Bernoulli **23**(1), 2017, 670–709
DOI: 10.3150/14-BEJ680

Nonasymptotic analysis of adaptive and annealed Feynman–Kac particle models

FRANÇOIS GIRAUD[1,2,*] and PIERRE DEL MORAL[3,**]

[1]*CEA-CESTA, 33114 Le Barp, France.*
[2]*INRIA Bordeaux Sud-Ouest, team ALEA, Domaine Universitaire, 351, cours de la Libération, 33405 Talence Cedex, France. E-mail: [*]francois.giraud@ens-cachan.org*
[3]*School of Maths and Stats. University of New South Wales, Sydney, Australia.*
E-mail: [**] pierre.del-moral@inria.fr

Sequential and quantum Monte Carlo methods, as well as genetic type search algorithms can be interpreted as a mean field and interacting particle approximations of Feynman–Kac models in distribution spaces. The performance of these population Monte Carlo algorithms is strongly related to the stability properties of nonlinear Feynman–Kac semigroups. In this paper, we analyze these models in terms of Dobrushin ergodic coefficients of the reference Markov transitions and the oscillations of the potential functions. Sufficient conditions for uniform concentration inequalities w.r.t. time are expressed explicitly in terms of these two quantities. We provide an original perturbation analysis that applies to annealed and adaptive Feynman–Kac models, yielding what seems to be the first results of this kind for these types of models. Special attention is devoted to the particular case of Boltzmann–Gibbs measures' sampling. In this context, we design an explicit way of tuning the number of Markov chain Monte Carlo iterations with temperature schedule. We also design an alternative interacting particle method based on an adaptive strategy to define the temperature increments. The theoretical analysis of the performance of this adaptive model is much more involved as both the potential functions and the reference Markov transitions now depend on the random evolution on the particle model. The nonasymptotic analysis of these complex adaptive models is an open research problem. We initiate this study with the concentration analysis of a simplified adaptive models based on reference Markov transitions that coincide with the limiting quantities, as the number of particles tends to infinity.

Keywords: adaptive models; Feynman–Kac models; interacting particle systems; sequential Monte Carlo methods

Introduction

Feynman–Kac (*abbreviate FK*) particle methods, also called sequential, quantum or diffusion Monte Carlo methods, are stochastic algorithms to sample from a sequence of complex high-dimensional probability distributions. These stochastic simulation techniques are of current use in numerical physics [1,5,33] to compute ground state energies in molecular systems. They are also used in statistics, signal processing and information sciences [6,15,17,20] to compute posterior distributions of a partially observed signal or unknown parameters. In the evolutionary computing literature, these Monte Carlo methods are used as natural population search algorithms for solving optimization problems. From the pure mathematical viewpoint, these advanced Monte Carlo methods are an interacting particle system (*abbreviate IPS*) interpretation of FK models.

For a more thorough discussion on these models, we refer the reader to the monograph [16], and the references therein. The principle (see also [17] and the references therein) is to approximate a sequence of target probability distributions $(\eta_n)_n$ by a large cloud of random samples termed particles or walkers. The algorithm starts with N independent samples from η_0 and then alternates two types of steps: an acceptance-rejection scheme equipped with a selection type recycling mechanism, and a sequence of free exploration of the state space.

In the recycling stage, the current cloud of particles is transformed by randomly duplicating and eliminating particles in a suitable way, similarly to a selection step in models of population genetics. In the Markov evolution step, particles move independently one each other (mutation step).

This method is often used for solving sequential problems, such as filtering (see, e.g., [6,15, 21,28]). In other interesting problems, these algorithms also turn out to be efficient to sample from a single target measure η. In this context, the central idea is to find a judicious interpolating sequence of measures $(\eta_k)_{0\leq k\leq n}$ with increasing sampling complexity, starting from some initial distribution η_0, up to the terminal one $\eta_n = \eta$. Consecutive measures η_k and η_{k+1} are sufficiently similar to allow for efficient importance sampling and/or acceptance-rejection sampling. The sequential aspect of the approach is then an "artificial way" to introduce the difficulty of sampling gradually. In this vein, important examples are provided by annealed models. More generally, a crucial point is that large population sizes allow to cover several modes simultaneously. This is an advantage compared to standard MCMC methods that are more likely to be trapped in local modes. These sequential samplers have been used with success in several application domains, including rare events simulation (see [9]), stochastic optimization and more generally Boltzmann–Gibbs measures sampling [17].

Up to now, IPS algorithms have been mostly analyzed using asymptotic (i.e., when number of particles N tends to infinity) techniques, notably through fluctuation theorems and large deviation principles (see, e.g., [6,11,14,15,19,20,23,25,35] and [16] for an overview).

Some nonasymptotic theorems have been recently developed [10,18], but unfortunately none of them apply to analyze annealed and adaptive FK particle models. On the other hand, these type of nonhomogeneous IPS algorithms are of current use for solving concrete problems arising in numerical physics and engineering sciences (see, e.g., [4,13,27,30,34,37–39]). By the lack of nonasymptotic estimates, these particle algorithms are used as natural heuristics.

The main contribution of this article is to analyze these two classes of time nonhomogeneous IPS models. Our approach is based on semigroup techniques and on an original perturbation analysis to derive several uniform estimates w.r.t. the time parameter.

More precisely, in the case of annealed type models, we estimate explicitly the stability properties of FK semigroup in terms of the Dobrushin ergodic coefficient of the reference Markov chain and the oscillations of the potential functions. We combine these techniques with nonasymptotic theorems on L^p-mean error bounds [25] and some useful concentration inequalities [22,26]. Then we provide parameter tuning strategies that allow to deduce some useful uniform concentration inequalities w.r.t. the time parameter. These results apply to nonhomogeneous FK models associated with cooling temperature parameters. In this situation, the sequence of measures η_n is associated with a nonincreasing temperature parameter. We mention that other independent approaches, such as Whiteley's [41] or Schweizer's [40], are based on, for example, drift conditions, hyper-boundedness, spectral gaps, or nonasymptotic biais and variance decompositions.

These approaches lead to convergence results that may also apply to noncompact state spaces. To our knowledge, these techniques are restricted to nonasymptotic variance theorems and they cannot be used to derive uniform and exponential concentration inequalities. It seems also difficult to extend these approaches to analyze the adaptive IPS model discussed in the present article. To solve these questions, we develop a perturbation technique of stochastic FK semigroups. In contrast to traditional FK semigroup, the adaptive particle scheme is now based on random potential functions that depend on a cooling schedule adapted to the variability and the adaptation of the random populations.

The rest of the article is organized as follows. In a preliminary section, we recall a few essential notions related to Dobrushin coefficients or FK semigroups. We also provide some important nonasymptotic results we use in the further development of the article. Section 2 is concerned with the semigroup stability analysis of these models. We also provide a couple of uniform L^p-deviations and concentration estimates. In Section 3, we apply these results to Boltzmann–Gibbs models associated with a decreasing temperature schedule. In this context, IPS algorithm can be interpreted as a sequence of interacting simulated annealing algorithms (*abbreviate ISA*). We design an explicit way of tuning the number of Markov chain Monte Carlo iterations with the temperature schedule. Finally, in Section 4, we propose an alternative ISA method based on an original adaptive strategy to design on the flow the temperature decrements. We provide a nonasymptotic study for a simplified model, based on a perturbation analysis. We end the article with L^p-deviation estimates as well as a couple of concentration inequalities.

Statement of some results

Feynman–Kac particle algorithms consist in evolving a particle system $\zeta_n = (\zeta_n^1, \ldots, \zeta_n^N)$ of size N, on a given state space E. Their interacting evolution is decomposed into two genetic type transitions: a selection step, associated with some positive potential function G_n; and a mutation step, where the selected particles evolve randomly according to a given Markov transition M_n (a more detailed description of these IPS algorithms is provided in Section 1.4). In this context, the occupation measures $\eta_n^N := \frac{1}{N}\sum_{1\le i\le N}\delta_{\zeta_n^i}$ are N-approximations of a sequence of measures η_n defined by the FK recursive formulae:

$$\eta_n(f) = \frac{\eta_{n-1}(G_n \times M_n.f)}{\eta_{n-1}(G_n)},$$

for all bounded measurable function f on E (a more detailed discussion on these evolution equations is provided in Section 1.3.1).

To describe with some precision the main results of the article, we consider the pair of parameters (g_n, b_n) defined below.

$$g_n := \sup_{x,y\in E}\frac{G_n(x)}{G_n(y)} \quad \text{and} \quad b_n = \beta(M_n) := \sup_{\substack{x,y\in E\\ A\subset E}}\big|M_n(x,A) - M_n(y,A)\big|.$$

The quantity $\beta(M_n)$ is called the Dobrushin ergodic coefficient of the Markov transition M_n. One of our first main results can be basically stated as follows.

Theorem 1. *We assume that*

$$\sup_{p\geq 1} g_p \leq M \quad \text{and} \quad \sup_{p\geq 1} b_p \leq \frac{a}{M(1+a)}$$

for some finite constant $M < \infty$ and some $a \in (0,1)$. In this situation, for any $n \geq 0$, $N \geq 1$, $y \geq 0$ and $f \in \mathcal{O}_1(E)$, the probability of the event

$$\eta_n^N(f) - \eta_n(f) \leq \frac{r_1^\star(1+h_0(y)) + r_2^\star\sqrt{Ny}}{N}$$

is greater than $1 - \mathrm{e}^{-y}$, where $r_1^\star$ and $r_2^\star$ are some constants that are explicitly defined in terms of (a, M), and $h_0(y) = 2y + 2\sqrt{y}$.

In Section 2.2, under the same assumptions of Theorem 1, we also prove uniform L^p-mean error bounds as well as new concentration inequalities for unnormalized particle models. We also extend the analysis to the situation where $g_n \underset{n\to+\infty}{\longrightarrow} 1$.

We already mentioned that the regularity conditions on b_n may appear difficult to check since the Markov kernels are often dictated by the application under study. However, we can deal with this problem as soon as we can simulate a Markov kernel K_n such that $\eta_n.K_n = \eta_n$. Indeed, to stabilize the system, the designer can "add" several MCMC evolution steps next to each M_n-mutation step. From a more formal viewpoint, the target sequence $(\eta_n)_n$ is clearly also solution of the FK measure-valued equations associated with the Markov kernels $M_n' = M_n.K_n^{m_n}$, where iteration numbers m_n are to be chosen loosely. This system is more stable since the corresponding b_n' satisfy

$$b_n' = \beta\big(M_n'\big) \leq b_n.\beta\big(K_n^{m_n}\big) \leq b_n.\beta(K_n)^{m_n}.$$

In such cases, Theorem 1 and its extension provide sufficient conditions on the iteration numbers m_n to ensure the convergence and the stability properties of the algorithm.

These results apply to stochastic optimization problems. Let $V : E \to \mathbb{R}$ be a bounded potential function, β_n a sequence which tends to infinity, and m a reference measure on E. It is well known that the sequence of Boltzmann–Gibbs measures

$$\eta_n(\mathrm{d}x) \propto \mathrm{e}^{-\beta_n.V(x)} m(\mathrm{d}x)$$

concentrates on V's global minima (in the sense of m-essinf(V)). In the above display, $\propto$ stands for the proportional sign. One central observation is that these measures can be interpreted as a FK flow of measures associated with potential functions $G_n = \mathrm{e}^{-(\beta_n-\beta_{n-1}).V}$ and Markov kernels $M_n = K_{\beta_n}^{m_n k_0}$ where K_{β_n} is a simulating annealing kernel (see Section 3.2) and m_n and k_0 are given iteration parameters. In the further development of this section, we let K be the proposal transition of the simulated annealing transition K_β. In this context, the IPS methods can be used to minimize V. The conditions on b_n and g_n can be turned into conditions on the temperature schedule β_n and the number of MCMC iterations m_n. Moreover, combining our results with standard concentration properties of Boltzmann–Gibbs measures, we derive some convergence results in terms of optimization performance. In this notation, our second main result is basically stated as follows.

Theorem 2. *Let us fix $a \in (0,1)$. We assume that for any $x \in E$, $K^{k_0}(x,\cdot) \geq \delta\nu(\cdot)$ for some measure ν on E, some $\delta > 0$ and some $k_0 \geq 1$. We also assume that the temperature increments $\Delta_p := \beta_p - \beta_{p-1}$ and the iteration numbers m_p satisfy the following conditions*:

$$\sup_{p\geq 1} \Delta_p \leq \Delta \quad \text{and} \quad m_p \geq \frac{\log(\mathrm{e}^{\Delta.\,\mathrm{osc}(V)}(1+a)/a)\mathrm{e}^{\mathrm{osc}(V).\beta_p}}{\delta}$$

for some constant Δ. For all $\varepsilon > 0$, let $p_n^N(\varepsilon)$ be the proportion of particles (ζ_n^i) so that $V(\zeta_n^i) \geq V_{\min} + \varepsilon$. Then, for any $n \geq 0$, $N \geq 1$, $y \geq 0$ and for all $\varepsilon' < \varepsilon$, the probability of the event

$$p_n^N(\varepsilon) \leq \frac{\mathrm{e}^{-\beta_n(\varepsilon-\varepsilon')}}{m_{\varepsilon'}} + \frac{r_1^\star(1+h_0(y)) + r_2^\star\sqrt{Ny}}{N}$$

is greater than $1-\mathrm{e}^{-y}$, with $m_{\varepsilon'} = m(V \leq V_{\min}+\varepsilon')$, $h_0(y) = 2y+2\sqrt{y}$, and the same constants $(r_1^\star, r_2^\star)$ as the ones stated in Theorem 1 (with $M = \mathrm{e}^{\Delta.\,\mathrm{osc}(V)}$).

It is instructive to compare the estimates in the above theorem with the performance analysis of the traditional simulated annealing model (*abbreviate SA*). Firstly, most of the literature on SA models is concerned with the weak convergence of the law of the random states of the algorithm. When the initial temperature of the scheme is greater than some critical value, using a logarithmic cooling schedule, it is well known that the probability for the random state to be in the global extrema levels tends to 1, as the time parameter tends to ∞. The cooling schedule presented in Theorem 2 is again a logarithmic one. In contrast to the SA model (see, e.g., Theorem 1 in [32], and Theorem 4.3.16 in [3]), Theorem 2 allows to quantify the performance analysis of the ISA model in terms of uniform concentration inequalities that does not depend on a critical parameter.

In practice, choosing the sequence of increments $\Delta_n = (\beta_n - \beta_{n-1})$ in advance can cause computational problems. To solve this problem, adaptive strategies, where increment Δ_n depends on the current set of particles ζ_{n-1}, are of common use in the engineering community (see, e.g., [13,27,34,37,39]). In this context, we propose to study the case where the increment Δ_n^N is chosen so that

$$\eta_{n-1}^N\left(\mathrm{e}^{-\Delta_n^N \cdot V}\right) = \varepsilon,$$

where $\varepsilon > 0$ is a given constant (see Section 4.2 for a detailed description of the algorithm). Computationally speaking, ε is the expectation of the proportion of particles which are not concerned with the recycling mechanism in the selection step. We interpret this particle process as a perturbation of a theoretical FK sequence η_n associated with a theoretical temperature schedule β_n. As we mentioned in the abstract and the Introduction, the theoretical analysis of this class of adaptive particle model is much more involved as the potential functions and the mutations transitions now depend on the evolution of the particle populations. We consider a simplified adaptive model associated with the limiting reference Markov transitions. A precise description of these adaptive particle algorithms, and their reduced simplified versions are provided in Sections 4.1, 4.2 and 4.3.

Our main result is the following L^p-mean error estimate, where (η_n^N) denotes the sequence of empirical measures associated with the particle system described in (4.3) page 698.

