Bernoulli **18**(1), 2012, 252–278
DOI: 10.3150/10-BEJ335

On adaptive resampling strategies for sequential Monte Carlo methods

PIERRE DEL MORAL[1], ARNAUD DOUCET[2] and AJAY JASRA[3]

[1] *Centre INRIA Bordeaux et Sud-Ouest & Institut de Mathématiques de Bordeaux, Université de Bordeaux I, 33405, France. E-mail: pierre.del-moral@inria.fr*
[2] *Department of Statistics, University of British Columbia, Vancouver BC, Canada V6T 1Z2. E-mail: a.doucet@stat.ubc.ca*
[3] *Department of Statistics and Applied Probability, National University of Singapore, Singapore 117546, Singapore. E-mail: staja@nus.edu.sg*

Sequential Monte Carlo (SMC) methods are a class of techniques to sample approximately from any sequence of probability distributions using a combination of importance sampling and resampling steps. This paper is concerned with the convergence analysis of a class of SMC methods where the times at which resampling occurs are computed online 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. By combining semigroup techniques with an original coupling argument, we obtain functional central limit theorems and uniform exponential concentration estimates for these algorithms.

Keywords: random resampling; sequential Monte Carlo methods

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 engineering, statistics and physics; see [1, 4, 7, 8] for many applications. Sequential Monte Carlo 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 2.2 for more details. These resampling steps are crucial and, without them, it is impossible to obtain time uniform convergence results for SMC estimates. However, resampling too often has a negative effect as it decreases the number of distinct particles. Hence, a resampling step should only be applied when necessary. Consequently, in most practical implementations of SMC, the times at which resampling occurs are selected by monitoring a criterion that assesses the quality of the current particle approximation. Whenever this criterion is above or below a given threshold, a resampling step is triggered. This approach was originally proposed in [9] and has been widely adopted ever since [1], Section 7.3.2.

For this class of adaptive SMC methods, the resampling times are computed online 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 [6] for an exception discussed later. The objective of this paper is to provide convergence results for this type of adaptive SMC algorithm. This is achieved using a coupling argument. Under some assumptions, the random resampling times converge almost surely as the number of particles goes to infinity toward some deterministic (but not explicitly known) resampling times. We show here that the difference, in probability, between the reference SMC algorithm based on these deterministic but unknown resampling times and the adaptive SMC algorithm is exponentially small in the number of particles. This allows us to straightforwardly transfer the convergence results of the reference SMC algorithm to the adaptive SMC algorithm. In particular, we establish functional central limit theorems and new exponential concentration estimates that improve over those presented in [4], Section 7.4.3. Note that some exponential concentration estimates have also been established in [1], Theorem 9.4.12, using different techniques and a weaker assumption. The constants appearing in [1], Theorem 9.4.12, are not explicit so the comparison between these two results is difficult. In a specific example, we found our bound to be significantly tighter but have not established it in a general case.

The rest of the paper is organized as follows. In Section 2, we present the class of adaptive SMC algorithms studied here and our main coupling result. A precise description of the sequence of distributions approximated by the reference SMC algorithm is given in Section 3 and a theoretical analysis of the reference SMC algorithm is presented in Section 4. In particular, we propose an original concentration analysis to obtain exponential estimates for SMC approximations. 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. Finally, we analyze the fluctuations of adaptive SMC algorithms in Section 6.

2. Adaptive SMC algorithms and main results

2.1. Notation and conventions

Let $\mathcal{M}(E)$, $\mathcal{P}(E)$ and $\mathcal{B}_b(E)$ denote, respectively, the set of bounded and signed measures, the subset of all probability measures on some measurable space $(E, \mathscr{E})$ and the Banach space of all bounded and measurable functions f on E when equipped with norm $\|f\| = \sup_{x\in E} |f(x)|$. $\mathrm{Osc}_1(E)$ is the set of $\mathscr{E}$-measurable functions f with oscillations $\mathrm{osc}(f) = \sup_{(x,y)\in E^2} \{|f(x) - f(y)|\} \le 1$. $\mu(f) = \int \mu(\mathrm{d}x) f(x)$ is the integral of a function $f \in \mathcal{B}_b(E)$, w.r.t. a measure $\mu \in \mathcal{M}(E)$. $\mu(A) = \mu(1_A)$ with $A \in \mathscr{E}$ and 1_A the indicator of A. δ_a is the Dirac measure. A bounded integral operator M from a measurable space $(E, \mathscr{E})$ into another $(F, \mathscr{F})$ is an operator $f \mapsto M(f)$ from $\mathcal{B}_b(F)$ into $\mathcal{B}_b(E)$ such that the functions $M(f)(x) = \int_F M(x, \mathrm{d}y) f(y)$ are measurable and bounded for any $f \in \mathcal{B}_b(F)$. A bounded integral operator M from $(E, \mathscr{E})$ into $(F, \mathscr{F})$ also generates a dual operator $\mu \mapsto \mu M$ from $\mathcal{M}(E)$ into $\mathcal{M}(F)$ defined by $(\mu M)(f) := \mu(M(f))$. If constants are written with an argument, then they depend only on this given argument. The tensor product of functions is written $\otimes$. For any generic sequence $\{z_n\}_{n \geq 0}$, we denote $z_{i:j} = (z_i, z_{i+1}, \ldots, z_j)$ for $i \leq j$.

2.2. Adaptive sequential Monte Carlo methods

SMC methods are a popular class of methods for sampling random variables distributed approximately according to the Feynman–Kac path measures

$$\eta_n^*(f_n) = \gamma_n^*(f_n)/\gamma_n^*(1) \qquad \text{with } \gamma_n^*(f_n) = \mathbb{E}(f_n(X_{0:n}) W_{0:n-1}(X_{1:n-1})), \tag{2.1}$$

$$\widehat{\eta}_n^*(f_n) = \widehat{\gamma}_n^*(f_n)/\widehat{\gamma}_n^*(1) \qquad \text{with } \widehat{\gamma}_n^*(f_n) = \mathbb{E}(f_n(X_{0:n}) W_{0:n}(X_{1:n})), \tag{2.2}$$

where $(X_n)_{n \geq 0}$ is a Markov chain on $(E_n, \mathscr{E}_n)_{n \geq 0}$ with transition kernels $(M_n)_{n > 0}$, $(G_n)_{n > 0}$ is a sequence of non-negative potential functions on $(E_n)_{n > 0}$ and the importance weight function is defined by

$$W_{p,q} : x_{p+1:q} \in E_{p+1} \times \cdots \times E_q \mapsto W_{p,q}(x_{p+1:q}) := \prod_{p < k \leq q} G_k(x_k). \tag{2.3}$$

The basic SMC method proceeds as follows. Given N particles distributed approximately according to η_{n-1}^*, these particles first evolve according to the transition kernel M_n. In a second stage, particles with low relative G_n-potential value are killed and those with a larger relative potential are duplicated. However, as noted in Section 1, resampling at each time step is wasteful and should only be performed when necessary.

This has motivated researchers to introduce new resampling strategies where the resampling step is only triggered when a criterion is satisfied; this is typically computed via the current particle approximation (see Section 2.3). Such adaptive SMC algorithms proceed as follows. Let t_n^N denote the nth resampling time of the adaptive SMC algorithm. After the nth resampling step, assume we have the following empirical measure approximation of $\widehat{\eta}_{t_n}^{*N}$ denoted

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

where $\widehat{\mathcal{Y}}_n^{(N,i)} := \widehat{Y}_{0:t_n^N}^{(N,i)}$. We propagate forward these N paths by generating $Y_{t_n^N+1:t_{n+1}^N}^{(N,i)}$ according to the transition kernel $M_{t_n^N+1:t_{n+1}^N} := M_{t_n^N+1} M_{t_n^N+2} \cdots M_{t_{n+1}^N}$ of the reference

Markov chain initialized at $\widehat{Y}^{(N,i)}_{t^N_n}$, up to the first time (t^N_{n+1}) the importance weights of the N path samples given by $W_{t^N_n,t^N_{n+1}}(Y^{(N,i)}_{t^N_n+1:t^N_{n+1}})$ become, in some sense, degenerate.

