sigma_n^2} \alpha_n^2(\epsilon) \right).$$

4.3 Approximating the criteria

By construction, the particle occupation measures $\mathbb{P}_{\hat{\eta}_n, (t_n, s)}^N$ approximate the measures $\mathbb{P}_{\hat{\eta}_n, (t_n, s)}$ introduced in Definition 3.2; that is, in some sense

$$\mathbb{P}_{\hat{\eta}_n^N, (t_n, s)}^N \simeq_{N \uparrow \infty} \mathbb{P}_{\hat{\eta}_n, (t_n, s)}.$$

Conversely, observe that $\mathbb{P}_{\hat{\eta}_n^N, (t_n, s)}$, and respectively $\mathbb{P}_{\hat{\eta}_n, (t_n, s)}$ are the marginals of the measures η_{n+1}^N , and respectively η_{n+1} w.r.t the $(s - t_n) + 1$ first coordinates. In other words, the measures $\mathbb{P}_{\hat{\eta}_n^N, (t_n, s)}$, and respectively $\mathbb{P}_{\hat{\eta}_n, (t_n, s)}$ are the projections of the measures η_{n+1}^N , and respectively η_{n+1} on the state space

$$E_s := E_{t_n} \times (E'_{t_n+1} \times \dots \times E'_s).$$

For instance, the following proposition is more or less a direct consequence of Theorem 4.6.

Proposition 4.8 *For any $N \geq 1$, $n \geq 0$, $t_n \leq s \leq t_{n+1}$ and any $\epsilon > 0$, the concentration inequality:*

$$\mathbb{P} \left(\left| \mathcal{H}_{t_n, s}^{(n)} \left(\mathbb{P}_{\hat{\eta}_n^N, (t_n, s)}^N \right) - \mathcal{H}_{t_n, s}^{(n)} \left(\mathbb{P}_{\hat{\eta}_n, (t_n, s)} \right) \right| \geq \epsilon \right) \leq (1 + \epsilon \sqrt{N/2}) \exp \left(-\frac{N\epsilon^2}{c(n)} \right)$$

holds for some finite constant $c(n) < \infty$ whose values only depend on the time parameter. In addition, when the measures $H_{t_n, s}^{(n)}$ have a finite support, the concentration inequality:

$$\mathbb{P} \left(\left| \mathcal{H}_{t_n, s}^{(n)} \left(\mathbb{P}_{\hat{\eta}_n^N, (t_n, s)}^N \right) - \mathcal{H}_{t_n, s}^{(n)} \left(\mathbb{P}_{\hat{\eta}_n, (t_n, s)} \right) \right| \geq \epsilon \right) \leq c_1(n) \exp \left(-\frac{N\epsilon^2}{c_2(n)} \right)$$

also holds, with a pair of finite constants $c_1(n), c_2(n) < \infty$ whose values only depend on the time parameter.

Proof:

By theorem 7.4.4 in [4], for any $N \geq 1$, $p \geq 1$, $n \geq 0$ and any test function $f_n \in \text{Osc}_1(E_{t_n})$, we have

$$\sup_{N \geq 1} \sqrt{N} \mathbb{E} \left(\left| \eta_n^N(f_n) - \eta_n(f_n) \right|^p \right)^{\frac{1}{p}} \leq b(p) c(n)$$

with some finite constant $c(n) < \infty$ and with the collection of constants $b(p)$ defined in (4.11). These estimates clearly imply that for any $t_n \leq s \leq t_{n+1}$, and any test function $h_n \in \text{Osc}_1(E_s)$

$$\sup_{N \geq 1} \sqrt{N} \mathbb{E} \left(\left| \mathbb{P}_{\hat{\eta}_n^N, (t_n, s)}^N(h_n) - \mathbb{P}_{\hat{\eta}_n, (t_n, s)}(h_n) \right|^p \right)^{\frac{1}{p}} \leq b(p) c(n).$$

Under our Lipschitz regularity condition (3.10) on the criteria type functionals $\mathcal{H}_{t_n, s}^{(n)}$ and using the generalized integral Minkowski inequality, it can be concluded that

$$\sup_{N \geq 1} \sqrt{N} \mathbb{E} \left(\left| \mathcal{H}_{t_n, s}^{(n)} \left(\mathbb{P}_{\hat{\eta}_n^N, (t_n, s)}^N \right) - \mathcal{H}_{t_n, s}^{(n)} \left(\mathbb{P}_{\hat{\eta}_n, (t_n, s)} \right) \right|^p \right)^{\frac{1}{p}} \leq b(p) c(n) \delta(H_{t_n, s}^{(n)}).$$

The proof of the exponential estimate follows exactly the same lines of arguments as the ones used in the proof of corollary 7.4.3 in [4]; thus it is omitted. The last assertion is a direct consequence of Theorem 4.6. This ends the proof of the proposition $\blacksquare$

4.4 An online adaptive SMC algorithm

The above proposition shows that the functional criteria $\mathcal{H}_{t_n, s}^{(n)} \left(\mathbb{P}_{\hat{\eta}_n, (t_n, s)} \right)$ can be approximated by $\mathcal{H}_{t_n, s}^{(n)} \left(\mathbb{P}_{\hat{\eta}_n^N, (t_n, s)}^N \right)$, up to an exponentially small probability. Therefore, as we cannot compute the deterministic resampling times (t_n) , it is necessary to approximate the reference particle model:

