On the convergence of Quantum and Sequential Monte Carlo methods

François Giraud, Pierre Del Moral

Abstract 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 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. Special attention is devoted to the particular case of Boltzmann-Gibbs measures' sampling. In this context, we design an explicit way of tuning the temperature schedule with the number of Markov Chain Monte Carlo iterations.

Introduction

Sequential and Quantum Monte Carlo methods (*abbreviate SMC and QMC*) are stochastic algorithms to sample from complex high-dimensional probability distributions. These stochastic simulation techniques are of current use in numerical physics [21, 2, 1] to compute ground state energies. They are also used in statistics, signal processing and information sciences [4, 12, 11, 14] to compute posterior distributions of partially observed signal or unknown parameters. In evolutionary computing literature, these Monte Carlo methods are used as natural population search algorithms for solving optimization problems. From the pure mathematical

François Giraud
CEA-CESTA, 33114 Le Barp, France; e-mail: francois.giraud@ens-cachan.org

Pierre Del Moral
INRIA Bordeaux Sud-Ouest, Domaine Universitaire, 351, cours de la Liberation, 33405 Talence Cedex, France & CMAP, Applied Mathematical Department, Ecole Polytechnique Paris e-mail: pierre.del-moral@inria.fr

viewpoint, these advanced Monte Carlo methods coincide with mean field particle interpretations of Feynman-Kac (*abbreviate FK*) models. For a thorough discussion on FK models we refer the reader to the monograph [10], and references therein. The principle (see also [12] 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 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 of each other (mutation step).

This method is often used for solving sequential problems, such as filtering (see e.g., [11]). 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. Large population sizes allow to cover several modes simultaneously. This is an advantage compared to standard MCMC methods. These sequential samplers have been used with success in several application domains, including rare events simulation (see [5]), stochastic optimization and Boltzmann-Gibbs measures sampling ([12]).

Up to now, SMC and QMC algorithms have been mostly analyzed using asymptotic (i.e. when number of particles N tends to infinity) techniques, notably through central limit theorems and large deviation principles (see for instance [9, 15], [17, 14, 16], [23], [7], [11], [4] and [10] for an overview). Our work relates to less studied non-asymptotic problems, and follows those based on Markov kernels' mixing properties (see for instance [6], [17] and [10]). We emphasize that other independent approaches, such as Whiteley's ([27]) or Schweizer's ([26]), based on, e.g., drift conditions, hyper-boundedness, or spectral gaps, lead to convergence results that may also apply to non-compact state spaces. To our knowledge, these techniques are restricted to non-asymptotic variance theorems and they cannot be used to derive uniform and exponential concentration inequalities.

The present work consists in estimating 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 non-asymptotic theorems on L^p error bounds ([17]) and some useful concentration inequalities ([18]). Another contribution is to provide parameter tuning strategies that allow to deduce some useful uniform concentration inequalities w.r.t. the time parameter. These results also apply to non-homogeneous FK models associated with cooling temperature parameters.

In a preliminary section, we recall a few essential notions related to Dobrushin coefficients or FK semigroups, as well as a couple of important non-asymptotic results we use in the further development of the article. The second part is concerned with the semigroup stability analysis of these models. We also provide a couple of uniform L^p-deviations and concentration estimates. We end the article with an application of these results to Boltzmann-Gibbs models associated with a decreasing temperature schedule. In this context, SMC and QMC algorithms can be interpreted as a sequence of interacting simulated annealing (*abbreviate ISA*) algorithms. The detailed proofs of the results presented in this article will be presented in a forthcoming publication dedicated to adaptive particle algorithms (see [20] for a preliminary version).

1 Preliminaries

1.1 Notations

Let (E,r) be a complete, separable metric space and let $\mathscr{E}$ be the σ-algebra of Borel subsets of E. Denote by $\mathscr{P}(E)$ the space of probability measures on E. Let $\mathscr{B}(E)$ be the space of bounded, measurable, real-valued functions on E. Let $\mathscr{B}_1(E) \subset \mathscr{B}(E)$ be the subset of all bounded by 1 functions.