Theorem 3. *For any $p \geq 1$, $n \geq 0$, $N \geq 1$ and any bounded by 1 function f, we have*

$$\mathbb{E}\big(\big|\eta_n^N(f) - \eta_n(f)\big|^p\big)^{1/p} \leq \frac{B_p}{\sqrt{N}} \sum_{k=0}^{n} \prod_{i=k+1}^{n} \big(b_i g_i (1 + c_i)\big),$$

with $c_n = \frac{V_{\max} e^{\Delta_n V_{\max}}}{\varepsilon \cdot \eta_{n-1}(V)}$, $\Delta_n = \beta_n - \beta_{n-1}$ and B_p defined below.

$$B_{2p}^{2p} = \frac{(2p)!}{2^p.p!}; \qquad B_{2p+1}^{2p+1} = \frac{(2p+1)!}{2^p.p!\sqrt{2p+1}}. \tag{0.1}$$

Under appropriate regularity conditions on the parameters b_n, g_n, c_n, we mention that these L^p-mean error bounds also provide uniform concentration inequalities.

The proofs of Theorems 1, 2, 3 and related uniform exponential estimates are detailed, respectively, in Sections 2.2, 3.2 and 4.4.

1. Some preliminaries

1.1. Basic notation

Let (E, r) be a complete, separable metric space and let $\mathcal{E}$ be the σ-algebra of Borel subsets of E. Denote by $\mathcal{P}(E)$ the space of probability measures on E. Let $\mathcal{B}(E)$ be the space of bounded, measurable, real-valued functions on E.

If $\mu \in \mathcal{P}(E)$, $f \in \mathcal{B}(E)$ and K, K_1, K_2 are Markov kernels on E, then $\mu(f)$ denotes the quantity $\int_E f(x)\mu(\mathrm{d}x)$, $K_1.K_2$ denotes the Markov kernel defined by

$$(K_1.K_2)(x, A) = \int_E K_1(x, \mathrm{d}y) K_2(y, A),$$

$K.f$ denotes the function defined by

$$K.f(x) = \int_E K(x, \mathrm{d}y) f(y)$$

and $\mu.K$ denotes the probability measure defined by

$$\mu.K(A) = \int_E K(x, A)\mu(\mathrm{d}x).$$

If G is a positive, bounded function on E, then $\psi_G : \mathcal{P}(E) \to \mathcal{P}(E)$ denotes the Boltzmann–Gibbs transformation associated with G, defined by

$$\forall \mu \in \mathcal{P}(E), \forall f \in \mathcal{B}(E) \qquad \psi_G(\mu)(f) = \frac{\mu(G \times f)}{\mu(G)}.$$

For any $f \in \mathcal{B}(E)$, let $\|f\|_\infty = \sup_{x\in E} |f(x)|$ and $\mathrm{osc}(f) = (f_{\max} - f_{\min})$. Let $\mathcal{B}_1(E) \subset \mathcal{B}(E)$ be the subset of functions f so that $\|f\|_\infty \leq 1$, and $\mathcal{O}_1(E) \subset \mathcal{B}(E)$ be the subset of functions f so that $\mathrm{osc}(f) \leq 1$. For any random variable $X : \Omega \to \mathbb{R}$ defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and any $p \geq 1$, $\|X\|_p$ stands for the L^p norm $\mathbb{E}(|X|^p)^{1/p}$. Let $\mathcal{P}_\Omega(E)$ be the set of random probability measures on E. For all $p \geq 1$, we denote by d_p the distance on $\mathcal{P}_\Omega(E)$ defined for all random measures $\hat{\mu}, \hat{\nu}$ by

$$d_p(\hat{\mu}, \hat{\nu}) = \sup_{f\in\mathcal{O}_1(E)} \left\|\hat{\mu}(f) - \hat{\nu}(f)\right\|_p.$$

Finally, for any $x \in E$, δ_x stands for the Dirac measure centered on x.

1.2. Dobrushin Ergodic coefficient

Let us recall here the definitions as well as some simple properties that will be useful in the following.

Definition 4. *Let $\mu, \nu \in \mathcal{P}(E)$. The total variation distance between μ and ν is defined by*

$$\|\mu - \nu\|_{\mathrm{tv}} = \sup\left\{\left|\mu(A) - \nu(A)\right|; A \in \mathcal{E}\right\},$$

or in an equivalent way

$$\|\mu - \nu\|_{\mathrm{tv}} = \sup\left\{\left|\mu(f) - \nu(f)\right|; f \in \mathcal{B}_1(E), f \geq 0\right\}.$$

Definition 5. *To each Markov kernel K on E, is associated its Dobrushin ergodic coefficient $\beta(K) \in [0, 1]$ defined by*

$$\beta(K) = \sup\left\{K(x, A) - K(y, A); x, y \in E, A \in \mathcal{E}\right\},$$

or in an equivalent way

$$\beta(K) = \sup\left\{\frac{\|\mu.K - \nu.K\|_{\mathrm{tv}}}{\|\mu - \nu\|_{\mathrm{tv}}}; \mu, \nu \in \mathcal{P}(E), \mu \neq \nu\right\}.$$

The parameter $\beta(K)$ characterizes mixing properties of the Markov kernel K. Note that function β is an operator norm, in the sense that $\beta(K_1.K_2) \leq \beta(K_1).\beta(K_2)$, for any couple of Markov kernels K_1, K_2. By definition, for any measures $\mu, \nu \in \mathcal{P}(E)$ and any Markov kernel K, we have $\|\mu.K - \nu.K\|_{\mathrm{tv}} \leq \beta(K).\|\mu - \nu\|_{\mathrm{tv}}$. Otherwise, for any function $f \in \mathcal{B}(E)$,

$$\mathrm{osc}(K.f) \leq \beta(K) \cdot \mathrm{osc}(f). \tag{1.1}$$

Further details on these ergodic coefficients can be found in the monograph [16]. Let us give here a lemma that we will need hereinafter (a proof is given in the Appendix, page 705).

Lemma 6. *Let $\mu, \nu \in \mathcal{P}(E)$ and G a positive, bounded function on E satisfying $\sup_{x,y\in E} \frac{G(x)}{G(y)} \leq g$, for some finite constant $g \geq 0$. In this situation, we have*

$$\left\|\Psi_G(\mu) - \Psi_G(\nu)\right\|_{\mathrm{tv}} \leq g.\|\mu - \nu\|_{\mathrm{tv}}.$$

1.3. Feynman–Kac models

We recall here some standard tools related to FK models. They provide useful theoretical background and notation to formalize and analyze IPS methods (see, e.g., [20,22,25] for further details).

1.3.1. *Evolution equations*

Consider a sequence of probability measures $(\eta_n)_n$, defined by an initial measure η_0 and recursive relations

$$\forall f \in \mathcal{B}(E) \qquad \eta_n(f) = \frac{\eta_{n-1}(G_n \times M_n.f)}{\eta_{n-1}(G_n)} \tag{1.2}$$

for positive functions $G_n \in \mathcal{B}(E)$ and Markov kernels M_n with $M_n(x,\cdot) \in \mathcal{P}(E)$ and $M_n(\cdot, A) \in \mathcal{B}_1(E)$. This is the sequence of measures we mainly wish to approximate with the IPS algorithm. In an equivalent way, $(\eta_n)_n$ can be defined by the relation

$$\eta_n = \phi_n(\eta_{n-1}),$$

where $\phi_n : \mathcal{P}(E) \to \mathcal{P}(E)$ is the FK transformation associated with potential function G_n and Markov kernel M_n and defined by

$$\phi_n(\eta_{n-1}) = \psi_{G_n}(\eta_{n-1}).M_n$$

with

$$\psi_{G_n}(\eta_{n-1})(\mathrm{d}x) := \frac{1}{\eta_{n-1}(G_n)} G_n(x)\eta_{n-1}(\mathrm{d}x).$$

The next formula provides an interpretation of the Boltzmann–Gibbs transformation in terms of a nonlinear Markov transport equation

$$\Psi_{G_n}(\eta_{n-1})(\mathrm{d}y) = (\eta_{n-1}S_{n,\eta_{n-1}})(\mathrm{d}y) := \int \eta_{n-1}(\mathrm{d}x)S_{n,\eta_{n-1}}(x,\mathrm{d}y)$$

with the Markov transition S_{n,η_n} defined below

$$S_{n,\eta_{n-1}}(x,\mathrm{d}y) = \varepsilon_n.G_n(x)\delta_x(\mathrm{d}y) + \big(1 - \varepsilon_n.G_n(x)\big)\Psi_{G_n}(\eta_{n-1})(\mathrm{d}y)$$

(for any constant $\varepsilon_n > 0$ so that $\varepsilon_n.G_n \leq 1$). This implies

$$\eta_n = \eta_{n-1}K_{n,\eta_{n-1}} \qquad \text{with } K_{n,\eta_{n-1}} = S_{n,\eta_{n-1}}M_n. \tag{1.3}$$

Therefore, η_n can be interpreted as the distributions of the random states $\overline{X}_n$ of a Markov chain whose Markov transitions

$$\mathbb{P}(\overline{X}_{n+1} \in \mathrm{d}y | \overline{X}_n = x) := K_{n+1,\eta_n}(x, \mathrm{d}y) \tag{1.4}$$

depend on the current distribution $\eta_n = \mathrm{Law}(\overline{X}_n)$.

We finally recall that the measures η_n admit the following functional representations:

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

(1 stands for the unit function) with the unnormalized FK measures γ_n defined by the formulae

$$\gamma_0 = \eta_0; \qquad \gamma_n(f) = \gamma_{n-1}(G_n \times M_n.f). \tag{1.5}$$

Comparing this definition with (1.2), it is clear that the normalizing constant $\gamma_n(1)$ satisfies

$$\gamma_n(1) = \prod_{p=1}^{n} \eta_{p-1}(G_p). \tag{1.6}$$

The special interest given to this quantity will be motivated in Section 3.1.

1.3.2. *Feynman–Kac semigroup*

An important point is that the semigroup transformations

$$\phi_{p,n} := \phi_n \circ \phi_{n-1} \circ \cdots \circ \phi_{p+1}$$

admit a comparable structure as each of the ϕ_k. To be more precise, for each integer p, let us define the unnormalized integral operator Q_p

$$\forall f \in \mathcal{B}(E) \qquad Q_p.f = G_p.M_p.f \tag{1.7}$$

and the composition operators $Q_{p,n}$ defined by the backward recursion

$$Q_{p,n} = Q_{p+1}.(Q_{p+2} \cdots Q_n) = Q_{p+1}.Q_{p+1,n}. \tag{1.8}$$

We use the convention $Q_{n,n} = \mathrm{Id}$ for $p = n$. Comparing these definitions with (1.5), it is clear that $\gamma_n = \gamma_{n-1}.Q_n$ and more generally

$$\gamma_n = \gamma_p.Q_{p,n}.$$

for any $p \leq n$. The semigroup $\phi_{p,n}$ can be expressed in terms of $Q_{p,n}$ with the following formulae:

$$\phi_{p,n}(\mu)(f) = \frac{\mu(Q_{p,n}.f)}{\mu(Q_{p,n}.1)}$$

for any $f \in \mathcal{B}(E)$ and $\mu \in \mathcal{P}(E)$. Finally, if we set

$$P_{p,n}.f = \frac{Q_{p,n}.f}{Q_{p,n}.1} \quad \text{and} \quad G_{p,n} = Q_{p,n}.1$$

then we find that

$$\phi_{p,n}(\mu)(f) = \frac{\mu(G_{p,n}.P_{p,n}.f)}{\mu(G_{p,n})},$$

or in other words: $\phi_{p,n}(\mu) = \psi_{G_{p,n}}(\mu).P_{p,n}$.

1.4. The interacting particle system model

The central idea is to approximate the measures η_n by simulating an interacting particle system $(\zeta_n)_n = (\zeta_n^1, \ldots, \zeta_n^N)_n$ of size N so that

$$\eta_n^N = \frac{1}{N} \sum_{1 \le i \le N} \delta_{\zeta_n^i} \underset{N \uparrow \infty}{\longrightarrow} \eta_n.$$

Of course, the main issue is to make precise and to quantify this convergence. The particle model is defined as follows.

We start with N independent samples $\zeta_0 = (\zeta_0^1, \ldots, \zeta_0^N)$ from η_0. Then the particle dynamics alternates two genetic type transitions.

During the first step, every particle ζ_n^i evolves to a new particle $\widehat{\zeta}_n^i$ randomly chosen with the distribution

$$S_{n+1,\eta_n^N}\big(\zeta_n^i, \mathrm{d}x\big) := \varepsilon_{n+1}.G_{n+1}\big(\zeta_n^i\big)\delta_{\zeta_n^i}(\mathrm{d}x) + \big(1 - \varepsilon_{n+1}.G_{n+1}\big(\zeta_n^i\big)\big)\Psi_{G_{n+1}}\big(\eta_n^N\big)(\mathrm{d}x)$$

with the updated measures

$$\Psi_{G_{n+1}}\big(\eta_n^N\big) = \sum_{j=1}^{N} \frac{G_{n+1}(\zeta_n^j)}{\sum_{k=1}^{N} G_{n+1}(\zeta_n^k)} \delta_{\zeta_n^j}.$$

This transition can be interpreted as an acceptance-rejection scheme with a recycling mechanism. In the second step, the selected particles $\widehat{\zeta}_n^i$ evolve randomly according to the Markov transitions M_{n+1}. In other words, for any $1 \le i \le N$, we sample a random state ζ_{n+1}^i with distribution $M_{n+1}(\widehat{\zeta}_n^i, \mathrm{d}x)$.

In view of (1.4), if we replace η_n^N by η_n, then ζ_n coincide with N independent copies of the Markov chains $\overline{X}_n$ defined in (1.3). On the other hand, by the law of large numbers, we have $\eta_0^N \simeq \eta_0$ so that

$$\eta_1^N \simeq \eta_0^N.K_{1,\eta_0^N} \simeq \eta_0.K_{1,\eta_0} = \eta_1.$$

Iterating this approximation procedure, the empirical measure η_n^N is expected to approximate η_n at any time $n \geq 0$. As for the unnormalized measures γ_n, we define

$$\gamma_n^N(1) = \prod_{p=1}^{n} \eta_{p-1}^N(G_p)$$

(mimicking formula (1.6)) and more generally $\gamma_n^N(f) = \eta_n^N(f) \times \prod_{p=1}^n \eta_{p-1}^N(G_p)$. Let us mention (see, e.g., [15]) that these particle models provide an unbiased estimate of the unnormalized measures; that is we have that

$$\forall f \in \mathcal{B}(E) \qquad \mathbb{E}\big(\gamma_n^N(f)\big) = \gamma_n(f).$$

In addition to the analysis of η_n^N's convergence, the concentration properties of the unbiased estimators $\gamma_n^N(1)$ around their limiting values $\gamma_n(1)$ will also be considered thereafter.

1.5. Some nonasymptotic results

To quantify the FK semigroup stability properties, it is convenient to introduce the following parameters.

Definition 7. *For any integers $p < n$, we set*

$$b_n := \beta(M_n) \quad \textit{and} \quad b_{p,n} := \beta(P_{p,n}),$$
$$g_n := \sup_{x,y \in E} \frac{G_n(x)}{G_n(y)} \quad \textit{and} \quad g_{p,n} := \sup_{x,y \in E} \frac{G_{p,n}(x)}{G_{p,n}(y)}.$$

The quantities $g_{p,n}$, and, respectively, $b_{p,n}$, reflect the oscillations of the potential functions $G_{p,n}$, and, respectively, the mixing properties of the Markov transition $P_{p,n}$ associated with the FK semigroup $\phi_{p,n}$ described in Section 1.3.2. Several contraction inequalities of $\phi_{p,n}$ w.r.t. the total variation norm or different types of relative entropies can be derived in terms of these two quantities (see, for instance, [16]).

Let

$$\begin{aligned} h_0 &:= x \mapsto 2(x + \sqrt{x}) \quad \text{and} \\ h_1 &:= x \mapsto \frac{x}{3} + \sqrt{2x}. \end{aligned} \tag{1.9}$$

The performance analysis developed in Sections 2.1 and 3.2 is partly based on the three nonasymptotic inequalities presented below.

Firstly, the following L^p-mean error bound for all $f \in \mathcal{B}_1(E)$ is inspired by [25] (see Section 2.2.2):

$$\mathbb{E}\big(\big|\eta_n^N(f) - \eta_n(f)\big|^p\big)^{1/p} \leq \frac{B_p}{\sqrt{N}} \sum_{k=0}^{n} g_{k,n} b_{k,n}, \tag{1.10}$$

where B_p are the constants introduced in (0.1), page 675. A detailed proof of this estimate is given in the Appendix, page 706.