Definition 4.9 *We use the letters $\mathcal{Y}$ and Y to define the particle model associated to the following approximation of the reference particle model. More precisely, the particle systems $\mathcal{Y}^{(N)} = (\mathcal{Y}^{(N, i)})$, $\hat{\mathcal{Y}}^{(N)} = (\hat{\mathcal{Y}}^{(N, i)})$, $Y_{s, t}^{(N, i)}$ and $\hat{Y}_{s, t}^{(N, i)}$ are defined as $\mathcal{X}^{(N)} = (\mathcal{X}^{(N, i)})$, $\hat{\mathcal{X}}^{(N)} = (\hat{\mathcal{X}}^{(N, i)})$, and $X_{s, t}^{(N, i)}$ and $\hat{X}_{s, t}^{(N, i)}$ by replacing in the inductive construction of the deterministic sequence $(t_n)_{n \geq 0}$ the measures $\mathbb{P}_{\hat{\eta}_n, (t_n, s)}$ by their current N -particle approximation measures*

$$\bar{\mathbb{P}}_{\hat{\eta}_n^N, (t_n^N, s)}^N(\cdot) := \frac{1}{N} \sum_{i=1}^N \delta_{\left(\hat{\mathcal{Y}}_n^{(N, i)}, \left(Y_{t_n^N+1}^{(N, i)}, Y_{t_n^N+2}^{(N, i)}, \dots, Y_s^{(N, i)} \right) \right)}(\cdot).$$

Here $\hat{\eta}_n^N(\cdot) = \frac{1}{N} \sum_{i=1}^N \delta_{\hat{\mathcal{Y}}_n^{(N, i)}}(\cdot)$ denotes the updated occupation measure of the particle model $\hat{\mathcal{Y}}_n^{(N)}$. We also assume that both models are constructed in such a way that they coincide on every time interval $0 \leq n \leq m$, as soon as the random times $t_n^N = t_n$, for every $0 \leq n \leq m$.

It is emphasized that the measures $\bar{\mathbb{P}}_{\hat{\eta}_n^N, (t_n^N, s)}^N$ generally differ from the reference empirical measures $\mathbb{P}_{\hat{\eta}_n^N, (t_n, s)}^N$ introduced in (4.5). Indeed, the reference measures $\mathbb{P}_{\hat{\eta}_n^N, (t_n, s)}^N$ are built using the deterministic times t_n based on the functional criteria $\mathcal{H}_{t_{n-1}, s}^{(n-1)}(\mathbb{P}_{\hat{\eta}_{n-1}, (t_{n-1}, s)})$, whilst the empirical measures $\bar{\mathbb{P}}_{\hat{\eta}_n^N, (t_n^N, s)}^N$ are inductively constructed using random times t_n^N based on the current empirical criteria $\mathcal{H}_{t_{n-1}^N, s}^{(n-1)}(\bar{\mathbb{P}}_{\hat{\eta}_{n-1}^N, (t_{n-1}^N, s)})$.

By construction, for the four functional criteria discussed in Section 3.6, we have $\mathcal{H}_{t_n^N, s}^{(n)}(\bar{\mathbb{P}}_{\hat{\eta}_n^N, (t_n^N, s)}^N) = C_{t_n^N, s}^N$ where $C_{t_n^N, s}^N$ are the empirical criteria discussed in 2.3.

5 Asymptotic analysis

5.1 A key approximation lemma

To go one step further in our discussion, it is convenient to introduce the following collection of events.

Definition 5.1 For any $\delta \in (0, 1)$, $m \geq 0$, $a_n \in \mathbb{R}$ and $N \geq 1$, we denote by $\Omega_m^N(\delta, (a_n)_{0 \leq n \leq m})$, the collection of events defined by:

$$\Omega_m^N(\delta, (a_n)_{0 \leq n \leq m})$$

$$:= \{\forall 0 \leq n \leq m, \forall t_n \leq s \leq t_{n+1}$$

$$\left| \mathcal{H}_{t_n, s}^{(n)}(\mathbb{P}_{\hat{\eta}_n^N, (t_n, s)}^N) - \mathcal{H}_{t_n, s}^{(n)}(\mathbb{P}_{\hat{\eta}_n, (t_n, s)}) \right| \leq \delta \left| \mathcal{H}_{t_n, s}^{(n)}(\mathbb{P}_{\hat{\eta}_n, (t_n, s)}) - a_n \right\}.$$

Lemma 5.2 On the event $\Omega_m^N(\delta, (a_n)_{0 \leq n \leq m})$, for any $n \leq m$ and for any $t_n \leq s \leq t_{n+1}$, we have

$$\mathcal{H}_{t_n, s}^{(n)}(\mathbb{P}_{\hat{\eta}_n, (t_n, s)}) > a_n \implies \mathcal{H}_{t_n, s}^{(n)}(\mathbb{P}_{\hat{\eta}_n^N, (t_n, s)}^N) > a_n$$

$$\mathcal{H}_{t_n, s}^{(n)}(\mathbb{P}_{\hat{\eta}_n, (t_n, s)}) < a_n \implies \mathcal{H}_{t_n, s}^{(n)}(\mathbb{P}_{\hat{\eta}_n^N, (t_n, s)}^N) < a_n.$$

Proof:

Observe that

$$\left| \mathcal{H}_{t_n, s}^{(n)}(\mathbb{P}_{\hat{\eta}_n^N, (t_n, s)}^N) - \mathcal{H}_{t_n, s}^{(n)}(\mathbb{P}_{\hat{\eta}_n, (t_n, s)}) \right| \leq \delta \left| \mathcal{H}_{t_n, s}^{(n)}(\mathbb{P}_{\hat{\eta}_n, (t_n, s)}) - a_n \right|$$

is equivalent to saying that

$$\mathcal{H}_{t_n, s}^{(n)}(\mathbb{P}_{\hat{\eta}_n, (t_n, s)}) - \delta \left| \mathcal{H}_{t_n, s}^{(n)}(\mathbb{P}_{\hat{\eta}_n, (t_n, s)}) - a_n \right| \leq \mathcal{H}_{t_n, s}^{(n)}(\mathbb{P}_{\hat{\eta}_n^N, (t_n, s)}^N)$$

and

$$\mathcal{H}_{t_n, s}^{(n)}(\mathbb{P}_{\hat{\eta}_n^N, (t_n, s)}^N) \leq \mathcal{H}_{t_n, s}^{(n)}(\mathbb{P}_{\hat{\eta}_n, (t_n, s)}) + \delta \left| \mathcal{H}_{t_n, s}^{(n)}(\mathbb{P}_{\hat{\eta}_n, (t_n, s)}) - a_n \right|.$$

Thus, if $\mathcal{H}_{t_n,s}^{(n)}(\mathbb{P}_{\hat{\eta}_n,(t_n,s)}) < a_n$ then

$$\begin{aligned} \mathcal{H}_{t_n,s}^{(n)}(\mathbb{P}_{\hat{\eta}_n^N,(t_n,s)}) &\leq \mathcal{H}_{t_n,s}^{(n)}(\mathbb{P}_{\hat{\eta}_n,(t_n,s)}) + \delta \left(a_n - \mathcal{H}_{t_n,s}^{(n)}(\mathbb{P}_{\hat{\eta}_n,(t_n,s)}) \right) \\ &\leq \delta a_n + (1 - \delta) \mathcal{H}_{t_n,s}^{(n)}(\mathbb{P}_{\hat{\eta}_n,(t_n,s)}) < a_n. \end{aligned}$$

Similarly, if we have $\mathcal{H}_{t_n,s}^{(n)}(\mathbb{P}_{\hat{\eta}_n,(t_n,s)}) > a_n$ then we also have

$$\begin{aligned} \mathcal{H}_{t_n,s}^{(n)}(\mathbb{P}_{\hat{\eta}_n^N,(t_n,s)}) &\geq \mathcal{H}_{t_n,s}^{(n)}(\mathbb{P}_{\hat{\eta}_n,(t_n,s)}) - \delta \left(\mathcal{H}_{t_n,s}^{(n)}(\mathbb{P}_{\hat{\eta}_n,(t_n,s)}) - a_n \right) \\ &\geq \delta a_n + (1 - \delta) \mathcal{H}_{t_n,s}^{(n)}(\mathbb{P}_{\hat{\eta}_n,(t_n,s)}) > a_n. \end{aligned}$$

This ends the proof of the lemma. $\blacksquare$

Proposition 5.3 *Assume that the threshold parameters a_n are chosen so that $\mathcal{H}_{t_n,s}^{(n)}(\mathbb{P}_{\hat{\eta}_n,(t_n,s)}) \neq a_n$, for any $n \geq 0$. In this situation, for any $\delta \in (0, 1)$, $m \geq 0$ and $N \geq 1$, we have*

$$\cap_{0 \leq n \leq m} \{ t_n^N = t_n \} \supset \Omega_m^N(\delta, (a_n)_{0 \leq n \leq m})$$

Proof:

This result is proved by induction on the parameter $m \geq 0$. Under our assumptions, for $m = 0$ we have $t_0^N = t_0 = 0$. Thus, by our coupling construction the pair of particle models coincide up to the time $(t_1^N \wedge t_1)$. Therefore,

$$\forall s < (t_1^N \wedge t_1) \quad \mathbb{P}_{\hat{\eta}_0^N,(t_0,s)}^N = \bar{\mathbb{P}}_{\hat{\eta}_0^N,(t_0^N,s)}^N.$$

By Lemma 5.2, on the event $\Omega_m^N(\delta, (a_n)_{0 \leq n \leq m})$ we have $t_1^N = t_1$. This proves the inclusion for $m = 0$ and $m = 1$. Suppose the result is true at rank m . Thus, on the event $\Omega_m^N(\delta, (a_n)_{0 \leq n \leq m})$ it is the case that $t_n^N = t_n$, for any $0 \leq n \leq m$. By our coupling construction, the pair of particle models coincides up to $(t_{m+1}^N \wedge t_{m+1})$; that is,

$$t_m^N = t_m \quad \text{and} \quad \forall s < (t_{m+1}^N \wedge t_{m+1}) \quad \mathbb{P}_{\hat{\eta}_m^N,(t_m,s)}^N = \bar{\mathbb{P}}_{\hat{\eta}_m^N,(t_m^N,s)}^N.$$

Once again, by Lemma 5.2, on the event $\Omega_{m+1}^N(\delta, (a_n)_{0 \leq n \leq m+1})$ it also follows that $t_{m+1}^N = t_{m+1}$. This ends the proof of the proposition. $\blacksquare$

5.2 Randomized criteria

The situation where the threshold parameters coincide with the adaptive criteria values $\mathcal{H}_{t_n,s}^{(n)}(\mathbb{P}_{\hat{\eta}_n,(t_n,s)}) = a_n$ cannot be handled using our analysis. Indeed, in this rather degenerate situation it is impossible, at least with our approach, to compare the random times $(t_n^N)_{n \geq 0}$ with the reference deterministic ones $(t_n)_{n \geq 0}$. The analysis of this situation is clearly more involved since it requires us to control both the empirical approximating criteria and the mean field approximation model.