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

$$K_1.K_2(x,A) = \int_E K_1(x,dy)K_2(y,A),$$

$K.f$ denotes the function defined by

$$K.f(x) = \int_E K(x,dy)f(y)$$

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

$$\mu.K(A) = \int_E K(x,A)\mu(dx).$$

For any $f \in \mathscr{B}(E)$, denote by $\mathrm{osc}(f)$ the quantity $(f_{\max} - f_{\min})$. For any $x \in E$, the Dirac measure centered on x is designated by δ_x.

1.2 The Feynman-Kac Measure-Valued Model

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

$$\forall f \in \mathscr{B}(E), \quad \eta_n(f) = \frac{\eta_{n-1}(G_n \times M_n.f)}{\eta_{n-1}(G_n)}$$

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

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

where $\phi_n : \mathscr{P}(E) \to \mathscr{P}(E)$ 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

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

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

$$\Psi_{G_n}(\eta_{n-1})(dy) = \left(\eta_{n-1}S_{n,\eta_{n-1}}\right)(dy) := \int \eta_{n-1}(dx) S_{n,\eta_{n-1}}(x,dy)$$

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

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

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

$$\eta_n = \eta_{n-1}K_{n,\eta_{n-1}} \quad \text{with} \quad K_{n,\eta_{n-1}} = S_{n,\eta_{n-1}}M_n$$

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

$$\mathbb{P}\left(\overline{X}_{n+1} \in dy \mid \overline{X}_n = x\right) := K_{n+1,\eta_n}(x,dy)$$

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

An important point (FK semigroup structure, see e.g., [17]) is that the semigroup transformations

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

admit a comparable structure as each of the ϕ_k, i.e. for any integers $p < n$, there exist a positive function $G_{p,n} \in \mathscr{B}(E)$ and a Markov kernel $P_{p,n}$ so that:

$$\forall f \in \mathscr{B}(E), \quad \forall \mu \in \mathscr{P}(E), \quad \phi_{p,n}(\mu).f = \frac{\mu\left(G_{p,n} \times P_{p,n}.f\right)}{\mu(G_{p,n})} \tag{1}$$

1.3 The Associated Interacting Particle System

In SMC and QMC algorithms, we approximate the measures η_n by simulating an interacting particle system $(\zeta_n)_n = \left(\zeta_n^1, \ldots, \zeta_n^N\right)_n$ of size N so that

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

Of course, the main issue is to make precise and to quantify this convergence.

We start with N independent samples $\zeta_0 = (\zeta_0^1, \ldots, \zeta_0^N)$ from η_0. 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

$$S_{\eta_n^N}(\zeta_n^i, dx) := \varepsilon_{n+1}.G_{n+1}(\zeta_n^i)\ \delta_{\zeta_n^i}(dx) + \left(1 - \varepsilon_{n+1}.G_{n+1}(\zeta_n^i)\right)\ \Psi_{G_{n+1}}(\eta_n^N)(dx)$$

with the updated measures

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

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 $1 \leq i \leq N$, we sample a random state ζ_{n+1}^i with distribution $M_{n+1}\left(\widehat{\zeta}_n^i, dx\right)$.

Denote respectively by $\mathbb{P}(\cdot)$ and $\mathbb{E}(\cdot)$ probabilities and expectations taken with respect to the random variables $(\zeta_n^i)_{n,i}$ and $(\hat{\zeta}_n^i)_{n,i}$.

1.4 Dobrushin Ergodic Coefficients

The Dobrushin coefficient $\beta(K) \in [0,1]$ of a Markov kernel K on E, is defined by:

$$\beta(K) = \sup\{K(x,A) - K(y,A) \mid x,y \in E,\ A \in \mathscr{E}\},$$

or in an equivalent way:

$$\beta(K) = \sup\{\|K(x,\cdot) - K(y,\cdot)\|_{tv} \mid x,y \in E\}$$

where $\|\mu-\nu\|_{tv}$ denotes the total variation distance between the measures μ and ν.