Secondly, the following concentration inequality is derived in [26] (see Theorem 1.2 and Section A3). For all $f \in \mathcal{O}_1(E)$ and any $y \geq 0$, we have

$$\mathbb{P}\left(\eta_n^N(f) - \eta_n(f) \geq \frac{r_n}{N}\big(1 + h_0(y)\big) + \overline{\beta}_n\sqrt{\frac{2y}{N}}\right) \leq \mathrm{e}^{-y}, \tag{1.11}$$

where r_n and $\overline{\beta}_n$ are constants so that

$$\begin{cases} r_n \leq 4\displaystyle\sum_{p=0}^{n} g_{p,n}^3 b_{p,n}, \\ \overline{\beta}_n^2 \leq 4\displaystyle\sum_{p=0}^{n} g_{p,n}^2 b_{p,n}^2. \end{cases}$$

To be more precise with the derivation of these estimates, the upper bound on r_n is based on the quadratic remainder estimates provided in the end of page 14 and on page 28 (line 17) in [26]. The function h_0 coincides with the function ϵ_0^{-1} used in [26], and we have used the estimate provided on page 16 (line 13). The upper bound on $\overline{\beta}_n^2$ is a direct consequence of the estimates provided on page 9 (line 2) and on page 28 (line 13) in [26].

Thirdly, the following concentration inequality for unnormalized particle models γ_n^N is provided in [22] (see Theorem 6.13).

$\forall \epsilon \in \{+1, -1\}$ and $\forall y \geq 0$:

$$\mathbb{P}\left(\frac{\epsilon}{n}\log\left(\frac{\gamma_n^N(1)}{\gamma_n(1)}\right) \geq \frac{\bar{r}(n)}{N}h_0(y) + \tau_n^\star \bar{\sigma}_n^2 h_1\left(\frac{y}{N.\bar{\sigma}_n^2}\right)\right) \leq \mathrm{e}^{-y}, \tag{1.12}$$

where quantities $\tau_n^\star$, $\bar{\sigma}_n^2$ and $\bar{r}(n)$ can be estimated this way:

- $\tau_n^\star = \sup_{0\leq q\leq n} \tau_{q,n}$, where $\tau_{q,n}$ satisfy

$$\tau_{q,n} \leq \frac{4}{n}\sum_{p=q}^{n-1} g_{q,p}.g_{p+1}.b_{q,p}; \tag{1.13}$$

- $\bar{\sigma}_n^2 = \sum_{q=0}^{n-1} \sigma_q^2 (\frac{\tau_{q,n}}{\tau_n^\star})^2$ where σ_q satisfy $\sigma_q \leq 1$;
- $\bar{r}(n)$ satisfy

$$\bar{r}(n) \leq \frac{8}{n}\sum_{0\leq q\leq p<n} g_{p+1}.g_{q,p}^3.b_{q,p}. \tag{1.14}$$

2. Nonasymptotic theorems

The formulae (1.10), (1.11) and (1.12) provide explicit nonasymptotic estimates in terms of the quantities $g_{p,n}$ and $b_{p,n}$. Written this way, they hardly apply to any IPS parameters tuning decision, since the only known or calculable objects are generally the reference Markov chain M_p and the elementary potential functions G_p. We thus have to estimate $g_{p,n}$ and $b_{p,n}$ with some precision in terms of the g_p and b_p. This task is performed in Section 2.1. In the second section, Section 2.2, we combine these estimates with the concentration results presented in Section 1.5 to derive some useful uniform estimates w.r.t. the time parameter.

2.1. Semigroup estimates

We start with a series of technical lemmas.

Lemma 8. *Let K be a Markov kernel and G a positive function on E satisfying $\sup_{x,y\in E}\frac{G(x)}{G(y)}\leq g$, for some finite constant g. In this situation, we have that*

$$\sup_{x,y\in E}\frac{K.G(x)}{K.G(y)}\leq 1+\beta(K)(g-1).$$

Proof. Let $x, y\in E$ be s.t. $K.G(x)\geq K.G(y)$. Let us write

$$\frac{K.G(x)}{K.G(y)}=\frac{K.G(x)-K.G(y)}{K.G(y)}+1\leq\frac{\beta(K)(G_{\max}-G_{\min})}{G_{\min}}+1.$$

We check the last inequality using the fact that

$$K.G(x)-K.G(y)\leq \operatorname{osc}(K.G)\leq\beta(K).\operatorname{osc}(G)=\beta(K).(G_{\max}-G_{\min}).$$

On the other hand, we have $K.G(y)=\int_u G(u)K(y,\mathrm{d}u)\geq G_{\min}$. The desired result is now obtained taking the supremum over all $(x,y)\in E^2$. Note that $\frac{\beta(K)(G_{\max}-G_{\min})}{G_{\min}}+1$ is exactly equal to $1+\beta(K)(g-1)$.

This ends the proof of the lemma. □

Lemma 9. *Let M be a Markov kernel, Q a not necessarily normalized integral operator satisfying $\sup_{x,y\in E}\frac{Q.1(x)}{Q.1(y)}\leq g$, for some finite constant $g\geq 1$ and f a bounded, nonnegative function. In this situation, the Markov kernel P defined by*

$$P.f(x):=\frac{M.Q.f(x)}{M.Q.1(x)}$$

satisfies the following property:

$$\beta(P)\leq g.\beta(M).\beta\big(P'\big).$$

In the above display formula, P' is the Markov transition defined by $P'.f(x):=\frac{Q.f(x)}{Q.1(x)}$.

Proof. Note that $P.f(x)$ can be written in this way $P.f(x) = \Psi_{Q.1}(\delta_x.M)(P'.f)$. Thus, for any $x, y \in E$, we have that

$$\begin{aligned}\big|P.f(x) - P.f(y)\big| &= \big|\big(\Psi_{Q.1}(\delta_x.M) - \Psi_{Q.1}(\delta_y.M)\big)\big(P'.f\big)\big| \\ &\leq \big\|\big(\Psi_{Q.1}(\delta_x.M) - \Psi_{Q.1}(\delta_y.M)\big)\big\|_{\mathrm{tv}}.\operatorname{osc}\big(P'.f\big).\end{aligned}$$

Lemma 6 implies that

$$\big\|\big(\Psi_{Q.1}(\delta_x.M) - \Psi_{Q.1}(\delta_y.M)\big)\big\|_{\mathrm{tv}} \leq g.\|\delta_x.M - \delta_y.M\|_{\mathrm{tv}} \leq g.\beta(M).$$

Using (1.1), we have $\operatorname{osc}(P'.f) \leq \beta(P').\operatorname{osc}(f)$.
This ends the proof of the lemma. □

Lemma 10. *For any integers $p \leq n$, we have*

$$g_{p,n} - 1 \leq \sum_{k=p+1}^{n} (g_k - 1) \prod_{i=p+1}^{k-1} (b_i g_i), \tag{2.1}$$

$$b_{p,n} \leq \prod_{k=p+1}^{n} b_k.g_{k,n}. \tag{2.2}$$

Proof. Let us prove (2.1). By definition, we have $G_{p,n} = Q_{p,n}.1$. Combining (1.7) and (1.8) applied to unit function, we have

$$Q_{p-1,n}(1) = Q_p.\big[(Q_{p+1}, \ldots, Q_n).1\big] = G_p \times M_p.(Q_{p,n}.1).$$

This implies that the functions $G_{p,n}$ satisfy the following "backward" relations:

$$G_{n,n} = 1; \qquad G_{p-1,n} = G_p \times M_p.G_{p,n}.$$

Then, for any $x, y \in E$, we deduce that

$$\frac{G_{p-1,n}(x)}{G_{p-1,n}(y)} = \underbrace{\frac{G_p(x)}{G_p(y)}}_{E_1} \times \underbrace{\frac{(M_p.G_{p,n})(x)}{(M_p.G_{p,n})(y)}}_{E_2}.$$

Notice that $E_1 \leq g_p$ (by definition), and by Lemma 8, we have $E_2 \leq 1 + \beta(M_p).(g_{p,n} - 1)$. This shows the following backward inequalities:

$$g_{n,n} = 1; \qquad g_{p-1,n} \leq g_p\big(1 + b_p(g_{p,n} - 1)\big). \tag{2.3}$$

We end the proof of (2.1) by induction.

To prove (2.2), we use the formulae

$$P_{p-1,n}.f = \frac{Q_{p-1,n}.f}{Q_{p-1,n}.1} = \frac{G_p \times M_p.Q_{p,n}.f}{G_p \times M_p.Q_{p,n}.1} = \frac{M_p.Q_{p,n}.f}{M_p.Q_{p,n}.1}.$$

Recalling that $P_{p,n}.f = \frac{Q_{p,n}.f}{Q_{p,n}.1}$, we apply Lemma 9 to check that $\beta(P_{p-1,n}) \leq \beta(M_p).g_{p,n}. \times \beta(P_{p,n})$, from which we conclude that

$$b_{p-1,n} \leq b_p.g_{p,n}.b_{p,n}.$$

We end the proof of (2.2) by induction.
This ends the proof of the lemma. □

We end this section with a useful technical lemma to control the quantity $g_{p,n}b_{p,n}$.

Lemma 11. *For any $p \leq n$, we have*

$$g_{p,n}b_{p,n} \leq \prod_{k=p+1}^{n} (b_k.g_{k-1,n}).$$

Proof. Using Lemma 10, we have

$$\begin{aligned} g_{p,n}b_{p,n} &\leq g_{p,n}. \prod_{k=p+1}^{n} b_k g_{k,n} \\ &= g_{p,n}.(b_{p+1}g_{p+1,n}).(b_{p+2}g_{p+2,n}) \cdots (b_{n-1}g_{n-1,n}).(b_n \underbrace{g_{n,n}}_{=1}) \\ &= (g_{p,n}b_{p+1}).(g_{p+1,n}b_{p+2}) \cdots (g_{n-1,n}b_n) = \prod_{k=p+1}^{n} b_k g_{k-1,n}. \end{aligned}$$

This ends the proof of the lemma. □

The term $g_{p,n}b_{p,n}$ is central in the L^p-mean error bound (1.10). By Lemma 11, we have

$$\sum_{p=0}^{n} \prod_{k=p+1}^{n} b_k g_{k-1,n} < +\infty \quad \Longrightarrow \quad \sum_{p=0}^{n} g_{p,n}b_{p,n} < +\infty.$$

This gives a sufficient condition for a uniform L^p bound w.r.t. time n. $g_{p,n}b_{p,n}$ is also involved in the estimates of all the quantities defined in Section 1.5 such as r_n, $\overline{\beta_n}^2$, and others. In addition, by Lemma 6, we have the stability property

$$\left\| \phi_{p,n}(\mu) - \phi_{p,n}(\nu) \right\|_{\mathrm{tv}} \leq g_{p,n}b_{p,n} \|\mu - \nu\|_{\mathrm{tv}}.$$

This shows that the term $g_{p,n}b_{p,n}$ is central to quantify the stability properties of the semigroup $\phi_{p,n}$.

2.2. Uniform concentration theorems

To obtain uniform bounds w.r.t., the time horizon in (1.10), Lemma 11 naturally leads to a sufficient condition of the following type:

$$\sup_{k\leq n} b_k.g_{k-1,n} \leq a \qquad \text{for some } a\in(0,1).$$

In this situation, we prove that $g_{p,n}b_{p,n}\leq a^{n-p}$ and, therefore, using (1.10),

$$\sup_{f\in\mathcal{B}_1(E)} \mathbb{E}\big(\big|\eta_n^N(f)-\eta_n(f)\big|^p\big)^{1/p} \leq \frac{B_p}{\sqrt{N}}\frac{1}{1-a},$$

with the constants B_p introduced in (0.1). We then fix the parameter $a\in(0,1)$ and we look for conditions on the b_p so that $b_k g_{k-1,n}\leq a$. This parameter a can be interpreted as a performance degree of the N-approximation model. In order to explicit relevant and applicable conditions, we study two typical classes of regularity conditions on the potential functions G_p. The first one relates to bounded coefficients g_p (Theorem 12). In the second one, the parameters g_p tend to 1 as $p\to\infty$ (Theorem 13).

Theorem 12. *We assume that*

$$\sup_{p\geq 1} g_p \leq M \quad \text{and} \quad \sup_{p\geq 1} b_p \leq \frac{a}{M(1+a)} \tag{2.4}$$

for some finite $M\geq 1$ and some $a\in(0,1)$. In this situation, we have the following uniform estimates:

- *The L^p-error bound*:

$$\sup_{n\geq 0} d_p\big(\eta_n^N,\eta_n\big) \leq \frac{B_p}{2(1-a)\sqrt{N}}. \tag{2.5}$$

- *For any $n\geq 0$, $N\geq 1$, $y\geq 0$ and $f\in\mathcal{O}_1(E)$, the probability of the event*

$$\eta_n^N(f)-\eta_n(f) \leq \frac{r_1^\star(1+h_0(y))+r_2^\star\sqrt{Ny}}{N} \tag{2.6}$$

  is greater than $1-\mathrm{e}^{-y}$, with the parameters

$$r_1^\star=\frac{4M^2(1+a)^2}{1-a} \quad \text{and} \quad r_2^\star=\frac{2\sqrt{2}}{\sqrt{1-a^2}}. \tag{2.7}$$

- *For any $n\geq 0$, $N\geq 1$, $\epsilon\in\{+1,-1\}$, and $y\geq 0$, the probability of the event*

$$\frac{\epsilon}{n}\log\left(\frac{\gamma_n^N(1)}{\gamma_n(1)}\right) \leq \frac{\tilde{r}_1}{N}h_0(y)+\tilde{r}_2.h_1\left(\frac{y}{n.N}\right) \tag{2.8}$$

  is greater than $1-\mathrm{e}^{-y}$, with the parameters $\tilde{r}_1=\frac{8M^3(1+a)^2}{1-a}$ and $\tilde{r}_2=\frac{4M}{1-a}$. The functions h_0 and h_1 are defined in (1.9), *page* 680.

Proof. Firstly, we prove the pair of inequalities

$$\begin{cases} g_{p,n} \leq M(1+a), \\ b_p . g_{p-1,n} \leq a. \end{cases} \tag{2.9}$$

If $g_p \leq M$ and $b_p \leq b := \frac{a}{M(1+a)}$, then by Lemma 10

$$g_{p,n} - 1 \leq \sum_{k=p+1}^{n} (M-1)(bM)^{k-p-1} = (M-1)\frac{1-(bM)^{n-p}}{1-bM} \leq \frac{M-1}{1-bM}.$$

We substitute $b = \frac{a}{M(1+a)}$ to find

$$g_{p,n} \leq \frac{M-1}{1-a/(1+a)} + 1 = (M-1)(1+a) + 1 \leq M(1+a).$$

For the second inequality, we have

$$\begin{aligned} b_p . g_{p-1,n} &\leq \frac{b(M-1)}{1-a/(1+a)} + b \\ &= (1+a)\frac{a(M-1)}{M(1+a)} + \frac{a}{M(1+a)} \\ &= \frac{1}{M}\left(aM - a + \frac{a}{1+a}\right) \\ &\leq a. \end{aligned}$$

This ends the proof of (2.9).

- By Lemma 11 and (2.9), we have $g_{p,n} b_{p,n} \leq a^{n-p}$. Combining this with (1.10), the L^p-error bound (2.5) is clear. See the relation (A.1), page 706 for the factor $1/2$.
- Let us prove (2.6), which is a consequence of the concentration inequality (1.11). Combining the estimations of r_n and $\overline{\beta_n}^2$ given in Section 1.5 with (2.9) and $g_{p,n} b_{p,n} \leq a^{n-p}$, we deduce that

$$r_n \leq \frac{4M^2(1+a)^2}{1-a} \quad \text{and} \quad \overline{\beta_n}^2 \leq \frac{4}{1-a^2}.$$

  Equation (2.6) is obtained by making the suitable substitutions in (1.11).
- The last concentration inequality (2.8) is a consequence of (1.12) and (2.9). Let us recall that Lemma 11 and (2.9) imply $g_{p,n} b_{p,n} \leq a^{n-p}$. Then, from estimations (1.13) and (1.14), we can easily show that the quantities $\tau_n^\star$ and $\bar{r}(n)$ satisfy

$$\tau_n^\star \leq \frac{4M}{n(1-a)} \quad \text{and} \quad \bar{r}(n) \leq \frac{8M^3(1+a)^2}{1-a}.$$

On the other hand, $\bar{\sigma}_n^2$ is trivially bounded by n. Then we find that

$$\frac{\bar{r}(n)}{N}h_0(y) + \tau_n^\star \bar{\sigma}_n^2 h_1\left(\frac{y}{N.\bar{\sigma}_n^2}\right) = \frac{\bar{r}(n)}{N}h_0(y) + \frac{y\tau_n^\star}{3N} + \sqrt{\frac{2y(\tau_n^\star \bar{\sigma}_n)^2}{N}}.$$

Finally, (2.8) is obtained from (1.12) by making the suitable substitutions.