At time t^N_{n+1} the weighted occupation measure of the system

$$\widetilde{\eta}^{*N}_{t^N_{n+1}}(\cdot) := \sum_{i=1}^{N} \frac{W_{t^N_n,t^N_{n+1}}(Y^{(N,i)}_{t^N_n+1:t^N_{n+1}})}{\sum_{j=1}^{N} W_{t^N_n,t^N_{n+1}}(Y^{(N,j)}_{t^N_n+1:t^N_{n+1}})} \delta_{\mathcal{Y}^{(N,i)}_{n+1}}(\cdot)$$

is a particle approximation of $\widehat{\eta}^{*}_{t^N_{n+1}}(\cdot)$, where $\mathcal{Y}^{(N,i)}_{n+1} := (\widehat{\mathcal{Y}}^{(N,i)}_{n}, Y^{(N,i)}_{t^N_n+1:t^N_{n+1}})$. After the resampling step, this measure is replaced by an empirical measure

$$\widehat{\eta}^{*N}_{t^N_{n+1}}(\cdot) := \frac{1}{N}\sum_{i=1}^{N} \delta_{\widehat{\mathcal{Y}}^{(N,i)}_{n+1}}(\cdot)$$

associated with N path particles $\widehat{\mathcal{Y}}^{(N,i)}_{n+1} := \widehat{Y}^{(N,i)}_{0:t^N_{n+1}}$ that are resampled from $\widetilde{\eta}^{*N}_{t^N_{n+1}}$; see, for example, [7] for alternative resampling schemes.

2.3. Some empirical criteria

Two well-known criteria used in the SMC literature to trigger the resampling mechanism are now discussed. In both cases, the resampling times $(t^N_n)_{n\geq 0}$ are random variables that depend on the current SMC approximation.

• *Squared coefficient of variation.* After the resampling step at time t^N_n, the particles explore the state space up to the first time $(s = t^N_{n+1})$ the squared coefficient of variation of the unnormalized weights is larger than some prescribed threshold a_n

$$C^N_{t^N_n,s} = \frac{1}{N}\sum_{i=1}^{N}\left(W_{t^N_n,s}(Y^{(N,i)}_{t^N_n+1:s}) \Big/ \frac{1}{N}\sum_{j=1}^{N} W_{t^N_n,s}(Y^{(N,j)}_{t^N_n+1:s})\right)^2 - 1 \geq a_n. \qquad (2.4)$$

This is equivalent to resampling when the effective sample size (ESS), defined as $ESS = N(1+C^N_{t^N_n,s})^{-1}$, is below a prescribed threshold as proposed in [9].

• *Entropy.* After the resampling step at time t^N_n, the particles explore the state space up to the first time $(s = t^N_{n+1})$ the relative entropy of the empirical particle measure w.r.t. its weighted version is larger than some threshold a_n

$$C^N_{t^N_n,s} := -\frac{1}{N}\sum_{i=1}^{N} \log W_{t^N_n,s}(Y^{(N,i)}_{t^N_n+1:s}) \geq a_n. \qquad (2.5)$$

2.4. Statement of some results

The following section provides a guide of the major definitions and results in this paper; these will be repeated at the relevant stages 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 $C^N_{t^N_n,s}$ by their limiting values $C_{t_n,s}$ as $N \uparrow \infty$, that is, $t_{n+1} := \inf\{t_n < s\colon C_{t_n,s} \geq a_n\}$. In all situations, set $t^N_0 = t_0 = 0$. For the criterion (2.4), the limiting criterion $C_{t_n,s}$ is given by

$$\frac{\mathbb{E}_{t_n,\widehat{\eta}^*_{t_n}}(W_{t_n,s}(X_{t_n+1:s})^2)}{\mathbb{E}_{t_n,\widehat{\eta}^*_{t_n}}(W_{t_n,s}(X_{t_n+1:s}))^2} - 1, \tag{2.6}$$

whereas for (2.5) it is given by

$$-\,\mathbb{E}_{t_n,\widehat{\eta}^*_{t_n}}(\log W_{t_n,s}(X_{t_n+1:s})). \tag{2.7}$$

Here, $\mathbb{E}_{t_n,\widehat{\eta}^*_{t_n}}$ is the expectation w.r.t. the law $\mathbb{P}_{t_n,\widehat{\eta}^*_{t_n}}$ of the random path of variables that starts at time t_n at the end point $X_{t_n} = \widehat{X}_{t_n}$ of $\widehat{\mathcal{X}}_n := \widehat{X}_{0:t_n}$ distributed according to $\widehat{\eta}^*_{t_n}$ and evolves according to the Markov kernels $M_{t_n+1:s}$. In [3], the limiting expression for the normalized effective sample size $N^{-1}ESS$ has been established. An alternative entropy criterion has also been proposed that can be applied when the potential functions $(G_n)_{n>0}$ are not strictly positive on $(E_n)_{n>0}$.

2.4.2. An exponential coupling theorem

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

$$(G) \quad \forall n \geq 1 \qquad q'_n := \sup_{(x,y)\in(E_n)^2} (G_n(x)/G_n(y)) < \infty. \tag{2.8}$$

We refer the reader to [4], Chapter 3, for a thorough discussion in the case where (G) does not apply. To state our results, we first require the following definition.

Definition 2.1. *Let $\mathcal{Y}^{(N)}_n := (\mathcal{Y}^{(N,1)}_n, \mathcal{Y}^{(N,2)}_n, \dots, \mathcal{Y}^{(N,N)}_n)$ and $\widehat{\mathcal{Y}}^{(N)}_n := (\widehat{\mathcal{Y}}^{(N,1)}_n, \widehat{\mathcal{Y}}^{(N,2)}_n, \dots, \widehat{\mathcal{Y}}^{(N,N)}_n)$ denote the N particles associated to the adaptive SMC algorithm resampling at times $(t^N_n)_{n\geq 0}$ and let $\mathcal{X}^{(N)}_n := (\mathcal{X}^{(N,1)}_n, \mathcal{X}^{(N,2)}_n, \dots, \mathcal{X}^{(N,N)}_n)$ and $\widehat{\mathcal{X}}^{(N)}_n := (\widehat{\mathcal{X}}^{(N,1)}_n, \widehat{\mathcal{X}}^{(N,2)}_n, \dots, \widehat{\mathcal{X}}^{(N,N)}_n)$ denote the N particles associated to the reference SMC algorithm resampling at times $(t_n)_{n\geq 0}$. We also suppose that $(\mathcal{X}^{(N)}_n, \widehat{\mathcal{X}}^{(N)}_n)$ and $(\mathcal{Y}^{(N)}_n, \widehat{\mathcal{Y}}^{(N)}_n)$ coincide on every time interval $0 \leq n \leq m$, once $t^N_n = t_n$, for every $0 \leq n \leq m$. This condition corresponds to the coupling of the two processes on the event $\bigcap_{0\leq n\leq m}\{t^N_n = t_n\}$.*

The first result is a non-asymptotic exponential concentration estimate. The probability measures η^N_n and η_n are introduced below. They can be thought of as analogues of $\eta^{*N}_{t_n}$ and $\eta^*_{t_n}$; see Section 3 for formal definitions.

Theorem 2.2. *For any $n \geq 0$, $f_n \in \operatorname{Osc}_1(E_0 \times \cdots \times E_{t_n})$, any $N \geq 1$ and any $0 \leq \varepsilon \leq 1/2$, there exist $c_1 < \infty$, $0 < c_2(n) < \infty$ such that we have the exponential concentration estimate*

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

for the empirical measures $\eta_n^N(\cdot) = \frac{1}{N}\sum_{i=1}^N \delta_{(\widehat{\mathcal{X}}_{n-1}^{(N,i)}, X_{t_{n-1}+1:t_n}^{(N,i)})}(\cdot)$*. In addition, under appropriate regularity conditions on* $(M_k)_{k>0}$ *and* $(G_k)_{k>0}$ *given in Section 4.2.1 the above estimates are valid for the marginal measures associated to the time parameters* $t_{n-1} \le p \le t_n$ *for some constant* $c_2(n) = c_2$*.*

The second result is an exponential coupling theorem.