The parameter $\beta(K)$ caracterizes mixing properties of the Markov kernel K. Note that the 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. Further details on these ergodic coefficients can be found in the monograph [10].

Estimating these coefficients is generally a difficult task (related to the large field of Markov chains' stability), since their definition involves a supremum over every pair $(x,y) \in E^2$ and every set $A \in \mathscr{E}$. However, here is a first remark: if a Markov kernel K satisfies the ergodic total variation convergence $K^m(x,\cdot) \to \mu$ uniformly w.r.t. $x \in E$ when m tends to infinity, then $\beta(K^m)$ tends to zero.

In the particular case of a finite state space E, the Dobrushin ergodic coefficient of a Markov kernel K on E is given by the formula

$$\beta(K) = \frac{1}{2}\ \sup\left\{\sum_{l\in E} |K(i,\{l\}) - K(j,\{l\})|\ ;\ i,j \in E\right\},$$

which implies it is calculable as soon as the probability of the elementary transitions $K(i,\{l\})$ are known. This formula can provide a approximation of $\beta(K)$ in the case of an infinite but simple (low dimensional) state space E, that one can discretize.

In practice, the property $\beta(K) < 1$ is easily met as soon as the state space E is compact. Typically, any Markov kernel of the form

$$K(x,dy) = h(x,y)m(dy)$$

where h is a positive, continuous function on E^2 and m a reference measure on E, satisfies $\beta(K) < 1$. Otherwise, in some other particular situations, one can explicitly estimate $\beta(K)$. For instance, if $E = \mathbb{R}^d$ and

$$K(x,dy) \propto e^{-\alpha|y-a(x)|}dy$$

for some $\alpha > 0$ and some bounded function $a\ :\ E \to E$, then for all $x,x' \in E$ we have

$$\frac{K(x,dy)}{K(x',dy)} = e^{\alpha\left(|y-a(x')|-|y-a(x)|\right)} \leq e^{\alpha\cdot osc(a)}$$

$\Rightarrow K(x,dy) \leq e^{-\alpha\cdot osc(a)}K(x',dy)$. This clearly implies $\beta(K) \leq (1-e^{-\alpha\cdot osc(a)})$.

The reader will also find in [17] an estimate of $\beta(K^2)$ in the following case

$$K(x,dy) \propto e^{-\frac{1}{2}|y-a(x)|^2}dy,$$

when the function a is constant outside some compact set $F \subset E$. Finally, the case of the Metropolis-Hastings kernel will be recalled page 11.

1.5 Some Non-Asymptotic Results

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

Definition 1. For any integers $p < n$, we set

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

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 (1). 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 [10]).

The performance analysis developed in this article is partly based on the two non-asymptotic inequalities presented below.

The following L^p error bound for all $f \in \mathscr{B}_1(E)$ is proved in [17]:

$$\mathbb{E}\left(\left|\eta_n^N(f) - \eta_n(f)\right|^p\right)^{1/p} \leq \quad \frac{B_p}{\sqrt{N}} \sum_{k=0}^{n} g_{k,n} b_{k,n} \tag{2}$$

where B_p designates an universal constant.

In the further development of the article we also use the following exponential concentration inequality derived in [18]. For all $f \in \mathscr{B}_1(E)$ and any $\varepsilon > 0$ we have:

$$\frac{-1}{N} \log \mathbb{P}\left(|\eta_n^N(f) - \eta_n(f)| \geq \frac{r_n}{N} + \varepsilon\right) \geq \frac{\varepsilon^2}{2}\left[b_n^\star \overline{\beta}_n + \frac{\sqrt{2} r_n}{\sqrt{N}} + \varepsilon\left(2r_n + \frac{b_n^\star}{3}\right)\right]^{-1} \tag{3}$$

where r_n, $\overline{\beta}_n$ and $b_n^\star$ are constants so that:

$$\begin{cases} r_n & \leq \sum_{p=0}^n 4 g_{p,n}^3 b_{p,n} \\ \overline{\beta}_n^{-2} & \leq \sum_{p=0}^n 4 g_{p,n}^2 b_{p,n}^2 \\ b_n^\star & \leq \sup_{0\leq p\leq n} 2 g_{p,n} b_{p,n} \end{cases}$$