This ends the proof of the theorem. □

Let us now consider the case where g_p decreases to 1 as $p \to \infty$. The idea of the forthcoming analysis is to find a condition on the b_p so that the $g_{p,n}$ are uniformly bounded w.r.t. n by $g_{p+1}^{1+\alpha}$ with

$$\alpha = \frac{a}{1-a} > 0 \qquad \left(\Longleftrightarrow a = \frac{\alpha}{1+\alpha}\right).$$

The concentration inequalities developed in Theorem 13 will be described in terms of the parameters $r_3^\star(n)$ and $\tilde{r}_3(n), \tilde{r}_4, \tilde{r}_5(n)$ defined below.

$$r_3^\star(n) = \frac{4u_1(n)}{1-a}, \tag{2.10}$$

$$\begin{cases} \tilde{r}_3(n) = \dfrac{16.u_2(n)}{1-a}, \\ \tilde{r}_4 = \dfrac{4}{3}\displaystyle\sum_{n\geq 0} g_{n+1}.a^n \leq \frac{4.g_1}{3(1-a)}, \\ \tilde{r}_5(n) = \dfrac{4\sqrt{2}.u_3(n)}{1-a}. \end{cases} \tag{2.11}$$

The sequences $u_1(n)$, $u_2(n)$ and $u_3(n)$ used in the above formulae are defined by

$$\begin{cases} u_1(n) = (1-a)\displaystyle\sum_{p=0}^{n} g_{n-p+1}^{2(1+\alpha)} a^p \underset{n\to\infty}{\longrightarrow} 1, \\ u_2(n) = \dfrac{1}{n}\displaystyle\sum_{p=1}^{n} g_p^{3+2\alpha} \underset{n\to\infty}{\longrightarrow} 1, \\ u_3(n) = \left(\dfrac{1}{n}\displaystyle\sum_{p=0}^{n-1} g_{p+1}^2\right)^{1/2} \underset{n\to\infty}{\longrightarrow} 1. \end{cases}$$

Notice that the sequence $u_1(n)$ tends to 1 by dominated convergence. Sequences $u_2(n)$ and $u_3(n)$ tend to 1 by Cesaro's theorem.

Theorem 13.
We assume that $g_p \downarrow 1$ as $p \to \infty$ and the sequence b_p satisfies for any $p \geq 1$

$$b_p \leq \frac{g_p^\alpha - 1}{g_p^{\alpha+1} - 1} \left(\underset{p \to +\infty}{\longrightarrow} a \right) \quad \text{and} \quad b_p \leq \frac{a}{g_p^{\alpha+1}} \left(\underset{p \to +\infty}{\longrightarrow} a \right). \tag{2.12}$$

In this situation, we have the following uniform estimates:

- *The L^p-error bound*

$$\sup_{n \geq 0} d_p\left(\eta_n^N, \eta_n\right) \leq \frac{B_p}{2(1-a)\sqrt{N}} \tag{2.13}$$

  with the constants B_p introduced in (0.1).
- *For any $n \geq 0$, $N \geq 1$, $y \geq 0$ and $f \in \mathcal{O}_1(E)$, the probability of the event*

$$\eta_n^N(f) - \eta_n(f) \leq \frac{r_3^\star(n)(1 + h_0(y)) + r_2^\star \sqrt{Ny}}{N} \tag{2.14}$$

  is greater than $1 - \mathrm{e}^{-y}$, with the bounded sequence $r_3^\star(n)$ defined in (2.10), the parameter $r_2^\star$ defined in (2.7) and the function h_0 defined in (1.9), page 680.
- *For any $n \geq 0$, $N \geq 1$, $\epsilon \in \{+1, -1\}$, and $y \geq 0$, the probability of the event*

$$\frac{\epsilon}{n} \log\left(\frac{\gamma_n^N(1)}{\gamma_n(1)} \right) \leq \tilde{r}_3(n) \left(\frac{y + \sqrt{y}}{N} \right) + \tilde{r}_4\left(\frac{y}{n.N} \right) + \tilde{r}_5(n) \sqrt{\frac{y}{n.N}} \tag{2.15}$$

  is greater than $1 - \mathrm{e}^{-y}$, with the parameters $\tilde{r}_3(n), \tilde{r}_4, \tilde{r}_5(n)$ defined in (2.11).

Proof. Firstly, we prove that

$$(2.12) \quad \Longrightarrow \quad \forall p \leq n, \qquad \begin{cases} g_{p,n} \leq (g_{p+1})^{1+\alpha}, \\ g_{p-1,n}.b_p \leq a. \end{cases} \tag{2.16}$$

The proof of the first inequality comes from a simple backward induction on p (with fixed n), using formula $g_{p-1,n} \leq g_p(1 + b_p(g_{p,n} - 1))$ (see (2.3)). For $p = n$, $g_{p,n}$ is clearly smaller than $g_{p+1}^{1+\alpha}$ because $g_{n,n} = 1$. The second assertion is now immediate.

Now we assume that $g_{p,n} \leq g_{p+1}^{1+\alpha}$. In this case, $g_{p-1,n} \leq g_p^{1+\alpha}$ is met as soon as

$$b_p \leq \frac{g_p^\alpha - 1}{g_{p+1}^{\alpha+1} - 1}.$$

Notice that this estimate is met as soon as $b_p \leq \frac{g_p^\alpha - 1}{g_p^{\alpha+1} - 1}$, the sequence $(g_p)_p$ being decreasing.

- Now that we proved (2.16) (which implies $g_{p,n} b_{p,n} \leq a^{n-p}$ by Lemma 11), the L^p-mean error bound (2.13) comes from a simple substitution in (1.10). See the relation (A.1), page 706 for the factor $1/2$.

- To prove (2.14), we focus on the quantities $\overline{\beta_n}$ and r_n arising in the concentration inequality (1.11). With (2.16) and $g_{p,n}b_{p,n} \leq a^{n-p}$, we readily verify that

$$\overline{\beta_n}^2 \leq \frac{4}{1-a^2}.$$

The term r_n can be roughly bounded by $\frac{4}{1-a} g_1^{2(1+\alpha)}$, but another manipulation provides a more precise estimate. Indeed, using the fact that $b_{p,n}.g_{p,n} \leq a^{n-p}$ and $g_{p,n} \leq g_{p+1}^{1+\alpha}$, we prove that

$$\begin{aligned} r_n &\leq 4\sum_{p=0}^{n} g_{p,n}^2 a^{n-p} \leq 4\sum_{p=0}^{n} g_{p+1}^{2(1+\alpha)} a^{n-p} \leq 4\sum_{p=0}^{n} g_{n-p+1}^{2(1+\alpha)} a^p \\ &\leq \frac{4.u_1(n)}{1-a}. \end{aligned}$$

We prove (2.14) by making the suitable substitutions in (1.11).

- Let us prove the last concentration inequality (2.15). It is mainly a consequence of the inequality (1.12). Starting from the following decomposition,

$$\frac{\bar{r}(n)}{N} h_0(y) + \tau_n^\star \bar{\sigma}_n^2 h_1\left(\frac{y}{N.\bar{\sigma}_n^2}\right) = \frac{2\bar{r}(n)}{N}(y+\sqrt{y}) + \frac{y\tau_n^\star}{3N} + \sqrt{\frac{2y(\tau_n^\star \bar{\sigma}_n)^2}{N}}, \tag{2.17}$$

we need to find some refined estimates of the quantities $\tau_n^\star$, $\bar{r}_n$ and $(\tau_n^\star.\bar{\sigma}_n)^2$. To estimate $\tau_n^\star$, we notice that $\forall q$, $g_{p+1} \leq g_{p-q+1}$, so that

$$\begin{aligned} \tau_{q,n} &\leq \frac{4}{n}\sum_{p=q}^{n-1} g_{p+1} a^{p-q} \leq \frac{4}{n}\sum_{p=q}^{n-1} g_{p-q+1} a^{p-q} \leq \frac{4}{n}\sum_{p=0}^{n-q-1} g_{p+1} a^p \\ &\leq \frac{4}{n}\sum_{p=0}^{n-1} g_{p+1} a^p. \end{aligned}$$

Finally, we have that $\tau_n^\star \leq \frac{4.U_0}{n}$, where $U_0 = \sum_{p\geq 0} g_{p+1}a^p \leq \frac{g_1}{1-a}$.

We estimate $\bar{r}_n$, using the following inequalities:

$$\begin{aligned} \bar{r}_n &\leq \frac{8}{n} \sum_{0\leq q\leq p<n} g_{p+1}.g_{q,p}^3.b_{q,p} \leq \frac{8}{n}\sum_{p=0}^{n-1}\sum_{q=0}^{p} g_{p+1}^{3+2\alpha} a^{p-q} \\ &\leq \frac{8}{1-a} \cdot \underbrace{\frac{1}{n}\sum_{p=1}^{n} g_p^{3+2\alpha}}_{=u_2(n)}. \end{aligned}$$

Let us conduct a last useful estimation:

$$\begin{aligned}\left(\tau_n^\star.\bar{\sigma}_n\right)^2 &\leq \sum_{q=0}^{n-1}\tau_{q,n}^2 \leq \sum_{q=0}^{n-1}\left(\frac{4}{n}\sum_{p=q}^{n-1} g_{p+1}.a^{p-q}\right)^2\\ &\leq \sum_{q=0}^{n-1}\left(\frac{4}{n}\sum_{p=0}^{n-q+1} g_{p+q+1}.a^{p}\right)^2\\ &\leq \sum_{q=0}^{n-1}\frac{16}{n^2}\cdot g_{q+1}^2\cdot\frac{1}{(1-a)^2}\\ &\leq \frac{16}{n.(1-a)^2}\times\underbrace{\frac{1}{n}\sum_{q=0}^{n-1} g_{q+1}^2}_{=(u_3(n))^2}.\end{aligned}$$

At last, we make the suitable substitutions in (2.17) and obtain the desired inequality (2.15). This ends the proof of the theorem. □

3. Interacting simulated annealing models

3.1. Some motivations

We consider the Boltzmann–Gibbs probability measure associated with an inverse "temperature" parameter $\beta \geq 0$ and a given potential function $V \in \mathcal{B}(E)$ defined by

$$\mu_\beta(\mathrm{d}x) = \frac{1}{Z_\beta}\mathrm{e}^{-\beta.V(x)}m(\mathrm{d}x), \tag{3.1}$$

where m stands for some reference measure, and Z_β is a normalizing constant. We let β_n a strictly increasing sequence (which may tend to infinity as $n \to \infty$). In this case, the measures $\eta_n = \mu_{\beta_n}$ can be interpreted as a FK flow of measures with potential functions $G_n = \mathrm{e}^{-(\beta_n-\beta_{n-1})V}$ and Markov transitions M_n chosen as being MCMC dynamics for the current target distributions. Indeed, we have

$$\mu_{\beta_n}(\mathrm{d}x) = \frac{\mathrm{e}^{-\beta_n.V(x)}}{Z_{\beta_n}}m(\mathrm{d}x) = \frac{Z_{\beta_{n-1}}}{Z_{\beta_n}}\underbrace{\mathrm{e}^{-(\beta_n-\beta_{n-1})V(x)}}_{=G_n(x)}\underbrace{\left(\frac{\mathrm{e}^{-\beta_{n-1}.V(x)}}{Z_{\beta_{n-1}}}m(\mathrm{d}x)\right)}_{=\mu_{\beta_{n-1}}(\mathrm{d}x)}.$$

This shows $\mu_{\beta_n} = \psi_{G_n}(\mu_{\beta_{n-1}})$. Let ϕ_n stand for the FK transformation associated with potential function G_n and Markov transition M_n. We have

$$\phi_n(\mu_{\beta_{n-1}}) = \psi_{G_n}(\mu_{\beta_{n-1}}) = \mu_{\beta_n}.M_n = \mu_{\beta_n}.$$

Sampling from these distributions is a challenging problem in many application domains. The simplest one is to sample from a complex posterior distribution on some Euclidian space $E = \mathbb{R}^d$, for some $d \geq 1$. Let x be a variable of interest, associated with a prior density $p(x)$ (easy to sample) with respect to Lebesgue measure $\mathrm{d}x$ on E, and y a vector of observations, associated with a calculable likelihood model $p(y|x)$. In this context, we recall that $p(y|x)$ is the density of the observations given the variable of interest. The density $p(x|y)$ of the posterior distribution η is given by Bayes' formula

$$p(x|y) \propto p(x).p(y|x).$$

In the case where η is highly multimodal, it is difficult to sample from it directly. As an example, classic MCMC methods tend to get stuck in local modes for very long times. As a result, they converge to their equilibrium η only on unpractical time-scales. To overcome this problem, a common solution is to approximate the target distribution η with a sequence of measures $\eta_0, \ldots, \eta_n$ with density

$$\eta_k(\mathrm{d}x) \propto p(x).p(y|x)^{\beta_k}\,\mathrm{d}x,$$

where $(\beta_k)_{0\leq k\leq n}$ is a sequence of number increasing from 0 to 1, so that η_0 is the prior distribution of density $p(x)$, easy to sample, and the terminal measure η_n is the target distribution η (see, for instance, [4,30,37,38]). If we take $V := x \mapsto -\log(p(y|x))$ and $m(\mathrm{d}x) := p(x)\,\mathrm{d}x$, then the η_k coincide with the Boltzmann–Gibbs measures μ_{β_k} defined in (3.1). In this context, IPS methods arise as being a relevant approach, especially if η is multimodal, since the use of a large number of particles allows to cover several modes simultaneously.

The normalizing constant $\gamma_n(1)$ coincide with the marginal likelihood $p(y)$. Computing this constant is another central problem in model selection arising in hidden Markov chain problems and Bayesian statistics.

Next, we present another important application in physics and chemistry, known as free energy estimation. The problem starts with an unnormalized density of the form

$$q(\omega|T,\alpha) = \exp\left(-\frac{H(\omega,\alpha)}{k.T}\right),$$

where $H(\omega,\alpha)$ is the energy function of state ω, k is Boltzmann's constant, T is the temperature and α is a vector of system characteristics. The free energy F of the system is defined by the quantity

$$F(T,\alpha) = -k.T.\log\big(z(T,\alpha)\big),$$

where $z(T,\alpha)$ is the normalizing constant of the system density. See, for instance, [7,12,29] for a further discussion on these ground state energy estimation problems.

Last, but not least, it is well known that Boltzmann–Gibbs measures' sampling is related to the problem of minimizing the potential function V. The central idea is that μ_β tends to concentrate on V's minimizers as the inverse temperature β tends to infinity. To be more precise, we provide an exponential concentration inequality in Lemma 14.

In this context, the IPS algorithm can be interpreted as a sequence of interacting simulated annealing (*abbreviate ISA*) algorithms. As they involve a population of N individuals, evolving

according to genetic type processes (selection, mutation), ISA methods also belong to the rather huge class of evolutionary algorithms for global optimization. These algorithms consist in exploring a state space with a population associated with an evolution strategy, that is, an evolution based on selection, mutation and crossover. See [31] or [36] and the references therein for an overview. As these algorithms involve complex, possibly adaptive strategies, their analysis is essentially heuristic, or sometimes asymptotic (see [8] for general convergence results on genetic algorithms). The reader will also find in [24] a proof for a ISA method of the a.s. convergence to the global minimum in the case of a finite state space, when the time n tends to infinity, and as soon as the population size N is larger than a critical constant that depends on the oscillations of the potential fitness functions and the mixing properties of the mutation transitions.