To avoid this technical problem, one natural strategy is to introduce randomized criteria thresholds. We further assume that the parameters $(a_n)_{n \geq 0}$ are sampled realizations of a collection of absolutely continuous random variables $(A_n)_{n \geq 0}$; in practice, this element can be incorporated into the simulation, with no real change in the performance. The main simplification of these randomized criteria comes from the fact that the parameters

$$\epsilon_m := \inf_{0 \leq n \leq m} \inf_{t_n \leq s \leq t_{n+1}} \left| \mathcal{H}_{t_n, s}^{(n)} (\mathbb{P}_{\hat{\eta}_n, (t_n, s)}) - a_n \right|$$

are strictly positive for almost every realizations $A_n = a_n$ of the threshold parameters.

Theorem 5.4 *For almost every realization of the random threshold parameters, and for any $\delta \in (0, 1)$, we have the following exponential estimates :*

$$\mathbb{P}(\exists 0 \leq n \leq m \quad t_n^N \neq t_n \mid (A_n)_{0 \leq n \leq m}) \leq c_1(m) \left(1 + \delta \epsilon_m \sqrt{\frac{N}{2}} \right) e^{-\frac{N \delta^2 \epsilon_m^2}{c_2(m)}}$$

for some constants $c_1(m), c_2(m) < \infty$ whose values only depend on the time parameter m . In addition, when the measures $H_{t_n, s}^{(n)}$ have a finite support, for any $\delta \in]0, 1/(2\epsilon_m)[$

$$\mathbb{P}(\exists 0 \leq n \leq m \quad t_n^N \neq t_n \mid (A_n)_{0 \leq n \leq m}) \leq c_1(m) e^{-N \delta^2 \epsilon_m^2 / c_2(m)}$$

holds, for a possibly different pair of finite constants $c_1(m), c_2(m) < \infty$.

Proof:

Using Proposition 4.8, we obtain the rather crude estimate

$$\begin{aligned} & \mathbb{P}(\Omega - \Omega_m^N(\delta, (A_n)_{0 \leq n \leq m}) \mid (A_n)_{0 \leq n \leq m} = (a_n)_{0 \leq n \leq m}) \\ & \leq \sum_{n=0}^m \sum_{s=t_n}^{t_{n+1}} \\ & \quad \mathbb{P}\left(\left|\mathcal{H}_{t_n, s}^{(n)}(\mathbb{P}_{\hat{\eta}_n, (t_n, s)}^N) - \mathcal{H}_{t_n, s}^{(n)}(\mathbb{P}_{\hat{\eta}_n, (t_n, s)})\right| \geq \delta \left|\mathcal{H}_{t_n, s}^{(n)}(\mathbb{P}_{\hat{\eta}_n, (t_n, s)}) - a_n\right|\right) \\ & \leq \sum_{n=0}^m \sum_{s=t_n}^{t_{n+1}} \mathbb{P}\left(\left|\mathcal{H}_{t_n, s}^{(n)}(\mathbb{P}_{\hat{\eta}_n, (t_n, s)}^N) - \mathcal{H}_{t_n, s}^{(n)}(\mathbb{P}_{\hat{\eta}_n, (t_n, s)})\right| \geq \delta \epsilon_m\right) \\ & \leq c_1(m) (1 + \delta \epsilon_m \sqrt{N/2}) \exp(-N \delta^2 \epsilon_m^2 / c_2(m)) \end{aligned}$$

for a pair of finite constants $c_1(m), c_2(m) < \infty$ whose values only depend on the time parameter. The final line is a direct consequence of Proposition 4.8 and an application of Proposition 5.3 completes the proof. $\blacksquare$

We conclude that for almost every realization $(A_n)_{0 \leq n \leq m} = (a_n)_{0 \leq n \leq m}$ the pair particle models $(\mathcal{X}_n^{(N)}, \hat{\mathcal{X}}_n^{(N)})_{0 \leq n \leq m}$ and $(\mathcal{Y}^{(N)}, \hat{\mathcal{Y}}_n^{(N)})_{0 \leq n \leq m}$ only differ on events $\Omega - \Omega_m^N(\delta, (a_n)_{0 \leq n \leq m})$ with exponentially small probabilities:

$$\begin{aligned} & \mathbb{P}(\exists 0 \leq n \leq m \quad (\mathcal{Y}^{(N)}, \hat{\mathcal{Y}}_n^{(N)}) \neq (\mathcal{X}_n^{(N)}, \hat{\mathcal{X}}_n^{(N)}) \mid (A_n)_{0 \leq n \leq m} = (a_n)_{0 \leq n \leq m}) \\ & \leq c_1(m) (1 + \delta \epsilon_m \sqrt{N/2}) \exp(-N \delta^2 \epsilon_m^2 / c_2(m)). \end{aligned}$$