Theorem 2.3. *Assume the threshold parameters* $(a_n)_{n\ge 0}$ *are sampled realizations of a collection of absolutely continuous random variables* $A = (A_n)_{n\ge 0}$*. Then, for almost every realization of the sequence* $(A_n)_{n\ge 0}$*,* $(\mathcal{X}_n^{(N)}, \widehat{\mathcal{X}}_n^{(N)})_{n\ge 0}$ *and* $(\mathcal{Y}_n^{(N)}, \widehat{\mathcal{Y}}_n^{(N)})_{n\ge 0}$ *are such that, for every* $m \ge 0$ *and any* $N \ge 1$*, there exist* $0 < c_1(m), c_2(m) < \infty$ *and almost surely* $\varepsilon(m, A) \equiv \varepsilon(m) > 0$ *such that*

$$\mathbb{P}(\exists 0 \le n \le m\ (\mathcal{Y}_n^{(N)}, \widehat{\mathcal{Y}}_n^{(N)}) \ne (\mathcal{X}_n^{(N)}, \widehat{\mathcal{X}}_n^{(N)}) | A) \le c_1(m) \mathrm{e}^{-N\varepsilon^2(m)/c_2(m)}.$$

Up to an event having an exponentially small occurrence probability, Theorem 2.3 allows us to transfer many estimates of the reference SMC algorithm $(\mathcal{X}_n^{(N)}, \widehat{\mathcal{X}}_n^{(N)})_{n\ge 0}$ resampling at deterministic times to the adaptive SMC algorithm $(\mathcal{Y}_n^{(N)}, \widehat{\mathcal{Y}}_n^{(N)})_{n\ge 0}$.

The proofs of Theorems 2.2 and 2.3 are detailed, respectively, in Sections 4.2.3 and 5.2.

3. Description of the models

3.1. Feynman–Kac distributions flow

We consider a sequence of measurable state spaces $(S_n, \mathscr{S}_n)_{n\ge 0}$, a probability measure $\eta_0 \in \mathcal{P}(S_0)$ and a sequence of Markov transitions $\mathcal{M}_n(x_{n-1}, \mathrm{d}x_n)$ from S_{n-1} into S_n for $n \ge 1$. Let $(\mathcal{X}_n)_{n\ge 0}$ be a Markov chain with initial distribution $\mathrm{Law}(\mathcal{X}_0) = \eta_0$ and elementary transitions $\mathbb{P}(\mathcal{X}_n \in \mathrm{d}y | \mathcal{X}_{n-1} = x) = \mathcal{M}_n(x, \mathrm{d}y)$. Let $(\mathcal{G}_n)_{n\ge 0}$ be a sequence of non-negative and bounded potential functions on S_n. To simplify the presentation, and to avoid unnecessary technicalities, it is supposed $\mathcal{G}_n \in (0,1)$ for $n \ge 1$ with

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

The Boltzmann–Gibbs transformation Ψ_n associated to $\mathcal{G}_n$ is the mapping

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

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

$$\Psi_n(\mu)(\mathrm{d}y) = (\mu \mathcal{S}_{n,\mu})(\mathrm{d}y) = \int_{S_n} \mu(\mathrm{d}x) \mathcal{S}_{n,\mu}(x, \mathrm{d}y)$$

with $\mathcal{S}_{n,\mu}(x, \mathrm{d}y) := \mathcal{G}_n(x)\delta_x(\mathrm{d}y) + (1 - \mathcal{G}_n(x))\mathcal{G}_n(y)\mu(\mathrm{d}y)/\mu(\mathcal{G}_n)$.

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

$$\forall n \geq 0 \qquad \eta_{n+1} = \widehat{\eta}_n \mathcal{M}_{n+1} \qquad \text{with } \widehat{\eta}_n := \Psi_n(\eta_n) = \eta_n \mathcal{S}_{n,\eta_n}. \tag{3.2}$$

It can be checked that the solution $(\eta_n, \widehat{\eta}_n)$ of these recursive updating prediction equations have the following functional representations:

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

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

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

3.2. Feynman–Kac semigroups

To analyze SMC methods, we introduce the Feynman–Kac semigroup associated to the flow of measures $(\gamma_n)_{n\geq 0}$ and $(\eta_n)_{n\geq 0}$. Let us start by denoting by $\mathcal{Q}_{n+1}(x_n, \mathrm{d}x_{n+1})$ the bounded integral operator from S_n into S_{n+1} defined by

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

Let $(\mathcal{Q}_{p,n})_{0\leq p\leq n}$ be the corresponding linear semigroup defined by $\mathcal{Q}_{p,n} = \mathcal{Q}_{p+1}\mathcal{Q}_{p+2}\cdots\mathcal{Q}_n$ with the convention $\mathcal{Q}_{n,n} = I$, the identity operator. Note that $\mathcal{Q}_{p,n}$ is alternatively defined by

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

Using the Markov property, it follows that

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

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

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

Finally, denote by $(\Phi_{p,n})_{0\leq p\leq n}$ the nonlinear semigroup associated to the flow of normalized measures $(\eta_n)_{n\geq 0}$: $\Phi_{p,n} = \Phi_n \circ \cdots \circ \Phi_{p+2} \circ \Phi_{p+1}$ with the convention $\Phi_{n,n} = I$, the identity operator and $\Phi_n(\mu) = \mu(\mathcal{G}_{n-1}\mathcal{M}_n)/\mu(\mathcal{G}_{n-1})$, $\mu \in \mathcal{P}(S_{n-1})$. Note that

$(\Phi_{p,n})_{0\leq p\leq n}$ can be alternatively defined in terms of $(\mathcal{Q}_{p,n})_{0\leq p\leq n}$ using

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

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 $M_n(x_{n-1}, \mathrm{d}x_n)$ and initial distribution $\eta_0 = \mathrm{Law}(X_0)$. In addition, introduce a sequence of non-negative potential functions $(G_n)_{n>0}$ on the state spaces $(E_n)_{n>0}$. To simplify the presentation, it is assumed that $G_n \in (0,1)$ $\forall n > 0$.

We associate to an increasing sequence of time parameters $(t_n)_{n\geq 0}$ the excursion-valued random variables X_0 for $n = 0$ and $X_{t_{n-1}+1:t_n}$ for $n \geq 1$. We also define the random path sequences

$$\mathcal{X}_n := X_{0:t_n} \in E'_{t_n}$$

with the convention $E'_n := E_0 \times \cdots \times E_n$. Note that $(\mathcal{X}_n)_{n\geq 0}$ forms a Markov chain

$$\mathcal{X}_{n+1} := (\mathcal{X}_n, X_{t_n+1:t_{n+1}}) \tag{3.8}$$

taking values in the excursion spaces $S_n := E'_{t_n}$. Now adopting the potential functions

$$\forall n \geq 1 \qquad \mathcal{G}_n(\mathcal{X}_n) := W_{t_{n-1}:t_n}(X_{t_{n-1}+1:t_n}) \tag{3.9}$$

in (3.4), we readily find that

$$\gamma_n(f_n) = \mathbb{E}\left[f_n(\mathcal{X}_n) \prod_{0<k<n} \mathcal{G}_k(\mathcal{X}_k)\right] = \mathbb{E}[f_n(X_{0:t_n}) W_{0:t_{n-1}}(X_{1:t_{n-1}})].$$

By definition of the potential functions $\mathcal{G}_n$ of the excursion Feynman–Kac model (3.9), it is easily proved that the condition $(\mathcal{G})$ (equation (3.1)) is satisfied as soon as (G) introduced in (2.8) holds true. More precisely, it holds that (G) implies $(\mathcal{G})$ with

$$q_n \leq M - \sup\left\{\frac{W_{t_{n-1}:t_n}(x_{t_{n-1}+1:t_n})}{W_{t_{n-1}:t_n}(y_{t_{n-1}+1:t_n})}\right\} \qquad \left(\leq \prod_{t_{n-1}<k\leq t_n} q'_k\right),$$

where the essential supremum $M - \sup\{\cdot\}$ is taken over all admissible paths $x_{t_{n-1}+1:t_n}$ and $y_{t_{n-1}+1:t_n}$ of the underlying Markov chain $(X_n)_{n\geq 0}$.