The results of the previous sections apply to the analysis of ISA optimization methods. Our approach is nonasymptotic since we estimate at each time n, and for a fixed population size N the distance between the theoretical Boltzmann–Gibbs measure η_n and its empirical approximation η_n^N. In some sense, as $\beta_n \to +\infty$, η_n tends to the Dirac measure $\delta_{x^\star}$ where $x^\star = \operatorname{argmin}_{x\in E} V(x)$, as soon as there is a single global minimum. So intuitively, we guess that if this distance admits a uniform bound w.r.t. time n, then for large time horizon we have $\eta_n^N \simeq \delta_{x^\star}$.

In this section, we propose to turn the conditions of Theorems 12 and 13 into conditions on the temperature schedule to use, and the number of MCMC steps that ensure a given performance degree. Then we combine the concentration results of Section 2.2 with Lemma 14 to analyze the convergence of the IPS optimization algorithm.

3.2. An ISA optimization model

We fix an inverse temperature schedule β_n and we set

- $\eta_n(\mathrm{d}x) = \mu_{\beta_n}(\mathrm{d}x) = \frac{1}{Z_{\beta_n}} \mathrm{e}^{-\beta_n V(x)} m(\mathrm{d}x)$;
- $G_n(x) = \mathrm{e}^{-\Delta_n . V(x)}$;
- and then $g_n = \mathrm{e}^{\Delta_n . \operatorname{osc}(V)}$,

where Δ_n are the increments of temperature $\Delta_n = \beta_n - \beta_{n-1}$. We let K_β the simulated annealing Markov transition with invariant measure μ_β and a proposition kernel $K(x, \mathrm{d}y)$ reversible w.r.t. $m(\mathrm{d}x)$. We recall that $K_\beta(x, \mathrm{d}y)$ is given by the following formulae:

$$K_\beta(x, \mathrm{d}y) = K(x, \mathrm{d}y).\min\bigl(1, \mathrm{e}^{-\beta(V(y)-V(x))}\bigr) \qquad \forall y \neq x,$$

$$K_\beta\bigl(x, \{x\}\bigr) = 1 - \int_{y\neq x} K(x, \mathrm{d}y).\min\bigl(1, \mathrm{e}^{-\beta(V(y)-V(x))}\bigr).$$

Under the assumption $K^{k_0}(x, \cdot) \geq \delta \nu(\cdot)$ for any x with some integer $k_0 \geq 1$, some measure ν and some $\delta > 0$, one can show (see, e.g., [3], Lemma 4.3.11) that

$$\beta\bigl(K_\beta^{k_0}\bigr) \leq \bigl(1 - \delta \mathrm{e}^{-\beta \overline{\Delta V}(k_0)}\bigr). \tag{3.2}$$

In the above display formula, $\overline{\Delta V}(k_0) = \sup_{x\in E} \|V(x) - V(X(x))\|_\infty$, where for all $x \in E$, $X(x)$ is a random variable of law $\delta_x . K^{k_0}$. In other words, $\overline{\Delta V}(k_0)$ is the maximum potential

gap one can obtain making k_0 elementary moves with the Markov transition K. It is clearly bounded by $\operatorname{osc}(V)$. One way to control the mixing properties of the ISA model is to consider the Markov transition $M_p = K_{\beta_p}^{k_0.m_p}$, the simulated annealing kernel iterated $k_0.m_p$ times. In this case, the user has a choice to make on two tuning parameters, namely the temperature schedule β_p, and the iteration numbers m_p. Note that for all $b \in (0,1)$, condition $b_p \leq b$ in turned into $(1-\delta \mathrm{e}^{-\beta_p \overline{\Delta V}(k_0)})^{m_p} \leq b$, and this last condition is satisfied if we have

$$m_p \geq \frac{\log(1/b)\mathrm{e}^{\overline{\Delta V}(k_0).\beta_p}}{\delta}. \tag{3.3}$$

We prove now a technical lemma that we will use in the following. It deals with Boltzmann–Gibbs measures' concentration properties.

Lemma 14. *For any $\beta > 0$, and for all $0 < \varepsilon' < \varepsilon$, the Boltzmann–Gibbs measure μ_β satisfies*

$$\mu_\beta(V \geq V_{\min} + \varepsilon) \leq \frac{\mathrm{e}^{-\beta(\varepsilon-\varepsilon')}}{m_{\varepsilon'}},$$

where $m_{\varepsilon'} = m(V \leq V_{\min} + \varepsilon') > 0$.

Proof. The normalizing constant Z_β of the definition (3.1) is necessary equal to $\int_E \mathrm{e}^{-\beta V}\,\mathrm{d}m$. Then we have

$$\begin{aligned}\mu_\beta(V \geq V_{\min} + \varepsilon) &= \frac{\int_{V \geq V_{\min}+\varepsilon} \mathrm{e}^{-\beta V}\,\mathrm{d}m}{\int_{V \geq V_{\min}+\varepsilon} \mathrm{e}^{-\beta V}\,\mathrm{d}m + \int_{V < V_{\min}+\varepsilon} \mathrm{e}^{-\beta V}\,\mathrm{d}m} \\ &\leq \underbrace{\left(\int_{V \geq V_{\min}+\varepsilon} \mathrm{e}^{-\beta V}\,\mathrm{d}m\right)}_{A_1} \underbrace{\left(\int_{V < V_{\min}+\varepsilon} \mathrm{e}^{-\beta V}\,\mathrm{d}m\right)^{-1}}_{A_2^{-1}}.\end{aligned}$$

Firstly, it is clear that $A_1 \leq \mathrm{e}^{-\beta(V_{\min}+\varepsilon)}$. Secondly, $\varepsilon' < \varepsilon$ implies $\{V \leq V_{\min} + \varepsilon'\} \subset \{V < V_{\min} + \varepsilon\}$, then we have

$$A_2 \geq \int_{V \leq V_{\min}+\varepsilon'} \mathrm{e}^{-\beta V}\,\mathrm{d}m \geq m\big(V \leq V_{\min} + \varepsilon'\big)\mathrm{e}^{-\beta(V_{\min}+\varepsilon')}.$$

We end the proof by making the appropriate substitutions. □

Combining Lemma 14, the theorems of Section 2.2 (with indicator test function $f = \mathbf{1}_{\{V \geq V_{\min}+\varepsilon\}}$), and the Dobrushin ergodic coefficient estimate (3.2) (and the associated remark (3.3)) we prove the following theorem, which is the complete version of Theorem 2.

Theorem 15. *Let us consider the N-particles ISA algorithm associated with the potential functions G_n and Markov transitions M_n defined in the beginning of this section. Let us fix $a \in (0,1)$.*

For any $\varepsilon > 0$, and $n \geq 0$, let $p_n^N(\varepsilon)$ denote the proportion of particles (ζ_n^i) s.t. $V(\zeta_n^i) \geq V_{\min} + \varepsilon$. We assume that the inverse temperature schedule β_p and the iteration numbers m_p satisfy one of these two conditions:

1. $\sup_{p \geq 1} \Delta_p \leq \Delta < \infty$ (*e.g.*, *linear temperature schedule*) *and*

$$m_p \geq \frac{\log(\mathrm{e}^{\Delta \cdot \mathrm{osc}(V)}(1+a)/a)\mathrm{e}^{\overline{\Delta V}(k_0).\beta_p}}{\delta}.$$

2. $\Delta_p \downarrow 0$ (*as* $p \to \infty$) *and* $m_p \geq (\mathrm{osc}(V).\Delta_p + \log(\frac{1}{a}))\frac{\mathrm{e}^{\overline{\Delta V}(k_0).\beta_p}}{\delta}$.

In this situation, for any $\varepsilon > 0$, $n \geq 0$, $N \geq 1$, $y \geq 0$, and $\varepsilon' \in (0, \varepsilon)$, the probability of the event

$$p_n^N(\varepsilon) \leq \frac{\mathrm{e}^{-\beta_n(\varepsilon-\varepsilon')}}{m_{\varepsilon'}} + \frac{r_i^\star(n)(1+h_0(y)) + r_2^\star\sqrt{Ny}}{N}$$

is greater than $1-\mathrm{e}^{-y}$. In the above display formula, the constant $r_2^\star$ is defined in (2.7), *page* 685, *the function h_0 in* (1.9), *page* 680 *and $r_i^\star(n)$ is defined by*:

- $r_i^\star(n) = r_1^\star$ (*see* (2.7), *page* 685 *with* $M = \mathrm{e}^{\Delta \cdot \mathrm{osc}(V)}$) *in the case of bounded* Δ_p;
- $r_i^\star(n) = r_3^\star(n)$ (*see* (2.10), *page* 687 *with* $g_n = \mathrm{e}^{\Delta_n \mathrm{osc}(V)}$) *in the second one.*

We distinguish two error terms. The first one, $(\frac{\mathrm{e}^{-\beta_n(\varepsilon-\varepsilon')}}{m_{\varepsilon'}})$, is related to the concentration of the Boltzmann–Gibbs measure around the set of global minima of V. The second one, $(\frac{r_i^\star(n)(1+h_0(y))+r_2^\star\sqrt{Ny}}{N})$, is related to the concentration of the occupation measure around the limiting Boltzmann–Gibbs measure. Besides the fact that Theorem 15 provides tuning strategies which ensure the performance of the ISA model, the last concentration inequality is explicit in the relative importance of other parameters, including the probabilistic precision y, the threshold t on the proportion of particles possibly out of the area of interest, the final temperature β_n and the population size N. As long as one can have rough estimates of $m_{\varepsilon'}$ and $\mathrm{osc}(V)$ (spatial scale and variation of the function V), a simple equation, deduced from this last theorem, such as $(\frac{\mathrm{e}^{-\beta_n(\varepsilon-\varepsilon')}}{m_{\varepsilon'}} = \frac{r_i^\star(n)(1+h_0(y))+r_2^\star\sqrt{Ny}}{N} = \frac{t}{2})$ provides some orders of magnitude that can be useful to roughly design an ISA model at first sight, which is generally a difficult task.

One natural way to choose Δ_p is to look at n_1, the number of transitions of type $K_\beta^{k_0}$ we need to proceed to move from β_p to β_n, and to compare it to n_2, the number of transitions we need to proceed to move from β_p to β_q, and then from β_q to β_n, with $\beta_p < \beta_q < \beta_n$. Roughly speaking, we have seen that the convergence condition was $b_p \leq \frac{a}{g_p}$, then, using (3.3), we have

- $n_1 \simeq (\mathrm{osc}(V).\Delta_{p,n} + \log(\frac{1}{a}))\frac{\mathrm{e}^{\overline{\Delta V}(k_0).\beta_n}}{\delta}$,
- $n_2 \simeq (\mathrm{osc}(V).\Delta_{p,q} + \log(\frac{1}{a}))\frac{\mathrm{e}^{\overline{\Delta V}(k_0).\beta_q}}{\delta} + (\mathrm{osc}(V).\Delta_{q,n} + \log(\frac{1}{a}))\frac{\mathrm{e}^{\overline{\Delta V}(k_0).\beta_n}}{\delta}$,

where $\Delta_{p,q} := \beta_p - \beta_q$ for $p > q$. After some approximation technique we find that

$$n_1 \leq n_2 \quad \Longleftrightarrow \quad \Delta_{p,q}\Delta_{q,n} \geq \frac{\log(1/a)}{\mathrm{osc}(V).\overline{\Delta V}(k_0)}.$$

This condition does not bring any relevant information in the case where $\Delta_p \longrightarrow 0$ except that the error decomposition $\eta_n^N - \eta_n = \sum_{p=0}^n \phi_{p,n}(\eta_p^N) - \phi_{p,n}\phi_p(\eta_{p-1}^N)$, underlying our analysis, is not adapted to the case where $\Delta_p \longrightarrow 0$ (which can be compared to the continuous time case). Nevertheless, this condition is interesting in the case of constant inverse temperature steps. In this situation, the critical parameter Δ_β is given by

$$\Delta_\beta = \sqrt{\frac{\log(1/a)}{\mathrm{osc}(V).\overline{\Delta V}(k_0)}}.$$

More precisely, above Δ_β the algorithm needs to run too many MCMC steps to stabilize the system. In the reverse angle, when the variation of temperature is too small, it is difficult to reach the desired target measure.

4. An adaptive temperature schedule in ISA

As we already mentioned in the Introduction, the theoretical tuning strategies developed in Section 3.2 are of the same order as the logarithmic cooling schedule of traditional simulated annealing [3,8,24]. In contrast to SA models, we emphasize that the performance of the ISA models are not based on a critical initial temperature parameter. Another advantage of the ISA algorithm is to provide at any time step an N-approximation of the target measure with a given temperature. In other words, the population distribution reflects the probability mass distribution of the Boltzmann–Gibbs measure at that time. Computationally speaking, the change of temperature parameter Δ_p plays an important role. For instance, if Δ_p is taken too large, the selection process is dominated by a minority of well fitted particles and the vast majority of the particles are killed. The particle set's diversity, which is one of the main advantage of the ISA method, is then lost. On the contrary, if Δ_p is taken too small, the algorithm does not proceed to an appropriate selection. It wastes time by sampling from MCMC dynamics while the set of particles has already reached its equilibrium. The crucial point is to find a relevant balance between maintaining diversity and avoiding useless MCMC operations.

Designing such a balance in advance is almost as hard as knowing the function V in advance. Therefore, it is natural to implement adaptive strategies that depend on the variability and the adaptation of the population particles (see, e.g., [13,27,34,37,39] for related applications). In the general field of evolutionary algorithms, elaborating adaptive selection strategies is a crucial question (see, e.g., [2]) and a challenging problem to design performant algorithms. In the case of ISA methods, the common ways to choose Δ_p are based on simple criteria such as the expected number of particle killed (see Section 4.2), or the variance of the weights (effective sample size). All of these criteria are based on the same intuitive idea; that is to achieve a reasonable selection. As a result, all of these adaptive ISA models tend to perform similarly.

In [18], the reader will find a general formalization of adaptive IPS algorithms. The idea is to define the adaptation as the choice of the times n at which the resampling occurs. These times are chosen according to some adaptive criteria, depending on the current particle set, or more generally on the past process. Under weak conditions on the criteria, it is shown how the adaptive process asymptotically converges to a static process involving deterministic interaction times

when the population size tends to infinity. A functional central limit theorem is then obtained for a large class of adaptive IPS algorithms.

The approach developed in the following section is radically different, with a special focus on nonasymptotic convergence results for the ISA algorithm defined in Section 4.2. The adaptation consists here in choosing the β increment Δ_{n+1}^N so that

$$\eta_n^N\left(\mathrm{e}^{-\Delta_{n+1}^N \cdot V}\right) = \varepsilon,$$

where $\varepsilon > 0$ is a given constant, at each iteration n. We show that the associated stochastic process can be interpretated as a perturbation of the limiting FK flow.

4.1. Feynman–Kac representation

Let $V \in \mathcal{B}(E)$. To simplify the analysis, without any loss of generality, we assume $V_{\min} = 0$. Let us fix $\varepsilon > 0$. For any measure $\mu \in \mathcal{P}(E)$, we let the function λ_μ defined by

$$\begin{aligned} &[0, +\infty) \longrightarrow (0, 1], \\ &\lambda_\mu = x \mapsto \mu\left(\mathrm{e}^{-x \cdot V}\right). \end{aligned}$$

This function is clearly decreasing ($\lambda_\mu(0) = 1$), convex, and differentiable infinitely. Moreover, if $\mu(\{V = 0\}) = 0$, then it satisfies $\lambda_\mu(x) \longrightarrow 0$ when $x \to +\infty$. Therefore, we can define its inverse function κ_μ:

$$\begin{aligned} &(0, 1] \longrightarrow [0, +\infty), \\ &\kappa_\mu = \varepsilon \mapsto x \qquad \text{so that } \mu\left(\mathrm{e}^{-x \cdot V}\right) = \varepsilon. \end{aligned}$$

This function is again decreasing, convex, infinitely differentiable, takes value 0 for $\varepsilon = 1$, and it satisfies $\kappa_\mu(\varepsilon) \longrightarrow +\infty$ when $\varepsilon \to 0^+$.