6 A functional central limit theorem

6.1 A direct approach

In this Section some direct consequences of the exponential coupling estimates, presented in Section 5.2, are discussed. For almost every realization $(A_n)_{0 \leq n \leq m} = (a_n)_{0 \leq n \leq m}$ and for any test function $f_n \in \text{Osc}_1(E_{t_n})$ the following decomposition holds (writing $\bar{\eta}_n^N$ for the online adaptive approximation introduced in Definition 4.9)

$$\sqrt{N} [\bar{\eta}_n^N - \eta_n] = \sqrt{N} [\eta_n^N - \eta_n] + \sqrt{N} [\bar{\eta}_n^N - \eta_n^N] 1_{\Omega - \Omega_m^N(\delta, (a_n)_{0 \leq n \leq m})}$$

with

$$\mathbb{E} \left(\sqrt{N} [\bar{\eta}_n^N - \eta_n^N] (f_n) 1_{\Omega - \Omega_m^N(\delta, (a_n)_{0 \leq n \leq m})} \right) \leq \underbrace{\sqrt{N} \mathbb{P}(\Omega - \Omega_m^N(\delta, (a_n)_{0 \leq n \leq m}))}_{\rightarrow N \uparrow \infty 0}.$$

Thus we can conclude directly that, for almost every realization $(A_n)_{0 \leq n \leq m} = (a_n)_{0 \leq n \leq m}$, the random fields

$$\bar{W}_n^N := \sqrt{N} [\bar{\eta}_n^N - \eta_n] = \quad \text{and} \quad W_n^N := \sqrt{N} [\eta_n^N - \eta_n]$$

converge in law, as $N \uparrow \infty$, to the same Gaussian and centered random field W_n .

6.2 Functional central limit theorems

To demonstrate the impact of this functional fluctuation result and to connect this work with related studies, we provide a brief discussion on the proof of this multivariate central limit theorem.

Let us begin by recalling a rather well-known functional fluctuation theorem of the local errors associated with the mean field particle approximation introduced in (4.9). This result was initially presented in [6] in the case of elementary genetic type models. It has been extended to general mean field particle models in Section 9.3 in [4].

Theorem 6.1 *For any fixed time horizon $n \geq 0$, the sequence $(V_p^N)_{0 \leq p \leq n}$ converges in law, as N tends to infinity, to a sequence of n independent, Gaussian and centered random fields $(V_p)_{0 \leq p \leq n}$ with, for any $f_p, g_p \in \mathcal{B}_b(E_p)$, and $1 \leq p \leq n$,*

$$\mathbb{E}(V_p(f_p)V_p(g_p)) = \eta_{p-1}K_{p,\eta_{p-1}}([f_p - K_{p,\eta_{p-1}}(f_p)][g_p - K_{p,\eta_{p-1}}(g_p)]). \quad (6.1)$$

Using arguments fairly similar to those in the proof of Lemma 4.4, we obtain the first order decomposition formula:

$$[\Phi_n(\mu) - \Phi_n(\eta)](f) = (\mu - \eta)\mathcal{D}_{n,\eta} + \mathcal{R}_n(\mu, \eta)$$

with the signed measure $\mathcal{R}_n(\mu, \eta)$ given by

$$\mathcal{R}_n(\mu, \eta)(f) := -\frac{1}{\mu(\mathcal{G}_{n,\eta})} [\mu - \eta]^{\otimes 2}(\mathcal{G}_{n,\eta} \otimes \mathcal{D}_{n,\eta}(f)) \quad \text{with} \quad \mathcal{G}_{n,\eta} := \mathcal{G}_{n-1}/\eta(\mathcal{G}_{n-1})$$

and the bounded integral operator

$$\mathcal{D}_{n,\eta}(f)(x) := \mathcal{G}_{n,\eta}(x) \times \mathcal{M}_n(f - \Phi_n(\eta)(f))(x).$$

Definition 6.2 Denote by $D_{p,n}$ the semi-group associated to the integral operators $D_n := \mathcal{D}_{n,\eta_{n-1}}$; that is,

$$D_{p,n} := D_{p+1} \dots D_{n-1} D_n.$$

For $p = n$, we use the convention $D_{n,n} = \text{Id}$, the identity operator.

It should be noted that the semi-group $D_{p,n}$ introduced above can be explicitly described in terms of the semi-group $\mathcal{Q}_{p,n}$, with the following formula

$$D_{p,n}(f) = \frac{\mathcal{Q}_{p,n}}{\eta_p(\mathcal{Q}_{p,n}(1))} (f - \eta_n(f)).$$

The next technical lemma provides a key first order decomposition of the random fields W_n^N in terms of the local fluctuation errors. Its proof follows an elementary induction on the time parameter n , and is in the appendix. Note that R_p is a remainder term that can be understood in the proof.

Lemma 6.3 For any $N \geq 1$ and any $0 \leq p \leq n$, we have

$$W_n^N = \sum_{p=0}^n V_p^N D_{p,n} + \mathcal{R}_n^N \quad \text{with} \quad \mathcal{R}_n^N := \sqrt{N} \sum_{p=0}^{n-1} R_{p+1}(\eta_p^N, \eta_p) D_{p+1,n}. \quad (6.2)$$

Using the $\mathbb{L}_m$ -mean error estimates presented in Section 4.2.2, it is easily proved that the sequence of remainder random fields $\mathcal{R}_n^N$, introduced in (6.2), converge in law, in the sense of finite distributions, to the null random field as N tends to ∞ . Therefore the fluctuations of W_n^N follows from Theorem 6.1.