3.4. Functional criteria

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

Definition 3.1. *We consider a sequence of functional criteria*

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

satisfying the following Lipschitz type regularity condition

$$|\mathcal{H}^{(n)}_{p,q}(\mu_1) - \mathcal{H}^{(n)}_{p,q}(\mu_2)| \leq \delta(H^{(n)}_{p,q}) \int |[\mu_1 - \mu_2](h)| H^{(n)}_{p,q}(\mathrm{d}h) \tag{3.10}$$

for some collection of bounded measures $H^{(n)}_{p,q}$ *on* $\mathcal{B}_b(E'_q)$ *such that*

$$\delta(H^{(n)}_{p,q}) := \int \mathrm{osc}(h) H^{(n)}_{p,q}(\mathrm{d}h) < \infty.$$

We illustrate this construction with the pair of functional criteria discussed in Section 2.4.1. When we consider (2.6), the functional

$$\mathcal{H}^{(n)}_{p,q}(\mu) = \mu\left(\left[\frac{W_{p,q}}{\mu(W_{p,q})} - 1\right]^2\right) \tag{3.11}$$

coincides with the squared coefficient of variation of the weights w.r.t. μ. When we consider (2.7), the functional

$$\mathcal{H}^{(n)}_{p,q}(\mu) = \mathrm{Ent}(\mathrm{d}\mu | W_{p,q}\mathrm{d}\mu) := -\mu(\log W_{p,q}) \tag{3.12}$$

measures the relative entropy distance between μ and the updated weighted measure. Under the condition $(\mathcal{G})$ stated in (3.1), it is an elementary exercise to check that the above pair of criteria satisfy (3.10). In the first case (3.11), we can take $H^{(n)}_{p,q} = c[\delta_{W^2_{p,q}} + \delta_{W_{p,q}}]$ for some constant c sufficiently large. In the second case (3.12), we can take $H^{(n)}_{p,q} = c\delta_{W_{p,q}}$, again for some c large enough.

3.5. Resampling times construction

We now explain how to define the sequence of resampling times $(t_n)_{n\geq 0}$. This requires introducing the measure $\mathbb{P}_{\eta,(p,n)} \in \mathcal{P}(E'_n)$ defined for any pair of integers $0 \leq p \leq n$ and any $\eta \in \mathcal{P}(E'_p)$ by

$$\begin{aligned}\mathbb{P}_{\eta,(p,n)}(dx_{0:n}) &:= \eta(dx_{0:p}) M_{p+1}(x_p, dx_{p+1}) \cdots M_n(x_{n-1}, dx_n) \\ &\in \mathcal{P}(E'_p \times (E_{p+1} \times \cdots \times E_n)) = \mathcal{P}(E'_n),\end{aligned} \tag{3.13}$$

where $dx_{0:n}$ denotes an infinitesimal neighborhood of a path sequence $x_{0:n} \in E'_n$.

Given $\mathcal{H}^{(n)}_{p,q}$, with $n \geq 0$ and $0 \leq p \leq q$, we define an increasing sequence of deterministic time steps $(t_n)_{n\geq 0}$ and a flow of Feynman–Kac measures $(\eta_n, \widehat{\eta}_n)$ by induction as follows. Suppose that the resampling time t_n is defined as well as $(\eta_n, \widehat{\eta}_n) \in \mathcal{P}(E'_{t_n})^2$. The resampling time t_{n+1} is defined as the first time $(s > t_n)$ the quantity $\mathcal{H}^{(n)}_{t_n,s}(\mathbb{P}_{\widehat{\eta}_n,(t_n,s)})$ hits the set $I_n = [a_n, \infty)$; that is, $t_{n+1} := \inf\{t_n < s\colon\ \mathcal{H}^{(n)}_{t_n,s}(\mathbb{P}_{\widehat{\eta}_n,(t_n,s)}) \in I_n\}$. Given t_{n+1}, we set

$$\eta_{n+1} = \mathbb{P}_{\widehat{\eta}_n,(t_n,t_{n+1})} \quad \text{and} \quad \widehat{\eta}_{n+1} = \Psi_{n+1}(\eta_{n+1}) \tag{3.14}$$

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

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

$$\eta_{n+1} = \mathbb{P}_{\widehat{\eta}_n,(t_n,t_{n+1})} = \widehat{\eta}_n \mathcal{M}_{n+1}. \tag{3.15}$$

This yields the recursion ((3.14) and (3.15)) $\Longrightarrow \eta_{n+1} = \Psi_n(\eta_n)\mathcal{M}_{n+1}$. Hence the flow of measures η_n and $\widehat{\eta}_n$ 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 SMC approximation of these distributions is studied in Section 4.

3.6. Some applications

In this section, we examine the inductive construction of the deterministic resampling times $(t_n)_{n\geq 0}$ introduced in Section 3.5 for the criteria (2.6) and (2.7).

- *Squared coefficient of variation.* In this case, we have

$$\mathcal{H}^{(n)}_{t_n,s}(\mathbb{P}_{\widehat{\eta}_n,(t_n,s)}) = \frac{\mathbb{E}_{n,\widehat{\eta}_n}(W_{t_n,s}(X_{t_n+1:s})^2)}{\mathbb{E}_{n,\widehat{\eta}_n}(W_{t_n,s}(X_{t_n+1:s}))^2} - 1.$$

The mappings $s \mapsto \mathcal{H}^{(n)}_{t_n,s}(\mathbb{P}_{\widehat{\eta}_n,(t_n,s)})$ are generally increasing. One natural way to control these variances is to choose an interval $I_n := [a_n, \infty)$, with $a_n > 0$, then

$$t_{n+1} := \inf\{t_n < s\colon\ \mathbb{E}_{n,\widehat{\eta}_n}(W_{t_n,s}(X_{t_n+1:s})^2) \geq [1 + a_n]\mathbb{E}_{n,\widehat{\eta}_n}(W_{t_n,s}(X_{t_n+1:s}))^2\}.$$

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

$$\mathcal{H}^{(n)}_{t_n,s}(\mathbb{P}_{\widehat{\eta}_n,(t_n,s)}) = \mathrm{Ent}(\mathbb{P}_{\widehat{\eta}_n,(t_n,s)}|\mathbb{Q}_{\widehat{\eta}_n,(t_n,s)}) = -\mathbb{E}_{n,\widehat{\eta}_n}(\log W_{t_n,s}(X_{t_n+1:s}))$$

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

$$\mathbb{Q}_{\eta,(p,n)}(dx_{0:n}) = \mathbb{P}_{\eta,(p,n)}(dx_{0:n}) \times W_{p,n}(x_{p+1:n}).$$

If we choose an interval $I_n := [a_n, \infty)$, with $a_n > 0$, then the resampling time t_{n+1} coincides with the first time the entropy distance goes above the level a_n; that is,

$$t_{n+1} := \inf\{t_n < s\colon\ \mathrm{Ent}(\mathbb{P}_{\widehat{\eta}_n,(t_n,s)}|\mathbb{Q}_{\widehat{\eta}_n,(t_n,s)}) \geq a_n\}.$$

4. Convergence analysis of the reference SMC algorithm

4.1. A reference SMC algorithm

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

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

with elementary transitions

$$\mathbb{P}(\mathcal{X}_{n+1}^{(N)} \in B_1 \times \cdots \times B_N | \mathcal{X}_n^{(N)}) = \int_{B_1 \times \cdots \times B_N} \prod_{i=1}^{N} \mathcal{K}_{n+1,\eta_n^N}(\mathcal{X}_n^{(N,i)}, dx_{n+1}^i), \tag{4.1}$$

where $B_i \in \mathscr{S}_{n+1}$ for every $i \in \{1, \ldots, N\}$ and

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

This integral decomposition shows that the SMC algorithm has a similar updating/prediction nature as the one of the 'limiting' Feynman–Kac model. More precisely, the deterministic two-step updating/prediction transitions in distribution spaces

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

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

$$\mathcal{X}_n^{(N)} \in S_n^N \xrightarrow{\text{resampling}} \widehat{\mathcal{X}}_n^{(N)} \in S_n^N \xrightarrow{\text{mutation}} \mathcal{X}_{n+1}^{(N)} \in S_{n+1}^N. \tag{4.4}$$

In our context, the SMC algorithm keeps track of all the paths of the sampled particles and the corresponding ancestral lines are denoted by $\widehat{\mathcal{X}}_n^{(N,i)} = \widehat{X}_{0:t_n}^{(N,i)}$ and $\mathcal{X}_n^{(N,i)} = X_{0:t_n}^{(N,i)} \in S_n$, where we recall that $S_n = E'_{t_n}$. By definition of the reference Markov model $\mathcal{X}_n$ given in (3.8), every path particle $\mathcal{X}_{n+1}^{(N,i)} \in S_{n+1}$ keeps track of the selected excursion $\widehat{\mathcal{X}}_n^{(N,i)} \in S_n$ and it evolves from its terminal state $\widehat{X}_{t_n,t_n}^{(N,i)}$ with $(t_{n+1} - t_n)$ elementary moves using the Markov transition $M_{t_n+1:t_{n+1}}$. More formally, we have that

$$\mathcal{X}_{n+1}^{(N,i)} = (\widehat{X}_{0:t_n}^{(N,i)}, X_{t_n+1:t_{n+1}}^{(N,i)}) = (\widehat{\mathcal{X}}_n^{(N,i)}, X_{t_n+1:t_{n+1}}^{(N,i)}).$$