2 General Feynman-Kac semigroup analysis

Equations (2) and (3) provide explicit non-asymptotic estimates in terms of the quantities $g_{p,n}$ and $b_{p,n}$. Written this way, they hardly apply to any SMC' 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 $b_{p,n}$ in terms of the g_p and b_p.
By construction (see Lemma 2.1 in [13]), $G_{p,n}$ and $P_{p,n}$ satisfy the following backward relations:

$$\begin{cases} G_{p-1,n} = G_p \times M_p.G_{p,n} \\ \\ P_{p-1,n}.f = \dfrac{M_p.(G_{p,n} \times P_{p,n}.f)}{M_p.G_{p,n}} \end{cases}$$

with the initial definitions $G_{n,n} = 1$ and $P_{n,n} = Id$. By combining these formulae with Dobrushin ergodic coefficient estimation techniques, we obtain the following lemma:

Lemma 1. *For any integers $p \leq n$, we have:*

$$\begin{aligned} g_{p,n} - 1 &\leq \sum_{k=p+1}^{n} (g_k - 1) \prod_{i=p+1}^{k-1} (b_i g_i) \\ b_{p,n} &\leq \prod_{k=p+1}^{n} b_k.g_{k,n} \end{aligned} \tag{4}$$

To obtain uniform bounds w.r.t. time n (in the case of the L^p norm), we notice that

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

This naturally leads to a sufficient condition of the following type:

$$b_k \times g_{k-1,n} \leq a \quad \text{with} \quad 0 < a < 1$$

for any $k < n$, which ensures:

$$\forall f \in \mathscr{B}_1(E), \quad \mathbb{E}\left(\left|\eta_n^N(f) - \eta_n(f)\right|^p\right)^{1/p} \leq \frac{B_p}{\sqrt{N}} \frac{1}{1-a} \tag{5}$$

More generally, this condition ensures uniform bounds for $\overline{\beta_n}^2$, $b_n^\star$ and r_n :

$$\overline{\beta_n}^2 \leq \frac{4}{1-a^2} \qquad b_n^\star \leq 2 \qquad r_n \leq \frac{4}{1-a} \cdot \sup_{p,n} g_{p,n}^2$$

bearing in mind that the $g_{p,n}$ are bounded in the cases of interest. We then fix $0 < a < 1$ and the objective is to find conditions on the b_p so that $b_k g_{k-1,n} \leq a$. This

parameter a is to be chosen according to the error we allow ourselves to commit, and the number N of particles involved, with bounds explicited above. In order to explicit relevant and applicable conditions, we study two typical cases of assumptions on the potential functions G_p. The first one being that the g_p are bounded (Theorem 1), the second one being that the g_p tend to 1 (Theorem 2).

Theorem 1. *Under the assumption $\forall p \in \mathbb{N}, g_p \leq M$, where M is a constant, condition*

$$b_p \leq \frac{a}{M(1+a)} \tag{6}$$

ensures the L^p error bound (5), as well as the following concentration inequality:

$$\forall y \geq 0, \quad \forall f \in \mathscr{B}_1(E), \quad \mathbb{P}\left(|\eta_n^N(f) - \eta_n(f)| \geq \frac{r_1^\star N + r_2^\star y}{N^2} \right) \leq e^{-y}$$

with

$$\begin{cases} r_1^\star = \frac{9}{2}\frac{M^2}{(1-a)^3} + \sqrt{\frac{8}{\sqrt{1-a^2}} + \frac{18M^2}{(1-a)^2\sqrt{N}}} \\ \\ r_2^\star = 18\frac{M^2}{(1-a)^2} + \sqrt{\frac{8}{\sqrt{1-a^2}} + \frac{18M^2}{(1-a)^2\sqrt{N}}} \end{cases}$$