Corollary 6.4 For any fixed time horizon $n \geq 0$, the sequence of random fields $(W_n^N)_{n \geq 0}$ converges in law, as N tends to infinity, to the sequence of Gaussian and centered random fields $(W_n)_{n \geq 0}$ given below

$$\forall n \geq 0 \quad W_n = \sum_{p=0}^n V_p D_{p,n}.$$

6.3 On the fluctuations of weighted occupation measures

We end this article with some comments on the fluctuations of weighted occupation measures on path spaces. Returning to the online adaptive particle model, given a pair of time steps $(t_n^N, t_{n+1}^N) = (t_n, t_{n+1})$ the N -particle measures

$$\bar{\eta}_{n+1}^N = \frac{1}{N} \sum_{i=1}^N \delta_{\left(\hat{\mathcal{Y}}_n^{(N,i)}, \left(Y_{t_n^N+1}^{(N,i)}, Y_{t_n^N+2}^{(N,i)}, \dots, Y_{t_{n+1}^N}^{(N,i)} \right) \right)}$$

can be used to approximate the flow of updated Feynman-Kac path distributions $(\hat{\eta}_{n+1,s})_{t_n \leq s \leq t_{n+1}}$ given for any bounded test function $f_{n+1} \in \mathcal{B}_b(S_{n+1})$ by

$$s \in [t_n, t_{n+1}] \mapsto \hat{\eta}_{n+1,s}(f_{n+1}) \propto \mathbb{E} \left[f_{n+1}(X'_0, \dots, X'_{t_{n+1}}) \prod_{0 < k \leq s} G'_k(X'_k) \right].$$

Indeed, if we choose the weighted functions

$$T_{n+1}^{(1)}(f_{n+1})(x'_0, \dots, x'_{t_{n+1}}) := f_{n+1}(x'_0, \dots, x'_{t_{n+1}}) \prod_{t_n < k \leq s} G'_k(x'_k)$$

then in some sense

$$\widehat{\eta}_{n+1,s}^N(f_{n+1}) := \frac{\bar{\eta}_{n+1}^N(T_{n+1}^{(1)}(f_{n+1}))}{\bar{\eta}_{n+1}^N(T_{n+1}^{(1)}(1))} \simeq_{N \uparrow \infty} \widehat{\eta}_{n+1,s}(f_{n+1}) := \frac{\eta_{n+1}(T_{n+1}^{(1)}(f_{n+1}))}{\eta_{n+1}(T_{n+1}^{(1)}(1))}$$

where η_{n+1} is the flow of Feynman-Kac measures on path spaces introduced in Section 3.3.

Since the adaptive interaction time is taken such that $t_{n+1}^N = t_{n+1}$, it holds that

$$\frac{1}{N} \sum_{i=1}^N \delta_{\widehat{\mathcal{Y}}_{n+1}^{(N,i)}} \simeq_{N \uparrow \infty} \widehat{\eta}_{n+1,t_{n+1}} = \widehat{\eta}_{n+1}.$$

In other words, the marginal type functions are chosen

$$T_{n+1}^{(0)}(f_{n+1})(x'_0, \dots, x'_{t_{n+2}}) := f_{n+1}(x'_0, \dots, x'_{t_{n+1}})$$

so

$$\eta_{n+2}(T_{n+1}^{(0)}(f_{n+1})) = \widehat{\eta}_{n+1}(f_{n+1}) \propto \mathbb{E} \left[f_{n+1}(X'_0, \dots, X'_{t_{n+1}}) \prod_{0 < k \leq t_{n+1}} G'_k(X'_k) \right]$$

and therefore

$$\bar{\eta}_{n+2}^N(T_{n+1}^{(0)}(f_{n+1})) = \frac{1}{N} \sum_{i=1}^N f_{n+1}(\widehat{\mathcal{Y}}_{n+1}^{(N,i)}) \simeq_{N \uparrow \infty} \eta_{n+2}(T_{n+1}^{(0)}(f_{n+1})).$$

From the previous discussion, for almost every realization $(A_n)_{0 \leq n \leq m} = (a_n)_{0 \leq n \leq m}$ a central limit theorem is easily derived for the collection of random fields given below

$$\widehat{W}_{n+1}^{N,(0)}(f_{n+1}) := \sqrt{N} \left[\bar{\eta}_{n+2}^N(T_{n+1}^{(0)}(f_{n+1})) - \eta_{n+2}(T_{n+1}^{(0)}(f_{n+1})) \right]$$

and

$$\widehat{W}_{n+1,s}^{N,(1)}(f_{n+1}) := \sqrt{N} \left[\widehat{\eta}_{n+1,s}^N(f_{n+1}) - \widehat{\eta}_{n+1,s}(f_{n+1}) \right]$$

as well as for the mixture of random field sequences

$$\widehat{W}_{n+1,s}^N = 1_{t_n^N \leq s < t_{n+1}^N} \widehat{W}_{n+1,s}^{N,(1)} + 1_{s=t_{n+1}^N} \widehat{W}_{n+1}^{N,(0)}. \quad (6.3)$$

The fluctuation analysis of these random fields strongly relies on the functional central limit theorem stated in corollary 6.4. In particular, the fluctuations of the random fields (6.3) depend on the fluctuations of a pair of random fields. From elementary stochastic calculus, the corresponding fluctuations cannot be deduced directly from the separate fluctuations of each components.

6.4 Related work

To our knowledge, the article [8] is the only reference providing convergence results for a specific adaptive SMC scheme in a state-space models context. The authors establish a central limit theorem using an inductive proof w.r.t deterministic time periods; this does not account for the randomness of the latter quantities. They implicitly avoid the degenerate situation where the threshold parameter coincides with the limiting functional criterion. The fluctuations of the sum of the weighted and the resampled empirical approximations is deduced from the fluctuations of each separate components without the use of any multivariate type central limit theorem.

Acknowledgement

The third author was affiliated to the Institute of Statistical Mathematics (Tokyo, Japan), whilst this project was initiated.