From this discussion, it is worth mentioning a further convention that the particle empirical measures $\eta_{n+1}^N(\cdot) = \frac{1}{N}\sum_{i=1}^N \delta_{(\widehat{\mathcal{X}}_n^{(N,i)}, X_{t_n+1:t_{n+1}}^{(N,i)})}(\cdot)$ are the terminal values at time $s = t_{n+1}$ of the flow of random measures

$$t_n \le s \le t_{n+1} \mapsto \mathbb{P}_{\widehat{\eta}_n^N,(t_n,s)}^N(\cdot) = \frac{1}{N} \sum_{i=1}^{N} \delta_{(\widehat{\mathcal{X}}_n^{(N,i)}, X_{t_n+1:s}^{(N,i)})}(\cdot). \tag{4.5}$$

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

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

with $\mathcal{Q}_{p,n}$ as in (3.5) and $\mathcal{P}_{p,n}$ in equation (3.6). These parameters can be expressed in terms of the mixing properties of the Markov transitions $\mathcal{M}_n$; see [4], Chapter 4. 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 w.r.t. 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. As a result, these non-asymptotic results can be converted into time uniform convergence results. To get a flavor of these uniform estimates, assume that the Markov transitions $\mathcal{M}_k$ satisfy the following regularity property.

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

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

with $\mathcal{M}_{p,p+m} := \mathcal{M}_{p+1}\mathcal{M}_{p+2}\cdots\mathcal{M}_{p+m}$.

We also introduce the following quantities:

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

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

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

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

The proof of these estimates relies on semigroup techniques; see [4], Chapter 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 $m=1$, $\delta = \bigwedge_n \delta_n > 0$ and $q = \bigvee_{n\geq 1} q_n$. In this case, $q_{p,p+n} \leq \delta^{-1} q$ and $\beta(\mathcal{P}_{p,p+n}) \leq (1-\delta^2)^n$ imply that

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

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

$$\bigwedge_p \delta_p := \delta > 0, \qquad \bigvee_p r_{p,p+m} := \overline{r} < \infty \quad \text{and} \quad \bigvee_p r_{p+1,p+m} := \underline{r} < \infty. \tag{4.7}$$

In this situation, $q_{p,p+n} \leq \delta^{-1}\overline{r}$ and $\beta(\mathcal{P}_{p,p+n}) \leq (1-\delta^2\underline{r}^{-1})^{\lfloor n/m \rfloor}$ and therefore

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

See [4], Chapter 3, for a discussion of when $(\mathcal{M})_m$ holds. We also mention that this mixing condition is never met for $\mathcal{X}_n = (X_p)_{0 \leq p \leq t_n}$ on $S_n = E'_{t_n}$ 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 $\forall k \geq 1$, $\forall (x,y) \in (E_k)^2$, $M_k(x,\cdot) \geq \delta' M_k(y,\cdot)$ 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*

$$\eta_n^N = \eta_{n-1}^N \mathcal{K}_{n,\eta_{n-1}^N} + \frac{1}{\sqrt{N}} V_n^N \qquad (\Longleftrightarrow V_n^N := \sqrt{N}[\eta_n^N - \eta_{n-1}^N \mathcal{K}_{n,\eta_{n-1}^N}]). \tag{4.9}$$

For $n=0$, the conventions $\mathcal{K}_{0,\eta_{-1}^N}(x,\mathrm{d}y) = \eta_0(\mathrm{d}y)$ 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 Khinchine type inequality for martingales with symmetric and independent increments. This is a well-known result.

Lemma 4.2 (Khinchine's inequality). *Let $L_n^\Delta := \sum_{0\leq p\leq n} \Delta_p$ 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*

$$\mathbb{E}(|L_n^\Delta|^m)^{1/m} \leq b(m)\mathbb{E}([L^\Delta]_n^{m'/2})^{1/m'} \qquad \text{with } [L^\Delta]_n := \sum_{0\leq p\leq n} \Delta_p^2, \tag{4.10}$$

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

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

with $(2m)_m = (2m)!/(2m-m)!$.

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*

$$\mathbb{E}(|V_n^N(f_n)|^m | \mathcal{F}_{n-1}^{(N)})^{1/m} \leq b(m) \operatorname{osc}(f_n), \tag{4.12}$$

where $(\mathcal{F}_n^{(N)})_{n \geq 0}$ is the filtration generated by the N-particle system.

Proof. By construction, we have

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

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

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

This yields the formula $V_n^N(f_n) = \mathbb{E}(L_{n,N}^{(N)}(f_n) | \mathcal{F}_n^{(N)})$, where $L_{n,N}^{(N)}(f_n)$ is the terminal value of the martingale sequence defined by

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

Then as

$$\begin{aligned}\mathbb{E}(|V_n^N(f_n)|^m | \mathcal{F}_{n-1}^{(N)})^{1/m} &= \mathbb{E}(|\mathbb{E}(L_{n,N}^{(N)}(f_n) | \mathcal{F}_n^{(N)})|^m | \mathcal{F}_{n-1}^{(N)})^{1/m} \\ &\leq \mathbb{E}(|L_{n,N}^{(N)}(f_n)|^m | \mathcal{F}_{n-1}^{(N)})^{1/m},\end{aligned}$$

one may apply Khinchine's inequality to conclude. □

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 $f_n \in \operatorname{Osc}_1(S_n)$, we have the first-order decomposition for the nonlinear semigroup $\Phi_{p,n}$ defined in (3.7):*

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

where

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

with $\mathcal{U}_{p,n,\eta}(f), \mathcal{V}_{p,n,\eta}(f), \mathcal{W}_{p,n,\eta}(f)$ a collection of functions in $\operatorname{Osc}_1(S_p)$ whose values only depend 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 $n \geq 0$, $f_n \in \mathrm{Osc}_1(S_n)$ and any $N \geq 1$,*

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

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

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

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 $\eta_{-1}^N(f) := f$ and the pair of random measures $(\mathcal{I}_n^N, \mathcal{J}_n^N)$ given for any $f_n \in \mathrm{Osc}_1(S_n)$ by

$$\mathcal{I}_n^N(f_n) := 2\sum_{p=0}^{n} q_{p,n}\beta(\mathcal{P}_{p,n}) V_p^N(\mathcal{U}_{p,n,\Phi_p(\eta_{p-1}^N)}(f_n)),$$

$$\mathcal{J}_n^N(f_n) := \sqrt{N}\sum_{p=0}^{n} \mathcal{R}_{p,n}(\eta_p^N, \Phi_p(\eta_{p-1}^N))(f_n).$$

Now, observe that

$$\mathbb{E}(W_n^N(f_n)) = \mathbb{E}(\mathcal{J}_n^N(f_n)). \tag{4.13}$$

Using Proposition 4.3, for any $f_n \in \mathrm{Osc}_1(S_n)$ it can be checked that $\mathbb{E}(|\mathcal{I}_n^N(f_n)|^m)^{1/m} \leq b(m)\sigma_{2,n}$. In a similar way, we find that

$$\sqrt{N}\mathbb{E}(|\mathcal{J}_n^N(f_n)|^m)^{1/m} \leq b(2m)^2\sigma_{1,n}. \tag{4.14}$$

The first part of the proof then follows from (4.13) and (4.14); the remainder of the proof is now clear. □

4.2.3. A concentration theorem

The following concentration theorem is the main result of this section.