Now, we let m be a reference measure on E s.t. $m(\{V = 0\}) = 0$. We consider the sequence $(\beta_n)_n$ and its associated Gibbs measures $\eta_n = \mu_{\beta_n} \propto \mathrm{e}^{-\beta_n V}.m$, defined recursively by the equation

$$\Delta_{n+1} := (\beta_{n+1} - \beta_n) = \kappa_{\eta_n}(\varepsilon). \tag{4.1}$$

In an equivalent way, we have

$$\lambda_{\eta_n}(\Delta_{n+1}) = \varepsilon \quad \text{or} \quad \eta_n\left(\mathrm{e}^{-\Delta_{n+1} \cdot V}\right) = \varepsilon.$$

The main objective of this section is to approximate these target measures. Formally speaking, (η_n) admits the FK structure described in Section 3.1, with potential functions $G_n(x) = \mathrm{e}^{-\Delta_n \cdot V(x)}$, and some dedicated MCMC Markov kernels M_n. We let g_n, b_n, be the associated oscillations, Dobrushin ergodic coefficients and the corresponding FK transformations ϕ_n.

The solving of the equation $\eta_n(\mathrm{e}^{-\Delta_{n+1}\cdot V}) = \varepsilon$ can be interpreted as a way to impose some kind of theoretical regularity in the FK flow. Indeed, according to the formula (1.6) (and definition (3.1)), it is equivalent to find Δ_{n+1} s.t.

$$\frac{\gamma_{n+1}(1)}{\gamma_n(1)} = \varepsilon \qquad \left(= \frac{Z_{\beta_n+\Delta_{n+1}}}{Z_{\beta_n}}\right).$$

In other words, the sequence Δ_n is defined so that the normalizing constants $\gamma_n(1)$ increase geometrically, with the ratio $\gamma_{n+1}(1)/\gamma_n(1) = \varepsilon$. Notice that these increments Δ_n are only theoretical, and the corresponding potential functions G_n are not explicitly known.

4.2. An adaptive interacting particle model

As in the classic IPS algorithm, we approximate the measures η_n by simulating an interacting particle system $(\zeta_n)_n = (\zeta_n^1, \ldots, \zeta_n^N)_n$ of size N so that

$$\eta_n^N = \frac{1}{N} \sum_{1\le i\le N} \delta_{\zeta_n^i} \to_{N\uparrow\infty} \eta_n.$$

We start with N independent samples from η_0 and then alternate selection and mutation steps, as described in Section 1.4. As we mentioned above, in contrast to the classic IPS model, the potential function G_{n+1} arising in the selection is not known. The selection step then starts by calculating the empirical increment Δ_{n+1}^N defined by

$$\Delta_{n+1}^N := \kappa_{\eta_n^N}(\varepsilon)$$

or $\lambda_{\eta_n^N}(\Delta_{n+1}^N) = \eta_n^N(\mathrm{e}^{-\Delta_{n+1}^N\cdot V}) = \varepsilon$. As the quantity $\eta_n^N(\mathrm{e}^{-\Delta\cdot V}) = \frac{1}{N}\sum_{1\le i\le N} \mathrm{e}^{-\Delta\cdot V(\zeta_n^i)}$ is easy to calculate for all $\Delta \ge 0$, one can calculate Δ_{n+1}^N by, for example, performing a dichotomy algorithm. If we consider the stochastic potential functions

$$G_{n+1}^N = \mathrm{e}^{-\Delta_{n+1}^N\cdot V}$$

then every particle ζ_n^i evolves to a new particle $\widehat{\zeta}_n^i$ randomly chosen with the following stochastic selection transition:

$$S_{n+1,\eta_n^N}^N\big(\zeta_n^i, \mathrm{d}x\big) := G_{n+1}^N\big(\zeta_n^i\big)\delta_{\zeta_n^i}(\mathrm{d}x) + \big(1 - G_{n+1}^N\big(\zeta_n^i\big)\big)\Psi_{G_{n+1}^N}\big(\eta_n^N\big)(\mathrm{d}x).$$

In the above display formula, $\Psi_{G_{n+1}^N}(\eta_n^N)$ stands for the updated measure defined by

$$\Psi_{G_{n+1}^N}\big(\eta_n^N\big) = \sum_{j=1}^N \frac{G_{n+1}^N(\zeta_n^j)}{\sum_{k=1}^N G_{n+1}^N(\zeta_n^k)}\delta_{\zeta_n^j}.$$

Note that $V_{\min} = 0$ ensures $0 < G_{n+1}^N \le 1$.

For the mutation step, we consider two models. The implementable one consists in performing Markov transitions $M_{n+1}^N(\widehat{\zeta}_n^i, \cdot)$, defined as $M_{n+1}(\widehat{\zeta}_n^i, \cdot)$ by replacing β_{n+1} by $\beta_{n+1}^N = \beta_n^N + \Delta_{n+1}^N$. These kernels must be defined to let the associated Boltzmann–Gibbs measures stable, for instance, using some simulated annealing kernels (see Section 3.2, page 692). Thus, conditionally to the previous particle set ζ_n, the new population of particles ζ_{n+1} is sampled from distribution

$$\begin{aligned} &\mathrm{Law}\big(\zeta_{n+1}^1, \ldots, \zeta_{n+1}^N | \zeta_n^1, \ldots, \zeta_n^N\big) \\ &\quad = \big(\delta_{\zeta_n^1}.S_{n+1,\eta_n^N}^N.M_{n+1}^N\big) \otimes \cdots \otimes \big(\delta_{\zeta_n^N}.S_{n+1,\eta_n^N}^N.M_{n+1}^N\big). \end{aligned} \tag{4.2}$$

We also consider a simplified model where the mutation transition is given by the limiting transition M_{n+1} (MCMC Markov kernel associated with theoretical temperature β_{n+1} defined in Section 4.1):

$$\begin{aligned} &\mathrm{Law}\big(\zeta_{n+1}^1, \ldots, \zeta_{n+1}^N | \zeta_n^1, \ldots, \zeta_n^N\big) \\ &\quad = \big(\delta_{\zeta_n^1}.S_{n+1,\eta_n^N}^N.M_{n+1}\big) \otimes \cdots \otimes \big(\delta_{\zeta_n^N}.S_{n+1,\eta_n^N}^N.M_{n+1}\big). \end{aligned} \tag{4.3}$$

The definition of Δ_{n+1}^N is to be interpreted as the natural approximation of the theoretical relation (4.1). On the other hand, it admits a purely algorithmic interpretation. As a matter of fact, conditionally to the nth generation of particles $(\zeta_n^1, \ldots, \zeta_n^N)$, the probability for any particle ζ_n^i to be accepted, that is, not affected by the recycling mechanism, is given by $G_{n+1}^N(\zeta_n^i) = \mathrm{e}^{-\Delta_{n+1}^N \cdot V(\zeta_n^i)}$. Then the expectation of the number of accepted particles is given by $\sum_i \mathrm{e}^{-\Delta_{n+1}^N \cdot V(\zeta_n^i)}$. But it turns out that this quantity is exactly $N \times \eta_n^N(\mathrm{e}^{-\Delta_{n+1}^N \cdot V})$, which is equal to $N.\varepsilon$ by definition of Δ_{n+1}^N. Therefore, ε is an approximation of the proportion of particles which remain in place during the selection step. In other words, at each generation n, the increment Δ_{n+1}^N is chosen so that the selection step kills less than $(1-\varepsilon).N$ particles. This type of tuning parameter is very important in practice to avoid degenerate behaviors.

4.3. A perturbation analysis

This section is mainly concerned with the convergence analysis of the simplified adaptive model (4.3). The analysis of the adaptive model (4.2) is much more involved, and our approach does not apply directly to study the convergence of this model.

Despite the adaptation, the sequence η_n^N can be analyzed as a random perturbation of the theoretical sequence η_n. Let us fix n and a population state ζ_n at time n. If ϕ_{n+1}^N denotes the FK transformation associated with potential G_{n+1}^N and kernel M_{n+1}, then, by construction, the measure η_{n+1}^N is close to $\phi_{n+1}^N(\eta_n^N)$. In particular, by the Khintchine's type inequalities presented in [16] (see Lemma 7.3.3 page 223), we have

$$\forall f \in \mathcal{B}_1(E) \qquad \mathbb{E}\big(\big|\eta_{n+1}^N(f) - \phi_{n+1}^N\big(\eta_n^N\big)(f)\big|^p | \zeta_n\big)^{1/p} \leq \frac{B_p}{\sqrt{N}}, \tag{4.4}$$

with the constants B_p introduced in (0.1). A simple, but important remark about the Boltzmann–Gibbs transformations is that for any measure μ and any positive functions G and $\tilde{G}$ we have

$$\psi_{\tilde{G}}(\mu) = \psi_G\big(\psi_{\tilde{G}/G}(\mu)\big).$$

Therefore, if we take $H_{n+1}^N := \frac{G_{n+1}^N}{G_{n+1}}$, then the perturbed transformation ϕ_{n+1}^N can be written in terms of the theoretical one ϕ_{n+1} by

$$\phi_{n+1}^N = \phi_{n+1} \circ \psi_{H_{n+1}^N}. \tag{4.5}$$

If we use an inductive approach, we face the following problem. Let η be a deterministic measure (η_n in our analysis) and $\hat{\eta}$ a random measure (η_n^N in our analysis), close to η under the d_p distance (induction hypothesis). We also consider a Markov kernel M and the potential functions

$$G = \mathrm{e}^{-\kappa_\eta(\varepsilon).V}, \qquad \hat{G} = \mathrm{e}^{-\kappa_{\hat{\eta}}(\varepsilon).V}, \qquad \hat{H} = \frac{\hat{G}}{G} \tag{4.6}$$

and we let ϕ (resp., $\hat{\phi}$) be the FK transformation associated with the potential function G (resp., $\hat{G}$). The question is now: how can we estimate $d_p(\phi(\eta), \hat{\phi}(\hat{\eta}))$ in terms of $d_p(\eta, \hat{\eta})$?

To answer to this question, we propose to achieve a two-step estimation. Firstly, we estimate the distance between $\hat{\eta}$ and $\psi_{\hat{H}}(\hat{\eta})$ (Lemma 16). Secondly, we analyze the stability properties of the transformation ϕ (Lemma 17). This strategy is summarized by the following synthetic diagram:

$$\begin{array}{ccc} \eta & \xrightarrow[\phi]{} & \phi(\eta) \\ \hat{\eta} & & \\ \downarrow & & \\ \psi_{\hat{H}}(\hat{\eta}) & \xrightarrow[\phi]{} & \hat{\phi}(\hat{\eta}). \end{array}$$

Lemma 16. *Let $\eta \in \mathcal{P}(E)$, $\hat{\eta} \in \mathcal{P}_\Omega(E)$ and let G, $\hat{G}$, $\hat{H}$ be the positive functions on E defined by the equations (4.6) for some $\varepsilon > 0$. If $\eta(\{V = 0\}) = \hat{\eta}(\{V = 0\}) = 0$ (a.s.), then we have*

$$d_p\big(\psi_{\hat{H}}(\hat{\eta}), \hat{\eta}\big) \leq \frac{V_{\max} \cdot \mathrm{e}^{\kappa_\eta(\varepsilon).V_{\max}}}{\varepsilon \cdot \eta(V)} \cdot d_p(\hat{\eta}, \eta).$$

Proof. We simplify the notation and we set $x := \kappa_\eta(\varepsilon)$ and $\hat{x} := \kappa_{\hat{\eta}}(\varepsilon)$.

We start with the following observation:

$$\psi_{\hat{H}}(\hat{\eta})(f) - \hat{\eta}(f) = \frac{\hat{\eta}(\hat{H}.f)}{\hat{\eta}(\hat{H})} - \hat{\eta}(f) = \underbrace{\frac{1}{\hat{\eta}(\hat{H})}}_{A_1} \underbrace{\hat{\eta}\big[\big(\hat{H} - \hat{\eta}(\hat{H})\big).f\big]}_{A_2} \tag{4.7}$$

for any $f \in \mathcal{B}(E)$. We notice that $\hat{H} = \hat{G}/G = \mathrm{e}^{(x-\hat{x})\cdot V}$, which leads to the lower bound $\hat{\eta}(\hat{H}) = \hat{\eta}(\mathrm{e}^{(x-\hat{x})\cdot V}) \geq \hat{\eta}(\mathrm{e}^{-\hat{x}\cdot V}) = \varepsilon$. The last equality comes from the definition of $\hat{x}$. We just proved: $|A_1| \leq \varepsilon^{-1}$. On the other hand, we have $\mathrm{osc}(\hat{H}) = |\mathrm{e}^{(x-\hat{x})\cdot V_{\max}} - 1|$, so that

$$\begin{aligned} |A_2| \leq \hat{\eta}\big(\big|\hat{H} - \hat{\eta}(\hat{H})\big|\big) \cdot \|f\|_\infty &\leq \mathrm{osc}(\hat{H}) \cdot \|f\|_\infty \\ &\leq \big|\mathrm{e}^{(x-\hat{x})\cdot V_{\max}} - 1\big| \cdot \|f\|_\infty. \end{aligned} \tag{4.8}$$

The quantity $\hat{u} := (\mathrm{e}^{(x-\hat{x})\cdot V_{\max}} - 1)$ is intuitively small. Next, we provide an estimate in terms of the functions λ_η and $\lambda_{\hat{\eta}}$. Given $\omega \in \Omega$, if $x \geq \hat{x}$, then we can write

$$\underbrace{\lambda_\eta(x)}_{=\varepsilon=\lambda_{\hat{\eta}}(\hat{x})} - \lambda_{\hat{\eta}}(x) = \lambda_{\hat{\eta}}(\hat{x}) - \lambda_{\hat{\eta}}(x) = \int_{\hat{x}}^{x} -\lambda'_{\hat{\eta}}(s)\,\mathrm{d}s.$$

Furthermore, for all $\mu \in \mathcal{P}(E)$ and $s \geq 0$, we have

$$\begin{aligned} -\lambda'_\mu(s) = \mu\big(V \cdot \mathrm{e}^{-sV}\big) &\geq \mu\big(V \cdot \mathrm{e}^{-sV_{\max}}\big) \\ &\geq \mu(V) \cdot \mathrm{e}^{-sV_{\max}}. \end{aligned}$$

Then we have

$$\begin{aligned} \lambda_\eta(x) - \lambda_{\hat{\eta}}(x) &\geq \hat{\eta}(V)\frac{-1}{V_{\max}}\big[\mathrm{e}^{-sV_{\max}}\big]_{\hat{x}}^{x} \\ &= \hat{\eta}(V)\frac{\mathrm{e}^{-xV_{\max}}}{V_{\max}}\big(\mathrm{e}^{(x-\hat{x})V_{\max}} - 1\big) \\ &= \hat{\eta}(V)\frac{\mathrm{e}^{-xV_{\max}}}{V_{\max}}\hat{u}. \end{aligned}$$

By symmetry, we have

$$x \leq \hat{x} \quad \Longrightarrow \quad \lambda_{\hat{\eta}}(x) - \lambda_\eta(x) \geq \hat{\eta}(V)\frac{\mathrm{e}^{-xV_{\max}}}{V_{\max}}(-\hat{u}).$$

This yields the almost sure upper bound

$$|\hat{u}| \cdot \hat{\eta}(V)\frac{\mathrm{e}^{-xV_{\max}}}{V_{\max}} \leq \big|\lambda_\eta(x) - \lambda_{\hat{\eta}}(x)\big|.$$

Using the decomposition $\hat{\eta}(V) = \eta(V) + (\hat{\eta}(V) - \eta(V))$, by simple manipulation, we prove that

$$|\hat{u}| \leq \frac{V_{\max}\mathrm{e}^{xV_{\max}}}{\eta(V)}\underbrace{\big|\lambda_\eta(x) - \lambda_{\hat{\eta}}(x)\big|}_{A_3} + \underbrace{\frac{|\hat{\eta}(V) - \eta(V)|}{\eta(V)}}_{A_4} \cdot \underbrace{|\hat{u}|}_{A_5}.$$

Considering the L^p norm of the right-hand side of this inequality, one can check that

- $A_3 = (\eta - \hat{\eta})(\mathrm{e}^{-x\cdot V})$ so, as $\operatorname{osc}(\mathrm{e}^{-x\cdot V}) \leq 1$, $\|A_3\|_p \leq d_p(\hat{\eta}, \eta)$;
- as $\operatorname{osc}(V) = V_{\max}$, $\|A_4\|_p \leq \frac{V_{\max}}{\eta(V)} \cdot d_p(\hat{\eta}, \eta)$;
- $A_5 = |\hat{u}| = \mathrm{e}^{xV_{\max}} |\mathrm{e}^{-\hat{x}V_{\max}} - \mathrm{e}^{-xV_{\max}}| \leq \mathrm{e}^{xV_{\max}}$.