Appendix

Proof of Lemma 4.2

We prove the lemma by induction on the parameter n . The result is clearly satisfied for $n = 0$. Suppose the estimate (4.10) is true at rank $(n - 1)$. To prove the result at rank n , use the binomial decomposition

$$(M_{n-1} + \Delta_n)^{2m} = \sum_{p=0}^{2m} \binom{2m}{p} M_{n-1}^{2m-p} (\Delta_n)^p.$$

Using the symmetry condition, all the odd moments of Δ_n are null. Consequently,

$$\mathbb{E} \left((M_{n-1} + \Delta_n)^{2m} \right) = \sum_{p=0}^m \binom{2m}{2p} \mathbb{E} \left(M_{n-1}^{2(m-p)} \right) \mathbb{E} \left(\Delta_n^{2p} \right).$$

Then, via the induction hypothesis, the above expression is upper-bounded by the quantity

$$\sum_{p=0}^m \binom{2m}{2p} 2^{-(m-p)} (2(m-p))_{(m-p)} \mathbb{E} \left([M]_{n-1}^{m-p} \right) \mathbb{E} \left(\Delta_n^{2p} \right).$$

To take the final step, use the fact that

$$\binom{2m}{2p} 2^{-(m-p)} (2(m-p))_{(m-p)} = \frac{2^{-m} (2m)_m}{2^{-p} (2p)_p} \binom{m}{p} \quad \text{and} \quad (2p)_p \geq 2^p$$

to conclude that

$$\begin{aligned} \mathbb{E} \left((M_{n-1} + \Delta_n)^{2m} \right) &\leq 2^{-m} (2m)_m \sum_{p=0}^m \binom{m}{p} \mathbb{E} \left([M]_{n-1}^{m-p} \right) \mathbb{E} \left(\Delta_n^{2p} \right) \\ &= 2^{-m} (2m)_m \mathbb{E} \left([M]_n^m \right). \end{aligned}$$

For odd integers, apply the Cauchy-Schwarz inequality twice to deduce that

$$\begin{aligned} \mathbb{E}(|M_n|^{2m+1})^2 &\leq \mathbb{E}(M_n^{2m}) \mathbb{E}(M_n^{2(m+1)}) \\ &\leq 2^{-(2m+1)} (2m)_m (2(m+1))_{(m+1)} \mathbb{E}([M]_n^{m+1})^{\frac{2m+1}{m+1}}. \end{aligned}$$

Since we also have that

$$(2(m+1))_{(m+1)} = \frac{(2(m+1))!}{(m+1)!} = 2 \frac{(2m+1)!}{m!} = 2 (2m+1)_{(m+1)}$$

and

$$(2m)_m = \frac{1}{2m+1} \frac{(2m+1)!}{m!} = \frac{1}{2m+1} (2m+1)_{(m+1)}$$

it can be concluded that

$$\mathbb{E}(|M_n|^{2m+1}) \leq 2^{-(m+1/2)} \frac{(2m+1)_{(m+1)}}{\sqrt{m+1/2}} \mathbb{E}([M]_n^{m+1})^{1-\frac{1}{2(m+1)}}.$$

This ends the proof of the lemma. ■

Proof of Lemma 4.4

Using (3.7), for any $f \in \mathcal{B}_b(S_{n+1})$ we find that

$$[\Phi_{p,n}(\mu) - \Phi_{p,n}(\eta)](f) = \frac{1}{\mu(\mathcal{G}_{p,n,\eta})} (\mu - \eta) \mathcal{D}_{p,n,\eta}(f)$$

with the bounded integral operator

$$\mathcal{D}_{p,n,\eta}(f)(x) := \mathcal{G}_{p,n,\eta}(x) \times \mathcal{P}_{p,n}(f - \Phi_{p,n}(\eta)(f))(x)$$

where $\mathcal{G}_{p,n,\eta}$ and $\mathcal{P}_{p,n}$ stand for the potential function and the Markov operator is given by

$$\mathcal{G}_{p,n,\eta} := \mathcal{Q}_{p,n}(1)/\eta(\mathcal{Q}_{p,n}(1)) \quad \text{and} \quad \mathcal{P}_{p,n}(f) = \mathcal{Q}_{p,n}(f)/\mathcal{Q}_{p,n}(1).$$

Now, since $\eta(\mathcal{G}_{p,n,\eta}) = 1$, it is now immediate to check that

$$[\Phi_{p,n}(\mu) - \Phi_{p,n}(\eta)] = (\mu - \eta) \mathcal{D}_{p,n,\eta} + \mathcal{R}_{p,n}(\mu, \eta)$$

with the measure $\mathcal{R}_{p,n}(\mu, \eta)$ given by

$$\mathcal{R}_{p,n}(\mu, \eta)(f) := -\frac{1}{\mu(\mathcal{G}_{p,n,\eta})} [\mu - \eta]^{\otimes 2}(\mathcal{G}_{p,n,\eta} \otimes \mathcal{D}_{p,n,\eta}(f)).$$

Using the fact that

$$\mathcal{D}_{p,n,\eta}(f)(x) = \mathcal{G}_{p,n,\eta}(x) \int [\mathcal{P}_{p,n}(f)(x) - \mathcal{P}_{p,n}(f)(y)] \mathcal{G}_{p,n,\eta}(y) \eta(dy)$$

we find

$$\forall f \in \text{Osc}_1(E_n) \quad \|\mathcal{D}_{p,n,\eta}(f)\| \leq q_{p,n} \beta(\mathcal{P}_{p,n}).$$