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

$$\mathbb{P}(|[\eta_n^N - \eta_n](f_n)| \geq \varepsilon) \leq 6 \exp\left(-\frac{N\varepsilon^2}{8\sigma_{1,n}}\right), \tag{4.15}$$

where the constant $\sigma_{1,n}$ is as 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 $(\underline{r}, \overline{r})$. In this situation, for any value of the time parameter n, for any $f_n \in \mathrm{Osc}_1(S_n)$, $N \geq 1$ and for any $\rho \in (0,1)$, the probability that

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

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

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

Using (4.14), for any $f_n \in \mathrm{Osc}_1(S_n)$ and $0 \leq s < 1/(2\sigma_{1,n})$, it follows that

$$\mathbb{E}(\exp\{s\sqrt{N}\mathcal{J}_n^N(f_n)\}) \leq \sum_{m \geq 0} \frac{(s\sigma_{1,n})^m}{m!} b(2m)^{2m} = \frac{1}{\sqrt{1-2s\sigma_{1,n}}}. \tag{4.16}$$

To simplify the presentation, set

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

where $\mathcal{U}_{p,n,\eta}(\cdot)$ was introduced in Lemma 4.4. By the definition of V_p^N

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

Recalling that $\mathbb{E}(\mathrm{e}^{tX}) \leq \mathrm{e}^{t^2c^2/2}$ for every real-valued and centered random variable X with $|X| \leq c$ (e.g., [4], Lemma 7.3.1), we prove that

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

Iterating the argument, we find that

$$\mathbb{E}(\mathrm{e}^{t\mathcal{I}_n^N(f_n)}) = \mathbb{E}\left(\exp\left\{t\sum_{p=0}^{n}\alpha_{p,n}V_p^N(f_{p,n}^N)\right\}\right) \le \exp\left(\frac{t^2\sigma_n^2}{2}\right) \tag{4.17}$$

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

From these upper bounds, the proof of the exponential estimates now follows standard arguments. Indeed, for any $0 \le s < 1/(2\sigma_{1,n})$ and any $\varepsilon > 0$, by (4.16) we have

$$\mathbb{P}(\sqrt{N}\mathcal{J}_n^N(f_n) \ge \varepsilon) \le \frac{1}{\sqrt{1-2s\sigma_{1,n}}}\exp\{-\varepsilon s\}.$$

Replacing ε by εN and choosing $s = 3/(8\sigma_{1,n})$ yields

$$\mathbb{P}(\mathcal{J}_n^N(f_n)/\sqrt{N} \ge \varepsilon) \le 2\exp\{-\varepsilon N/(3\sigma_{1,n})\}.$$

To estimate the probability tails of $\mathcal{I}_n^N(f_n)$, we use (4.17) and the fact that $\varepsilon > 0$ and $t \ge 0$

$$\mathbb{P}(\mathcal{I}_n^N(f_n) \ge \varepsilon) \le \exp\left\{-\left(\varepsilon t - \frac{t^2}{2}\sigma_n^2\right)\right\}.$$

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

$$\forall \varepsilon > 0 \qquad \mathbb{P}(\mathcal{I}_n^N(f)/\sqrt{N} \ge \varepsilon) \le \exp\left(-\frac{N\varepsilon^2}{2\sigma_n^2}\right).$$

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 $\alpha \in [0,1]$

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

From previous calculations,

$$\mathbb{P}([\eta_n^N - \eta_n](f_n) \ge \varepsilon) \le \exp\left(-\frac{N\varepsilon^2\alpha^2}{2\sigma_n^2}\right) + 2\exp\left(-\frac{N\varepsilon(1-\alpha)}{3\sigma_{1,n}}\right). \tag{4.18}$$

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

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

It remains to observe that $q_{p,n} \geq 1$ and $\beta(\mathcal{P}_{p,n}) \leq 1 \Longrightarrow \sigma_n^2 \leq \sigma_{1,n}$ and

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

so that

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

The end of the proof of (4.15) is now easily completed. We now 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 $(\underline{r}, \overline{r})$. By (4.8) the following uniform concentration estimate holds

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

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

Remark 4.7. Returning to the end of the proof of Theorem 4.6, the exponential concentration estimates can be marginally improved by choosing, in (4.18), the parameter $\alpha = \alpha_n(\varepsilon) \in [0,1]$ such that $a_n(\varepsilon)\alpha^2 = b_n(1-\alpha)$, with $a_n(\varepsilon) := \frac{\varepsilon}{2\sigma_n^2}$, $b_n = \frac{1}{3\sigma_{1,n}}$ and $\sigma_n^2 := 4\sum_{p=0}^n q_{p,n}^2 \beta(\mathcal{P}_{p,n})^2$. Elementary manipulations yield

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

and therefore

$$\forall \varepsilon \geq 0 \qquad \mathbb{P}(|(\eta_n^N - \eta_n)(f)| \geq \varepsilon) \leq 6 \exp\left(-N \frac{\varepsilon^2}{2\sigma_n^2} \alpha_n^2(\varepsilon)\right).$$

For small values of ε, this bound improves that in Section 7.4.3 of [4], which is of the form

$$\forall \varepsilon \geq 0 \qquad \mathbb{P}(|(\eta_n^N - \eta_n)(f)| \geq \varepsilon) \leq (1 + \varepsilon\sqrt{N}) \exp\left(-N \frac{\varepsilon^2}{2\tilde{\sigma}_n^2}\right)$$

with

$$\tilde{\sigma}_n^2 := 4\left(\sum_{p=0}^n q_{p,n} \beta(\mathcal{P}_{p,n})\right)^2 \geq \sigma_n^2.$$

4.3. Approximating the criteria

By construction, the particle occupation measures $\mathbb{P}^N_{\widehat{\eta}_n,(t_n,s)}$ approximate the measures $\mathbb{P}_{\widehat{\eta}_n,(t_n,s)}$ introduced in (3.13); that is, in some sense, $\mathbb{P}^N_{\widehat{\eta}^N_n,(t_n,s)} \simeq_{N\uparrow\infty} \mathbb{P}_{\widehat{\eta}_n,(t_n,s)}$. Conversely, observe that $\mathbb{P}^N_{\widehat{\eta}^N_n,(t_n,s)}$, respectively $\mathbb{P}_{\widehat{\eta}_n,(t_n,s)}$, are the marginals of the measures η^N_{n+1}, respectively η_{n+1}, w.r.t. the $(s-t_n)+1$ first coordinates. In other words, the measures $\mathbb{P}^N_{\widehat{\eta}^N_n,(t_n,s)}$, respectively $\mathbb{P}_{\widehat{\eta}_n,(t_n,s)}$, are the projections of the measures η^N_{n+1}, respectively η_{n+1}, on the state space $E'_s = E'_{t_n} \times (E_{t_n+1} \times \cdots \times E_s)$.

For instance, the following proposition is essentially 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 $\varepsilon > 0$, the concentration inequality:*

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

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

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

also holds, with a pair of finite constants $c_1(n), c_2(n) < \infty$.

Proof. By [4], Theorem 7.4.4, for any $N \geq 1$, $p \geq 1$, $n \geq 0$ and any test function $f_n \in \mathrm{Osc}_1(E'_{t_n})$

$$\sup_{N\geq 1} \sqrt{N}\mathbb{E}(|\eta^N_n(f_n) - \eta_n(f_n)|^p)^{1/p} \leq b(p)c(n)$$

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 $h_n \in \mathrm{Osc}_1(E'_s)$,

$$\sup_{N\geq 1} \sqrt{N}\mathbb{E}(|\mathbb{P}^N_{\widehat{\eta}^N_n,(t_n,s)}(h_n) - \mathbb{P}_{\widehat{\eta}_n,(t_n,s)}(h_n)|^p)^{1/p} \leq b(p)c(n).$$

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

$$\sup_{N\geq 1} \sqrt{N}\mathbb{E}(|\mathcal{H}^{(n)}_{t_n,s}(\mathbb{P}^N_{\widehat{\eta}^N_n,(t_n,s)}) - \mathcal{H}^{(n)}_{t_n,s}(\mathbb{P}_{\widehat{\eta}_n,(t_n,s)})|^p)^{1/p} \leq b(p)c(n)\delta(H^{(n)}_{t_n,s}).$$