Making the appropriate substitutions, we have

$$\|\hat{u}\|_p \leq \frac{2V_{\max}\mathrm{e}^{x\cdot V_{\max}}}{\eta(V)} \cdot d_p(\hat{\eta}, \eta).$$

Combining this result with (4.7) and (4.8), we check that

$$\left\|\psi_{\hat{H}}(\hat{\eta})(f) - \hat{\eta}(f)\right\|_p \leq \frac{2V_{\max}\mathrm{e}^{xV_{\max}}}{\varepsilon \cdot \eta(V)} \cdot d_p(\hat{\eta}, \eta) \cdot \|f\|_\infty.$$

We finally go from $\|f\|_\infty$ to $\frac{\operatorname{osc}(f)}{2}$ by noticing that $\psi_{\hat{H}}(\hat{\eta})(f) - \hat{\eta}(f)$ is equal to 0 for any constant function f, and by considering the above inequality taken for $\tilde{f} = f - \frac{f_{\max}+f_{\min}}{2}$, which satisfies $\|\tilde{f}\|_\infty = \frac{\operatorname{osc}(f)}{2}$.

This ends the proof of the lemma. □

Lemma 17. *Let $\eta \in \mathcal{P}(E)$, $\hat{\eta} \in \mathcal{P}_\Omega(E)$, and let ϕ be a FK transformation associated with a positive function G and a Markov kernel M. If we set $g := \sup_{x,y\in E} G(x)/G(y)$ and $b := \beta(M)$, then we have*

$$d_p\big(\phi(\hat{\eta}), \phi(\eta)\big) \leq g \cdot b \cdot d_p(\hat{\eta}, \eta).$$

Proof. Let us fix $f \in \mathcal{B}(E)$. We have

$$\begin{aligned}\phi(\hat{\eta})(f) - \phi(\eta)(f) &= \frac{\hat{\eta}(G \times M.f)}{\hat{\eta}(G)} - \phi(\eta)(f) \\ &= \frac{\hat{\eta}(G \times [M.(f - \phi(\eta)(f))])}{\hat{\eta}(G)}.\end{aligned}$$

Let $\tilde{f} = M.(f - \phi(\eta)(f))$. By property (1.1), $\tilde{f}$ satisfies $\operatorname{osc}(\tilde{f}) = \operatorname{osc}(M.f) \leq b \cdot \operatorname{osc}(f)$.

Additionally we have $\eta(G \times \tilde{f}) = \eta(G \times M.f) - \eta(G \times \frac{\eta(G\times M.f)}{\eta(G)}) = 0$. So we obtain

$$\phi(\hat{\eta})(f) - \phi(\eta)(f) = \frac{\hat{\eta}(G \times \tilde{f})}{\hat{\eta}(G)} - \underbrace{\frac{\eta(G \times \tilde{f})}{\hat{\eta}(G)}}_{=0} = \frac{1}{\hat{\eta}(G)}(\hat{\eta} - \eta)(G \times \tilde{f})$$

$$= \frac{G_{\max}}{\hat{\eta}(G)}(\hat{\eta} - \eta)\left(\frac{G}{G_{\max}} \times \tilde{f}\right).$$

Firstly, we notice that $|\frac{G_{\max}}{\hat{\eta}(G)}| \leq \frac{G_{\max}}{G_{\min}} \leq g$. On the other hand, we notice that $\tilde{f}$ can be rewritten as

$$\tilde{f} = M.\big(f - \psi_G(\eta)(M.f)\big) = (M.f) - \psi_G(\eta)(M.f).$$

It follows that $\tilde{f}_{\max} \geq 0$ and $\tilde{f}_{\min} \leq 0$. Under these conditions, $\operatorname{osc}(\frac{G}{G_{\max}} \times \tilde{f}) \leq \operatorname{osc}(\tilde{f}) \leq b \cdot \operatorname{osc}(f)$. We conclude that

$$\begin{aligned}\left(\mathbb{E}\left|\frac{G_{\max}}{\hat{\eta}(G)}(\hat{\eta}-\eta)\left(\frac{G}{G_{\max}} \times \tilde{f}\right)\right|^p\right)^{1/p} &\leq g \cdot \mathbb{E}\left[\left|(\hat{\eta}-\eta)\left(\frac{G}{G_{\max}} \times \tilde{f}\right)\right|^p\right]^{1/p} \\ &\leq g \cdot b \cdot d_p(\hat{\eta}, \eta) \cdot \operatorname{osc}(f).\end{aligned}$$

This ends the proof of the lemma. □

Remark 18. *Lemma 17 holds in the case where η is also a random measure (we now note $\tilde{\eta} \in \mathcal{P}_\Omega(E)$) if one can find a σ-algebra $\mathcal{F}$ such that*:

1. *$\tilde{\eta}$ is $\mathcal{F}$-measurable*;
2. *the ($\mathcal{F}$-measurable) random variable*

$$d_p^{\mathcal{F}}(\hat{\eta}, \tilde{\eta}) := \sup_{\tilde{f} \in \mathcal{O}_1^{\Omega,\mathcal{F}}(E)} \mathbb{E}\left[\left|\hat{\eta}(\tilde{f}) - \tilde{\eta}(\tilde{f})\right|^p | \mathcal{F}\right]^{1/p}$$

is uniformly bounded on Ω. In the above definition, $\mathcal{O}_1^{\Omega,\mathcal{F}}(E)$ denotes the set of random, $\mathcal{F}$-measurable functions $\tilde{f} : E \to \mathbb{R}$ satisfying $\operatorname{osc}(\tilde{f}) \leq 1$ a.s.

More precisely, under these conditions, we have

$$d_p\big(\phi(\hat{\eta}), \phi(\tilde{\eta})\big) \leq g \cdot b \cdot \left\| d_p^{\mathcal{F}}(\hat{\eta}, \tilde{\eta}) \right\|_\infty.$$

Proof. We fix $f \in \mathcal{B}(E)$ and use the same line of arguments as in the proof of Lemma 17. Since the function $\tilde{f} = M.(f - \phi(\eta)(f))$ is $\mathcal{F}$-measurable and satisfies

$$\operatorname{osc}\left(\frac{G}{G_{\max}} \times \tilde{f}\right) \leq \operatorname{osc}(\tilde{f}) \leq b \cdot \operatorname{osc}(f) \qquad \text{a.s.},$$

we have

$$\begin{aligned}&\left(\mathbb{E}\left|\frac{G_{\max}}{\hat{\eta}(G)}(\hat{\eta}-\tilde{\eta})\left(\frac{G}{G_{\max}} \times \tilde{f}\right)\right|^p\right)^{1/p} \\ &\leq g \cdot \mathbb{E}\left(\mathbb{E}\left[\left|(\hat{\eta}-\tilde{\eta})\left(\frac{G}{G_{\max}} \times \tilde{f}\right)\right|^p \middle| \mathcal{F}\right]\right)^{1/p} \\ &\leq g \cdot b \cdot \operatorname{osc}(f) \cdot \mathbb{E}\left(\mathbb{E}\left[\left|(\hat{\eta}-\tilde{\eta})\left(\frac{G/G_{\max} \times \tilde{f}}{b \cdot \operatorname{osc}(f)}\right)\right|^p \middle| \mathcal{F}\right]\right)^{1/p} \\ &\leq g \cdot b \cdot \operatorname{osc}(f) \cdot \left\| d_p^{\mathcal{F}}(\hat{\eta}, \tilde{\eta}) \right\|_\infty.\end{aligned}$$

This ends the proof. □

4.4. Nonasymptotic convergence results

This section is mainly concerned with the proof of Theorem 3 stated on page 675. We also deduce some concentration inequalities of the ISA adaptive model. In this section, (η_n) denotes the sequence of theoretical measures defined in Section 4.1, and (η_n^N) denotes the sequence of empirical measures associated with the particle system described in (4.3), page 698. We start with the proof of Theorem 3.

Proof of Theorem 3. We fix $p \geq 1$ and we let $\tilde{e}_n = \sum_{k=0}^{n} \prod_{i=k+1}^{n} b_i g_i (1 + c_i)$. We notice that this sequence can also be defined with the recurrence relation $\tilde{e}_{n+1} = 1 + g_{n+1} b_{n+1} (1 + c_{n+1}) \cdot \tilde{e}_n$ starting at $\tilde{e}_0 = 1$. We also consider the following parameter:

$$e_n := \frac{2\sqrt{N}}{B_p} \cdot \sup_{f \in \mathcal{O}_1(E)} \left\| \eta_n^N(f) - \eta_n(f) \right\|_p.$$

We use an inductive proof to check that the proposition $\mathbf{IH}(n) = \{e_n \leq \tilde{e}_n\}$ is met at any rank n. As η_0^N is obtained with N independent samples from η_0, $\mathbf{IH}(0)$ is given by the Khintchine's inequality (see the relation (A.1), page 706 for the factor $1/2$). Now suppose that $\mathbf{IH}(n)$ is satisfied.

According to the identity (4.5), we can write the following decomposition:

$$\begin{aligned} \eta_{n+1}^N - \eta_{n+1} &= \left(\eta_{n+1}^N - \phi_{n+1}^N\left(\eta_n^N\right)\right) + \left(\phi_{n+1}^N\left(\eta_n^N\right) - \eta_{n+1}\right) \\ &= \underbrace{\left(\eta_{n+1}^N - \phi_{n+1}^N\left(\eta_n^N\right)\right)}_{A_1} + \underbrace{\left(\phi_{n+1}\left(\psi_{H_{n+1}^N}\left(\eta_n^N\right)\right) - \phi_{n+1}(\eta_n)\right)}_{A_2}. \end{aligned} \tag{4.9}$$

Using (4.4), page 698, we know that for all function $f \in \mathcal{O}_1(E)$, we have

$$\begin{aligned} \left\| A_1(f) \right\|_p &= \mathbb{E}\left(\mathbb{E}\left[\left|\eta_{n+1}^N(f) - \phi_{n+1}^N\left(\eta_n^N\right)(f)\right|^p | \zeta_n\right]\right)^{1/p} \\ &\leq \frac{B_p}{2\sqrt{N}} \end{aligned}$$

(see the relation (A.1), page 706 for the factor $1/2$), so $\frac{2\sqrt{N}}{B_p} \|A_1(f)\|_p \leq 1$.

To estimate A_2, we start by decomposing $(\psi_{H_{n+1}^N}(\eta_n^N) - \eta_n)$ in this way:

$$\psi_{H_{n+1}^N}\left(\eta_n^N\right) - \eta_n = \underbrace{\left(\psi_{H_{n+1}^N}\left(\eta_n^N\right) - \eta_n^N\right)}_{Q_1} + \underbrace{\left(\eta_n^N - \eta_n\right)}_{Q_2}.$$

By the induction hypothesis, we have $\frac{2\sqrt{N}}{B_p} \cdot \sup_{f \in \mathcal{O}_1(E)} \|Q_2(f)\|_p \leq \tilde{e}_n$. Therefore, by Lemma 16, we find that

$$\frac{2\sqrt{N}}{B_p} \cdot \sup_{f \in \mathcal{O}_1(E)} \left\| Q_1(f) \right\|_p \leq c_{n+1} \cdot \tilde{e}_n.$$

Thus, the measures $\psi_{H_{n+1}^N}(\eta_n^N)$ and η_n satisfy

$$\frac{2\sqrt{N}}{B_p} \cdot \sup_{f\in\mathcal{O}_1(E)} \left\|\psi_{H_{n+1}^N}\left(\eta_n^N\right)(f) - \eta_n(f)\right\|_p \leq (1+c_{n+1})\cdot \tilde{e}_n.$$

Applying Lemma 17, we also have

$$\frac{2\sqrt{N}}{B_p} \cdot \sup_{f\in\mathcal{O}_1(E)} \left\|A_2(f)\right\|_p \leq g_{n+1}b_{n+1}(1+c_{n+1})\cdot \tilde{e}_n.$$

Back to (4.9), we conclude that $e_{n+1} \leq 1 + g_{n+1}b_{n+1}(1+c_{n+1})\cdot \tilde{e}_n = \tilde{e}_{n+1}$.
This ends the proof of the theorem. □

We are now in position to obtain a sufficient condition for uniform concentration and L^p-mean error bounds w.r.t. time for the simplified adaptive model discussed in Sections 4.2 and 4.3.

Corollary 19. *If the condition $b_n g_n(1+c_n) \leq a$ is satisfied for some $a < 1$ and any n, then we have the uniform error bounds*

$$d_p\left(\eta_n^N, \eta_n\right) \leq \frac{B_p}{2(1-a)\sqrt{N}} \tag{4.10}$$

for any p, with the constants B_p introduced in (0.1). In addition, for any $f \in \mathcal{B}_1(E)$, we have the following concentration inequalities:

$$\forall s \geq 0 \qquad \mathbb{P}\left(\left|\eta_n^N(f) - \eta_n(f)\right| \geq s\right) \leq r_1(\sqrt{N}.s)\mathrm{e}^{-r_2 N s^2} \tag{4.11}$$

and

$$\forall y \geq 0 \qquad \mathbb{P}\left(\left|\eta_n^N(f) - \eta_n(f)\right| \geq \frac{r(1+\sqrt{2y})}{\sqrt{N}}\right) \leq \mathrm{e}^{-y}, \tag{4.12}$$

with the parameters

$$\begin{cases} r_1 = \mathrm{e}^{1/2}(1-a), \\ r_2 = \dfrac{1}{2}(1-a)^2, \\ r = \dfrac{1}{1-a}. \end{cases}$$

Proof. The inequality (4.10) is a direct consequence of Theorem 3 (see the relation (A.1), page 706 for the factor 1/2).

Let us fix n, $f \in \mathcal{B}_1(E)$ and set

$$X := \left|\eta_n^N(f) - \eta_n(f)\right| \quad \text{and} \quad \epsilon_N := \frac{1}{(1-a)\sqrt{N}}.$$

In this notation, we have $\|X\|_p \leq B_p \cdot \epsilon_N$ for any $p \geq 1$. Let us fix $s \geq 0$. By Markov inequality, for all $t \geq 0$ we have

$$\mathbb{P}(X \geq s) = \mathbb{P}\left(\mathrm{e}^{tX} \geq \mathrm{e}^{ts}\right) \leq \mathrm{e}^{-st}\mathbb{E}\left(\mathrm{e}^{tX}\right). \tag{4.13}$$

Using the formulae (0.1), page 675, we estimate the Laplace transform $\mathbb{E}(\mathrm{e}^{tX})$.

$$\begin{aligned}\mathbb{E}\left(\mathrm{e}^{tX}\right) &= \sum_{p\geq 0}\mathbb{E}\left(\frac{t^p.X^p}{p!}\right)\\ &\leq \sum_{p\geq 0}\frac{t^{2p}\epsilon_N^{2p}}{(2p)!}\frac{(2p)!}{2^p.p!} + \sum_{p\geq 0}\frac{t^{2p+1}\epsilon_N^{2p+1}}{(2p+1)!}\frac{(2p+1)!}{2^p.p!\sqrt{2p+1}}\\ &\leq (1+t\epsilon_N)\mathrm{e}^{t^2\epsilon_N^2/2}.\end{aligned}$$

Taking the inequality (4.13) with $t = \frac{1}{\epsilon_N}(\frac{s}{\epsilon_N} - 1)$, we obtain

$$\mathbb{P}(X \geq s) \leq \frac{s}{\epsilon_N}\mathrm{e}^{-(1/2)[(s/\epsilon_N)^2-1]},$$

which is equivalent to the first concentration inequality (4.11).