Finally, for any $f \in \text{Osc}_1(S_n)$ observe that

$$|\mathcal{R}_{p,n}(\mu, \eta)(f)| \leq (4q_{p,n}^3 \beta(\mathcal{P}_{p,n})) \quad |[\mu - \eta]^{\otimes 2} (\bar{\mathcal{G}}_{p,n,\eta} \otimes \bar{\mathcal{D}}_{p,n,\eta}(f))|$$

with the pair of functions

$$\bar{\mathcal{G}}_{p,n,\eta} := \frac{\mathcal{G}_{p,n,\eta}}{2q_{p,n}} \quad \text{and} \quad \bar{\mathcal{D}}_{p,n,\eta}(f) := \frac{D_{p,n,\eta}(f)}{2q_{p,n}\beta(\mathcal{P}_{p,n})} \in \text{Osc}_1(S_p).$$

This ends the proof of the lemma. ■

Proof of Lemma 6.3

The lemma is proved by induction on the time parameter n . For $n = 0$, we readily find that $W_n^N = V_0^N = \sqrt{N} [\eta_0^N - \Phi_0(\eta_{-1}^N)]$, with the convention $\Phi_0(\eta_{-1}^N) = \eta_0$. Assuming that the decomposition is satisfied at rank n it may be proved that

$$\begin{aligned} W_{n+1}^N &= V_{n+1}^N + \sqrt{N} [\Phi_{n+1}(\eta_n^N) - \Phi_{n+1}(\eta_n)] \\ &= V_{n+1}^N + W_n^N D_{n+1} + \sqrt{N} R_{n+1}(\eta_n^N, \eta_n) \\ &= V_{n+1}^N + \sum_{p=0}^n V_p^N D_{p,n+1} \\ &\quad + \sqrt{N} \sum_{p=0}^{n-1} R_{p+1}(\eta_p^N, \eta_p) D_{p+1,n+1} + \sqrt{N} R_{n+1}(\eta_n^N, \eta_n). \end{aligned}$$

Using the convention $D_{n+1,n+1} = I$ it follows that

$$\begin{aligned} V_{n+1}^N + \sum_{p=0}^n V_p^N D_{p,n+1} &= \sum_{p=0}^{n+1} V_p^N D_{p,n+1} \\ \sum_{p=0}^{n-1} R_{p+1}(\eta_p^N, \eta_p) D_{p+1,n+1} + R_{n+1}(\eta_n^N, \eta_n) &= \sum_{p=0}^n R_{p+1}(\eta_p^N, \eta_p) D_{p+1,n+1} \end{aligned}$$

from which we conclude that (6.2) is satisfied at rank $(n + 1)$. This ends the proof of the lemma. ■

References

- [1] F. CEROU & A. GUYADER (2007). Adaptive multilevel splitting for rare event analysis. *Stoch. Anal. Appl.*, 25, 417-433.
- [2] G.E. CROOKS (1998) Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems. *J. Stat. Phys.*, 90, 1481-1487.
- [3] P. DEL MORAL (1998). Measure valued processes and interacting particle systems. Application to non linear filtering problems. *Ann. Appl. Probab.*, 8, 438-495.

-
- [4] P. DEL MORAL (2004). *Feynman-Kac formulae. Genealogical and interacting particle systems*, Springer-Verlag: New York.
 - [5] P. DEL MORAL, & J. JACOD (2000). Interacting particle filtering with discrete-time observations: asymptotic behaviour in the Gaussian case, in *Stochastics in finite and infinite dimensions Trends in mathematics*, Birkhäuser, pp.101-123.
 - [6] P. DEL MORAL, & L. MICLO (2000). Branching and Interacting Particle Systems Approximations of Feynman-Kac Formulae. *Séminaire de Probabilités XXXIV, Lecture Notes in Mathematics, Springer-Verlag Berlin*, Vol. 1729, 1-145.
 - [7] P. DEL MORAL, A. DOUCET, & A. JASRA (2006). Sequential Monte Carlo samplers. *J. R. Statist. Soc. B*, 68, 411–436.
 - [8] R. DOUC, & É. MOULINES (2008). Limit theorems for weighted samples with applications to sequential Monte Carlo. *Ann. Statist.*, to appear.
 - [9] A. DOUCET, N. DE FREITAS, & N. GORDON (2001). *Sequential Monte Carlo methods in practice*, Springer Verlag New York, Series: Statistics for Engineering and Information Science.
 - [10] C. JARZYNSKI (1997). Equilibrium free energy differences from nonequilibrium measurements: a master equation approach, *Phys. Rev. E* 56(5), pp. 5018-5035.
 - [11] A. JASRA, D.A. STEPHENS, A. DOUCET & T. TSAGARIS (2008). Inference for Lévy driven stochastic volatility models via adaptive sequential Monte Carlo, Technical Report, McGill University.
 - [12] J.S. LIU (2001) Monte Carlo methods for scientific computing, Springer Verlag New York.
 - [13] J.S. LIU & R. CHEN (1995). Blind deconvolution via sequential imputation. *J. Amer. Statist. Assoc.*, 90, 567-576.
 - [14] M. ROUSSET & G. STOLTZ (2006). Equilibrium sampling from non-equilibrium dynamics. *J. Stat. Phys.*, 123, 1251-1272.



Centre de recherche INRIA Bordeaux – Sud Ouest
Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex (France)

Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
<http://www.inria.fr>
ISSN 0249-6399