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. □

4.4. An online adaptive SMC algorithm

The above proposition shows that the functional criteria $\mathcal{H}^{(n)}_{t_n,s}(\mathbb{P}_{\widehat{\eta}_n,(t_n,s)})$ can be approximated by $\mathcal{H}^{(n)}_{t_n,s}(\mathbb{P}^N_{\widehat{\eta}^N_n,(t_n,s)})$, 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. *The particle systems $\mathcal{Y}^{(N)}=(\mathcal{Y}^{(N,i)})$, $\widehat{\mathcal{Y}}^{(N)}=(\widehat{\mathcal{Y}}^{(N,i)})$, $Y^{(N,i)}_{s,t}$ and $\widehat{Y}^{(N,i)}_{s,t}$ are defined as $\mathcal{X}^{(N)}=(\mathcal{X}^{(N,i)})$, $\widehat{\mathcal{X}}^{(N)}=(\widehat{\mathcal{X}}^{(N,i)})$, and $X^{(N,i)}_{s,t}$ and $\widehat{X}^{(N,i)}_{s,t}$ by replacing in the inductive construction of the deterministic sequence $(t_n)_{n\geq 0}$ the measures $\mathbb{P}_{\widehat{\eta}_n,(t_n,s)}$ by their current N-particle approximation measures $\overline{\mathbb{P}}^N_{\widehat{\eta}^N_n,(t^N_n,s)}(\cdot) := \frac{1}{N}\sum_{i=1}^N \delta_{(\widehat{\mathcal{Y}}^{(N,i)}_n, Y^{(N,i)}_{t^N_n+1:s})}(\cdot)$. Here $\widehat{\eta}^N_n(\cdot)=\frac{1}{N}\sum_{i=1}^N \delta_{\widehat{\mathcal{Y}}^{(N,i)}_n}(\cdot)$ denotes the updated occupation measure of the particle system $\widehat{\mathcal{Y}}^{(N)}_n$. We also assume that both models are constructed in such a way that they coincide on every time interval $0\leq n\leq m$, once the random times $t^N_n=t_n$, for every $0\leq n\leq m$.*

It is emphasized that the measures $\overline{\mathbb{P}}^N_{\widehat{\eta}^N_n,(t^N_n,s)}$ differ from the reference empirical measures $\mathbb{P}^N_{\widehat{\eta}^N_n,(t_n,s)}$ in (4.5). Indeed, the reference measures $\mathbb{P}^N_{\widehat{\eta}^N_n,(t_n,s)}$ are built using the deterministic times t_n based on the functional criteria $\mathcal{H}^{(n-1)}_{t_{n-1},s}(\mathbb{P}_{\widehat{\eta}_{n-1},(t_{n-1},s)})$, whilst the empirical measures $\overline{\mathbb{P}}^N_{\widehat{\eta}^N_n,(t^N_n,s)}$ are inductively constructed using random times t^N_n based on $\mathcal{H}^{(n-1)}_{t^N_{n-1},s}(\overline{\mathbb{P}}^N_{\widehat{\eta}^N_{n-1},(t^N_{n-1},s)})$.

By construction, for the pair of functional criteria discussed in Section 3.6, we have that $\mathcal{H}^{(n)}_{t^N_n,s}(\overline{\mathbb{P}}^N_{\widehat{\eta}^N_n,(t^N_n,s)})=C^N_{t^N_n,s}$, where $C^N_{t^N_n,s}$ are the empirical criteria discussed in Section 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 $\Omega^N_m(\delta,(a_n)_{0\leq n\leq m})$, the collection of events defined by:*

$$\Omega^N_m(\delta,(a_n)_{0\leq n\leq m}) := \{\forall 0\leq n\leq m, \forall t_n\leq s\leq t_{n+1}$$
$$|\mathcal{H}^{(n)}_{t_n,s}(\mathbb{P}^N_{\widehat{\eta}^N_n,(t_n,s)})-\mathcal{H}^{(n)}_{t_n,s}(\mathbb{P}_{\widehat{\eta}_n,(t_n,s)})|\leq \delta|\mathcal{H}^{(n)}_{t_n,s}(\mathbb{P}_{\widehat{\eta}_n,(t_n,s)})-a_n|\}.$$

The proof of the following result is straightforward and hence omitted.

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

$$\mathcal{H}^{(n)}_{t_n,s}(\mathbb{P}_{\widehat{\eta}_n,(t_n,s)})>a_n \quad\Longrightarrow\quad \mathcal{H}^{(n)}_{t_n,s}(\mathbb{P}^N_{\widehat{\eta}^N_n,(t_n,s)})>a_n.$$

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

$$\bigcap_{0\le n\le m}\{t_n^N=t_n\}\supset\Omega_m^N(\delta,(a_n)_{0\le n\le m}).$$

Proof. This result is proved by induction on $m\ge 0$. 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, we have

$$\forall s<(t_1^N\wedge t_1)\qquad \mathbb{P}^N_{\widehat{\eta}_0^N,(t_0,s)}=\overline{\mathbb{P}}^N_{\widehat{\eta}_0^N,(t_0^N,s)}.$$

By Lemma 5.2, on the event $\Omega_m^N(\delta,(a_n)_{0\le n\le m})$ 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 $\Omega_m^N(\delta,(a_n)_{0\le n\le m})$ it is the case that $t_n^N=t_n$, for any $0\le n\le m$. By our coupling construction, the pair of particle models coincide up to $(t_{m+1}^N\wedge t_{m+1})$; that is,

$$t_m^N=t_m\quad\text{and}\quad\forall s<(t_{m+1}^N\wedge t_{m+1})\qquad \mathbb{P}^N_{\widehat{\eta}_m^N,(t_m,s)}=\overline{\mathbb{P}}^N_{\widehat{\eta}_m^N,(t_m^N,s)}.$$

Once again, by Lemma 5.2, on the event $\Omega_{m+1}^N(\delta,(a_n)_{0\le n\le m+1})$ it also follows that $t_{m+1}^N=t_{m+1}$. □

5.2. Randomized criteria

The situation where the threshold parameters coincide with the adaptive criteria values $\mathcal{H}^{(n)}_{t_n,s}(\mathbb{P}_{\widehat{\eta}_n,(t_n,s)})=a_n$ cannot be dealt with using our analysis. This situation is more involved since it requires us to control both the empirical approximating criteria and the particle approximation. It should be noted, however, that this is not a difficulty in many applications where the probability of this event is zero. Nonetheless, to avoid this technical problem, one natural strategy is to introduce randomized criteria thresholds. We further assume that the parameters $(a_n)_{n\ge 0}$ are sampled realizations of a collection of absolutely continuous random variables $(A_n)_{n\ge 0}$. The main simplification of these randomized criteria comes from the fact that the parameters $\varepsilon_m:=\inf_{0\le n\le m}\inf_{t_n\le s\le t_{n+1}}|\mathcal{H}^{(n)}_{t_n,s}(\mathbb{P}_{\widehat{\eta}_n,(t_n,s)})-a_n|$ are strictly positive for almost every realization $A_n=a_n$ 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:*

$$\mathbb{P}(\exists 0\le n\le m\ t_n^N\neq t_n|(A_n)_{0\le n\le m})\le c_1(m)\left(1+\delta\varepsilon_m\sqrt{\frac{N}{2}}\right)\exp\left(-N\delta^2\varepsilon_m^2/c_2(m)\right)$$

for some constants $c_1(m), c_2(m) < \infty$. *In addition, when the measures* $H^{(n)}_{t_n,s}$ *have a finite support, for any* $\delta \in (0, 1/(2\varepsilon_m))$,

$$\mathbb{P}(\exists 0 \leq n \leq m t^N_n \neq t_n | (A_n)_{0\leq n\leq m}) \leq c_1(m) \exp\left(-N\delta^2\varepsilon^2_m / c_2(m)\right)$$

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

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

for a pair of finite constants $c_1(m), c_2(m) < \infty$. The final line is a direct consequence of Proposition 4.8 and an application of Proposition 5.3 completes the proof. □

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

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

6. A functional central limit theorem

6.1. A direct approach

In this section some direct consequences of the exponential coupling estimates 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 $f_n \in \mathrm{Osc}_1(E'_{t_n})$ the following decomposition holds (writing $\overline{\eta}^N_n$ for the online adaptive approximation introduced in Definition 4.9):

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

with

$$\mathbb{E}(\sqrt{N}[\overline{\eta}^N_n - \eta^N_n](f_n) 1_{\Omega - \Omega^N_m(\delta, (a_n)_{0\leq n\leq m})}) \leq \underbrace{\sqrt{N}\mathbb{P}(\Omega - \Omega^N_m(\delta, (a_n)_{0\leq n\leq m}))}_{\stackrel{N\uparrow\infty}{\longrightarrow} 0}.$$

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

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

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

6.2. Functional central limit theorems

To demonstrate the impact of this functional fluctuation result we provide a brief discussion on the proof of the multivariate central limit theorem. We first recall the functional fluctuation theorem of the local errors associated with the mean field particle approximation introduced in (4.9). This result was initially presented in [5] and extended 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 $1 \leq p \leq n$,*

$$\mathbb{E}(V_p(f_p)V_p(g_p)) = \eta_{p-1}\mathcal{K}_{p,\eta_{p-1}}([f_p - \mathcal{K}_{p,\eta_{p-1}}(f_p)][g_p - \mathcal{K}_{p,\eta_{p-1}}(g_p)]). \tag{6.1}$$

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

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

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

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

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, $D_{p,n} := D_{p+1}\cdots D_{n-1}D_n$. For $p = n$, we use the convention $D_{n,n} = Id$, the identity operator.*

The semigroup $D_{p,n}$ can be explicitly described in terms of the semigroup $\mathcal{Q}_{p,n}$ via

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

The next lemma provides a first-order decomposition of the random fields W_n^N in terms of the local fluctuation errors. Its proof is in the Appendix. Note that R_p can be understood in the proof.