For the second one, let $u = \frac{s}{\epsilon_N}$. Since $\log(u) \leq u - 1$ we have

$$\mathbb{P}(X \geq s) = \mathbb{P}(X \geq \epsilon_N \cdot u) \leq u \cdot \mathrm{e}^{-(u^2/2-1/2)} = \mathrm{e}^{-(u^2/2-\log u-1/2)} \leq \mathrm{e}^{-(u^2/2-u+1/2)}.$$

Let T the function defined by

$$\begin{aligned}&[1,+\infty) \longrightarrow [0,+\infty),\\ &T = v \mapsto \frac{v^2}{2} - v + \frac{1}{2}\end{aligned}$$

T is bijective, and its inverse function is $T^{-1} : y \mapsto 1 + \sqrt{2y}$. Thus, we have

$$\forall y \geq 0 \qquad \mathbb{P}\left(X \geq \epsilon_N \cdot T^{-1}(y)\right) \leq \mathrm{e}^{-y},$$

which is equivalent to (4.12).

This ends the proof of the corollary. □

Appendix

A.1. Proof of Lemma 6, page 677

For all $\eta \in \mathcal{P}(E)$, let S_η be the Markov kernel defined by

$$S_\eta(x, \mathrm{d}y) = \frac{G(x)}{G_{\max}}\delta_x(\mathrm{d}y) + \left(1 - \frac{G(x)}{G_{\max}}\right)\Psi_G(\eta)(\mathrm{d}y).$$

This kernel satisfies $\eta.S_\eta = \Psi_G(\eta)$, since for all $f \in \mathcal{B}(E)$ we have

$$\begin{aligned}\eta.S_\eta(f) &= \frac{\eta(G\times f)}{G_{\max}} + \left(1-\frac{\eta(G)}{G_{\max}}\right).\underbrace{\Psi_G(\eta)(f)}_{=\eta(G\times f)/\eta(G)}\\ &= \frac{\eta(G\times f)}{G_{\max}} + \frac{\eta(G\times f)}{\eta(G)} - \frac{\eta(G).\eta(G\times f)}{G_{\max}.\eta(G)}\\ &= \Psi_G(\eta)(f).\end{aligned}$$

We use the decomposition

$$\Psi_G(\mu)-\Psi_G(\nu) = \mu.S_\mu - \nu.S_\nu = (\mu-\nu).S_\mu + \nu.(S_\mu - S_\nu).$$

On the one hand, $\|(\mu-\nu).S_\mu\|_{\mathrm{tv}} \le \|\mu-\nu\|_{\mathrm{tv}}.\beta(S_\mu) \le \|\mu-\nu\|_{\mathrm{tv}}$.

On the other hand, for all nonnegative function $f \in \mathcal{B}_1(E)$,

$$(S_\mu - S_\nu)(f)(x) = \underbrace{\left(1-\frac{G(x)}{G_{\max}}\right)}_{\le 1-1/g}\underbrace{\left[\Psi_G(\mu)(f)-\Psi_G(\nu)(f)\right]}_{|-|\le \|\Psi_G(\mu)-\Psi_G(\nu)\|_{\mathrm{tv}}}$$

so $|\nu.(S_\mu-S_\nu)(f)| \le (1-\frac{1}{g})\|\Psi_G(\mu)-\Psi_G(\nu)\|_{\mathrm{tv}}$.

Thus,

$$\begin{aligned}&\left\|\Psi_G(\mu)-\Psi_G(\nu)\right\|_{\mathrm{tv}} \le \|\mu-\nu\|_{\mathrm{tv}} + \left(1-\frac{1}{g}\right)\left\|\Psi_G(\mu)-\Psi_G(\nu)\right\|_{\mathrm{tv}}\\ \Rightarrow\quad &\left\|\Psi_G(\mu)-\Psi_G(\nu)\right\|_{\mathrm{tv}} \le g.\|\mu-\nu\|_{\mathrm{tv}}.\end{aligned}$$

This ends the proof of the lemma.

A.2. Proof of the L^p-mean error bound (1.10) (page 680)

We fix $p \ge 1$. Let us start with this remark: for all $\hat\mu, \hat\nu \in \mathcal{P}_\Omega(E)$, $\hat\mu(f)-\hat\nu(f) = 0$ for all constant function f, so we have

$$\begin{aligned}\sup_{\|f\|_\infty\le 1}\left\|\hat\mu(f)-\hat\nu(f)\right\|_p &= 2\sup_{\|f\|_\infty\le 1/2}\left\|\hat\mu(f)-\hat\nu(f)\right\|_p\\ &= 2\sup_{\mathrm{osc}(f)=1}\left\|\hat\mu(f)-\hat\nu(f)\right\|_p \qquad (A.1)\\ &= 2\cdot d_p(\hat\mu,\hat\nu).\end{aligned}$$

In the context of the interacting particle system described in Section 1.4, page 679, we use the decomposition

$$\eta_n^N - \eta_n = \sum_{k=0}^n \phi_{k,n}\left(\eta_k^N\right) - \phi_{k,n}\left(\phi_k\left(\eta_{k-1}^N\right)\right) \qquad (A.2)$$

(with the convention $\phi_0(\eta_{-1}^N) = \eta_0$). For all $k \in \{1, 2, \ldots, n\}$, let $\mathcal{F}_{k-1}$ be the σ-algebra generated by the variable ζ_{k-1} (we set $\mathcal{F}_0 = \{\varnothing, E\}$). Remark 18, page 702, implies

$$d_p\big(\phi_{k,n}\big(\eta_k^N\big), \phi_{k,n}\big(\phi_k\big(\eta_{k-1}^N\big)\big)\big) \leq g_{k,n} \cdot b_{k,n} \cdot \big\| d_p^{\mathcal{F}_{k-1}}\big(\eta_k^N, \phi_k\big(\eta_{k-1}^N\big)\big)\big\|_\infty, \tag{A.3}$$

where $d_p^{\mathcal{F}_{k-1}}(\eta_k^N, \phi_k(\eta_{k-1}^N))$ is the random variable defined by

$$d_p^{\mathcal{F}_{k-1}}\big(\eta_k^N, \phi_k\big(\eta_{k-1}^N\big)\big) := \sup_{\tilde{f} \in \mathcal{O}_1^{\Omega, \mathcal{F}_{k-1}}(E)} \mathbb{E}\big[\big|\eta_k^N(\tilde{f}) - \phi_k\big(\eta_{k-1}^N\big)(\tilde{f})\big|^p | \mathcal{F}_{k-1}\big]^{1/p}.$$

In the above definition, $\mathcal{O}_1^{\Omega, \mathcal{F}_{k-1}}(E)$ denotes the set of random, $\mathcal{F}_{k-1}$-measurable functions $\tilde{f} : E \to \mathbb{R}$ satisfying $\operatorname{osc}(\tilde{f}) \leq 1$ a.s.

Given the conditioning relation

$$\begin{aligned}&\operatorname{Law}\big(\xi_k^1, \ldots, \xi_k^N | \xi_{k-1}^1, \ldots, \zeta_{k-1}^N\big)\\ &\quad = (\delta_{\zeta_{k-1}^1} . S_{k, \eta_{k-1}^N} . M_k) \otimes \cdots \otimes (\delta_{\zeta_{k-1}^N} . S_{k, \eta_{k-1}^N} . M_k)\end{aligned}$$

and the Khintchine's type inequalities presented in [16], we have for all random, $\mathcal{F}_{k-1}$-measurable function $\tilde{f}$ satisfying $\|\tilde{f}\|_\infty \leq 1$ a.s.:

$$\mathbb{E}\big(\big|\eta_k^N(\tilde{f}) - \phi_k\big(\eta_{k-1}^N\big)(\tilde{f})\big|^p | \mathcal{F}_{k-1}\big)^{1/p} \leq \frac{B_p}{\sqrt{N}},$$

with the constants B_p introduced in (0.1), page 675. This is equivalent to

$$d_p^{\mathcal{F}_{k-1}}\big(\eta_k^N, \phi_k\big(\eta_{k-1}^N\big)\big) \leq \frac{B_p}{2\sqrt{N}}.$$

Finally, we combine this result with (A.3) and the decomposition (A.2) to obtain

$$d_p\big(\eta_n^N, \eta_n\big) \leq \frac{B_p}{2\sqrt{N}} \sum_{k=0}^n g_{k,n} b_{k,n}.$$

This ends the proof.

References

[1] Assaraf, R., Caffarel, M. and Khelif, A. (2000). Diffusion Monte Carlo with a fixed number of walkers. *Phys. Rev. E* (3).

[2] Baker, J.E. (1985). Adaptive selection methods for genetic algorithms. *ICGA*1 101–111.

[3] Bartoli, N. and Del Moral, P. (2001). *Simulation & Algorithmes Stochastiques*. Toulouse, France: Cépaduès.

[4] Bertrand, C., Hamada, Y. and Kado, H. (2001). MRI prior computation and parallel tempering algorithm: A probabilistic resolution of the MEG/EEG inverse problem. *Brain Topography* **14** 57–68.

[5] Caffarel, M. and Assaraf, R. (2000). A pedagogical introduction to quantum Monte-Carlo. In *Mathematical Models and Methods for Ab Initio Quantum Chemistry*. *Lecture Notes in Chemistry* **74** 45–73. Berlin: Springer. MR1855574
[6] Cappé, O., Moulines, E. and Rydén, T. (2005). *Inference in Hidden Markov Models*. *Springer Series in Statistics*. New York: Springer. MR2159833
[7] Ceperley, D.M. (1995). Path integrals in the theory of condensed helium. *Rev. Modern Phys*. **67** 279–355.
[8] Cerf, R. (1994). Une théorie asymptotique des algorithmes génétiques. Ph.D. thesis, Université de Montpellier 2, France.
[9] Cérou, F., Del Moral, P., Furon, T. and Guyader, A. (2012). Sequential Monte Carlo for rare event estimation. *Stat*. *Comput*. **22** 795–808. MR2909622
[10] Cérou, F., Del Moral, P. and Guyader, A. (2011). A nonasymptotic theorem for unnormalized Feynman–Kac particle models. *Ann*. *Inst*. *Henri Poincaré Probab*. *Stat*. **47** 629–649. MR2841068
[11] Chopin, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann*. *Statist*. **32** 2385–2411. MR2153989
[12] Ciccotti, G. and Hoover, W.G., eds. (1986). *Molecular-Dynamics Simulation of Statistical-Mechanical Systems*. *Proceedings of the International School of Physics "Enrico Fermi"*, *XCVII*. Amsterdam: North-Holland. MR0960486
[13] Clapp, T. (2000). Statistical Methods in the Processing of Communications Data. Ph.D. thesis, Cambridge Univ., Engineering Dept.
[14] Dawson, D.A. and Del Moral, P. (2005). Large deviations for interacting processes in the strong topology. In *Statistical Modeling and Analysis for Complex Data Problems* (P. Duchesne and B. Rémillard, eds.). *GERAD* 25*th Anniv*. *Ser*. **1** 179–208. New York: Springer. MR2189537
[15] Del Moral, P. (1996). Nonlinear filtering: Interacting particle solution. *Markov Process*. *Related Fields* **2** 555–579. MR1431187
[16] Del Moral, P. (2004). *Feynman–Kac Formulae*: *Genealogical and Interacting Particle Systems with Applications*. *Probability and Its Applications* (*New York*). New York: Springer. MR2044973
[17] Del Moral, P., Doucet, A. and Jasra, A. (2006). Sequential Monte Carlo samplers. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **68** 411–436. MR2278333
[18] Del Moral, P., Doucet, A. and Jasra, A. (2012). On adaptive resampling strategies for sequential Monte Carlo methods. *Bernoulli* **18** 252–278. MR2888706
[19] Del Moral, P. and Guionnet, A. (1998). Large deviations for interacting particle systems: Applications to non-linear filtering. *Stochastic Process*. *Appl*. **78** 69–95. MR1653296
[20] Del Moral, P. and Guionnet, A. (1999). Central limit theorem for nonlinear filtering and interacting particle systems. *Ann*. *Appl*. *Probab*. **9** 275–297. MR1687359
[21] Del Moral, P. and Guionnet, A. (2001). On the stability of interacting processes with applications to filtering and genetic algorithms. *Ann*. *Inst*. *Henri Poincaré Probab*. *Stat*. **37** 155–194. MR1819122
[22] Del Moral, P., Hu, P. and Wu, L. (2012). On the concentration properties of interacting particle processes. *Foundations and Trends in Machine Learning* **3** 225–389.
[23] Del Moral, P. and Ledoux, M. (2000). Convergence of empirical processes for interacting particle systems with applications to nonlinear filtering. *J*. *Theoret*. *Probab*. **13** 225–257. MR1744985
[24] Del Moral, P. and Miclo, L. (1999). On the convergence and applications of generalized simulated annealing. *SIAM J*. *Control Optim*. **37** 1222–1250. MR1691939
[25] 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
[26] Del Moral, P. and Rio, E. (2011). Concentration inequalities for mean field particle models. *Ann*. *Appl*. *Probab*. **21** 1017–1052. MR2830611

[27] Deutscher, J., Blake, A. and Reid, I. (2000). Articulated body motion capture by annealed particle filtering. *IEEE Conference on Computer Vision and Pattern Recognition* **2** 126–133.
[28] Doucet, A., de Freitas, N. and Gordon, N., eds. (2001). *Sequential Monte Carlo Methods in Practice*. *Statistics for Engineering and Information Science*. New York: Springer. MR1847783
[29] Frankel, D. and Smit, B. (1996). *Understanding Molecular Simulation*. New York: Academic Press.
[30] Giraud, F., Minvielle, P., Sancandi, M. and Del Moral, P. (2012) Rao-blackwellised interacting Markov chain Monte Carlo for electromagnetic scattering inversion. Available at arXiv:stat.AP/1209.4006v2.
[31] Goldberg, D.E. (1989). *Genetic Algorithms in Search*, *Optimization*, *and Machine Learning*. Reading, MA: Addison-Wesley.
[32] Hajek, B. (1988). Cooling schedules for optimal annealing. *Math. Oper. Res.* **13** 311–329. MR0942621
[33] Hetherington, J.H. (1984). Observations on the statistical iteration of matrices. *Phys. Rev. A* (3) **30** 2713–2719. MR0763130
[34] Jasra, A., Stephens, D.A., Doucet, A. and Tsagaris, T. (2011). Inference for Lévy-driven stochastic volatility models via adaptive sequential Monte Carlo. *Scand. J. Stat.* **38** 1–22. MR2760137
[35] Künsch, H.R. (2005). Recursive Monte Carlo filters: Algorithms and theoretical analysis. *Ann. Statist.* **33** 1983–2021. MR2211077
[36] Le Riche, R., Schoenauer, M. and Sebag, M. (2007). Un état des lieux de l'optimisation évolutionnaire et de ses implications en sciences pour l'ingénieur, chapter in modélisation numérique:. In *Défis et Perspectives*, *Vol.* 2, *Traité Mécanique et Ingénierie des Matériaux* (P. Breitkopf and C. Knopf-Lenoir, eds.) 187–259. Hermes.
[37] Minvielle, P., Doucet, A., Marrs, A. and Maskell, S. (2010). A Bayesian approach to joint tracking and identification of geometric shapes in video sequences. *Image and Vision Computing*. **28** 111–123.
[38] Neal, R.M. (2001). Annealed importance sampling. *Stat. Comput.* **11** 125–139. MR1837132
[39] Schäfer, C. and Chopin, N. (2013). Sequential Monte Carlo on large binary sampling spaces. *Stat. Comput.* **23** 163. DOI:10.1007/s11222-011-9299-z.
[40] Schweizer, N. (2012). Non-asymptotic error bounds for sequential MCMC and stability of Feynman–Kac propagators. Working paper, University of Bonn.
[41] Whiteley, N. (2012). Sequential Monte Carlo samplers: Error bounds and insensitivity to initial conditions. *Stoch. Anal. Appl.* **30** 774–798. MR2966098

Received September 2012 and revised September 2014