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

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

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 $\mathcal{R}_n^N$ in (6.2) converge in law, in the sense of finite distributions, to the null random field as $N \uparrow \infty$. Therefore the fluctuations of W_n^N follow 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 $N \uparrow \infty$, to a sequence of Gaussian and centered random fields $(W_n)_{n\geq 0}$, where $\forall n \geq 0$ $W_n = \sum_{p=0}^n V_p D_{p,n}$.*

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 $(t_n^N, t_{n+1}^N) = (t_n, t_{n+1})$ the N-particle measures $\overline{\eta}_{n+1}^N = \frac{1}{N}\sum_{i=1}^N \delta_{(\widehat{\mathcal{Y}}_n^{(N,i)},(Y^{(N,i)}_{t_n^N+1}, Y^{(N,i)}_{t_n^N+2},\ldots,Y^{(N,i)}_{t^N_{n+1}}))}$ can be used to approximate the flow of updated Feynman–Kac path distributions $(\widehat{\eta}_{n+1,s})_{t_n\leq s\leq t_{n+1}}$ given for any bounded test function $f_{n+1} \in \mathcal{B}_b(S_{n+1})$ by

$$s \in [t_n, t_{n+1}] \mapsto \widehat{\eta}_{n+1,s}(f_{n+1}) \propto \mathbb{E}[f_{n+1}(X_{0:t_{n+1}})W_{0:s}(X_{1:s})].$$

Indeed, if we choose

$$T_{n+1}^{(1)}(f_{n+1})(x_{0:t_{n+1}}) := f_{n+1}(x_{0:t_{n+1}})W_{t_n:s}(x_{t_n+1:s}),$$

then in some sense

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

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, if the marginal type functions are chosen such that

$$T_{n+1}^{(0)}(f_{n+1})(x_{0:t_{n+2}}) := f_{n+1}(x_{0:t_{n+1}})$$

so

$$\eta_{n+2}(T_{n+1}^{(0)}(f_{n+1})) = \widehat{\eta}_{n+1}(f_{n+1}) \propto \mathbb{E}[f_{n+1}(X_{0:t_{n+1}})W_{0:t_{n+1}}(X_{1:t_{n+1}})],$$

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

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 (CLT) is easily derived for the collection of random fields

$$\widehat{W}^{N,(0)}_{n+1}(f_{n+1}) := \sqrt{N}[\overline{\eta}^N_{n+2}(T^{(0)}_{n+1}(f_{n+1})) - \eta_{n+2}(T^{(0)}_{n+1}(f_{n+1}))],$$
$$\widehat{W}^{N,(1)}_{n+1,s}(f_{n+1}) := \sqrt{N}[\widehat{\eta}^N_{n+1,s}(f_{n+1}) - \widehat{\eta}_{n+1,s}(f_{n+1})]$$

as well as for the mixture of random field sequences

$$\widehat{W}^N_{n+1,s} = 1_{t^N_n \leq s < t^N_{n+1}} \widehat{W}^{N,(1)}_{n+1,s} + 1_{s=t^N_{n+1}} \widehat{W}^{N,(0)}_{n+1}. \tag{6.3}$$

The fluctuation analysis of these random fields relies on the functional CLT stated in Corollary 6.4. In particular, the fluctuations of the random fields (6.3) depend on those of a pair of random fields.

6.4. Related work

Reference [6] is the only published paper discussing a convergence result for an adaptive SMC scheme. The authors establish a CLT using an inductive proof w.r.t. deterministic time periods. They avoid the degenerate situation where the threshold parameter coincides with the limiting functional criterion. More recently, this problem has also been addressed in [2], Chapter 4. However, the author does not account for the randomness of the resampling times in his analysis.

Appendix

Proof of Lemma 4.4. Via (3.7), for any $f \in \mathcal{B}_b(S_{n+1})$ we find that

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

where $\mathcal{G}_{p,n,\eta} := \mathcal{Q}_{p,n}(1)/\eta(\mathcal{Q}_{p,n}(1))$ and $\mathcal{P}_{p,n}(f) = \mathcal{Q}_{p,n}(f)/\mathcal{Q}_{p,n}(1)$. Now, since $\eta(\mathcal{G}_{p,n,\eta}) = 1$, it follows that

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

Using the fact that

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

we find

$$\forall f \in \mathrm{Osc}_1(S_n) \qquad \|\mathcal{D}_{p,n,\eta}(f)\| \le q_{p,n}\beta(\mathcal{P}_{p,n}).$$

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

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

with $\overline{\mathcal{G}}_{p,n,\eta} := \mathcal{G}_{p,n,\eta}/2q_{p,n}$ and $\overline{\mathcal{D}}_{p,n,\eta}(f) := D_{p,n,\eta}(f)/2q_{p,n}\beta(\mathcal{P}_{p,n}) \in \mathrm{Osc}_1(S_p)$. □

Proof of Lemma 6.3. The lemma is proved by induction on n. For $n=0$, it follows that $W_n^N = V_0^N = \sqrt{N}[\eta_0^N - \Phi_0(\eta_{-1}^N)]$, with $\Phi_0(\eta_{-1}^N) = \eta_0$. Assuming the formula at n

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

Letting $D_{n+1,n+1} = I$, it follows that (6.2) is satisfied at rank $(n+1)$ due to

$$V_{n+1}^N + \sum_{p=0}^{n} V_p^N D_{p,n+1} = \sum_{p=0}^{n+1} V_p^N D_{p,n+1},$$

$$\sum_{p=0}^{n-1} R_{p+1}(\eta_p^N, \eta_p) D_{p+1,n+1} + R_{n+1}(\eta_n^N, \eta_n) = \sum_{p=0}^{n} R_{p+1}(\eta_p^N, \eta_p) D_{p+1,n+1}.$$

□

Acknowledgements

We would like to thank the Associate Editor and the three referees for many helpful comments that vastly improved the paper.

References

[1] Cappé, O., Moulines, E. and Rydén, T. (2005). *Inference in Hidden Markov Models*. New York: Springer. MR2159833

[2] Cornebise, J. (2009). Adaptive sequential Monte Carlo methods. Ph.D. thesis, Dept. Applied Mathematics, Univ. Paris 6.

[3] Cornebise, J., Moulines, É. and Olsson, J. (2008). Adaptive methods for sequential importance sampling with application to state space models. *Stat. Comput.* **18** 461–480. MR2461889

[4] Del Moral, P. (2004). *Feynman–Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. New York: Springer. MR2044973

[5] Del Moral, P. and Miclo, L. (2000). Branching and interacting particle systems approximations of Feynman–Kac formulae with applications to non-linear filtering. In *Séminaire de Probabilités* **XXXIV**. *Lecture Notes in Math.* **1729** 1–145. Berlin: Springer. MR1768060

[6] Douc, R. and Moulines, E. (2008). Limit theorems for weighted samples with applications to sequential Monte Carlo methods. *Ann. Statist.* **36** 2344–2376. MR2458190

[7] Doucet, A., de Freitas, N. and Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice*. New York: Springer. MR1847783

[8] Liu, J.S. (2001). *Monte Carlo Strategies in Scientific Computing*. New York: Springer. MR1842342

[9] Liu, J.S. and Chen, R. (1995). Blind deconvolution via sequential imputation. *J. Amer. Statist. Assoc.* **90** 567–576.

Received May 2009 and revised September 2010