Let us now consider the case where g_p tends decreasingly to 1. We define

$$\alpha = \frac{a}{1-a} > 0 \quad \text{so that} \quad a = \frac{\alpha}{1+\alpha}$$

Theorem 2. *Under the assumption $g_p \longrightarrow_{p\to\infty} 1$ (decreasingly), if the sequence b_p satisfies for any $p \geq 1$,*

$$b_p \leq \frac{g_p^\alpha - 1}{g_p^{\alpha+1} - 1} \longrightarrow a \quad \text{and} \quad b_p \leq \frac{a}{g_p^{\alpha+1}} \longrightarrow a$$

then the L^p error bound (5) is satisfied, as well as the following concentration inequalities :

$$\forall y \geq 0, \quad \forall f \in \mathscr{B}_1(E), \quad \mathbb{P}\left(|\eta_n^N(f) - \eta_n(f)| \geq \frac{r_3^\star(n).N + r_4^\star(n).y}{N^2} \right) \leq e^{-y}$$

In the above displayed formulae $r_3^\star(n)$ and $r_4^\star(n)$ are defined below in terms of a sequence u_n which tends to 1, as n tends to ∞:

$$\begin{cases} r_3^\star(n) = \frac{9.u_n}{2(1-a)} + \sqrt{\frac{8}{\sqrt{1-a^2}} + \frac{18.u_n}{\sqrt{N}}} \\ \\ r_4^\star(n) = \frac{18.u_n}{1-a} + \sqrt{\frac{8}{\sqrt{1-a^2}} + \frac{18.u_n}{\sqrt{N}}} \end{cases}$$

Such conditions on the b_p can appear to be difficult to reach since the Markov kernels may be imposed 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, the algorithm designer can add MCMC evolution steps next to each M_n-mutation step, to stabilize the system. From the 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'_p satisfy:

$$b'_p \leq b_p.\beta(K_p^{m_p}) \leq b_p.\beta(K_p)^{m_p}.$$

In such cases, Theorems 1 and 2 provide sufficient conditions on iteration numbers m_p to ensure the convergence of the algorithm.

3 The Particular Case of Boltzmann-Gibbs Measures, Interacting Simulated Annealing

Let $V \in \mathscr{B}(E)$. For all $\beta \geq 0$, denote Boltzmann-Gibbs probability measure associated with "temperature" β and potential function V by:

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

where m is a reference measure, and Z_β a normalizing constant. It is well known that Boltzmann-Gibbs measures' sampling is related to the problem of minimizing the potential function V, since μ_β tends to concentrate on V's minimizers as temperature β tends to infinity. One illustration is the following inequality, satisfied for all $0 < \varepsilon' < \varepsilon$:

$$\mu_\beta\left(V \geq V_{\min} + \varepsilon\right) \leq \frac{e^{-\beta(\varepsilon-\varepsilon')}}{m_{\varepsilon'}} \tag{7}$$

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

Besides, let fix a "temperature schedule", being a strictly increasing sequence β_n so that $\beta_n \longrightarrow +\infty$. The sequence $(\eta_n)_n := (\mu_{\beta_n})_n$ admits a FK structure associated with potential functions $G_n = e^{-(\beta_n - \beta_{n-1}).V}$ and Markov kernels M_n chosen as being MCMC dynamics for the current target distributions. In this context, the SMC algorithm, used as a strategy to minimize V, can be interpreted as a sequence of

interacting simulated annealing (*abbreviate ISA*) algorithms.
We propose in this section to turn the previously raised conditions on b_p and g_p into conditions on the temperature schedule to use, and the number of MCMC steps. We will then combine the concentration results of section 2 with inequality (7) to obtain results in terms of optimization performance.

Let us fix a "temperature schedule" (β_n) and denote:

- $\eta_n(dx) = \mu_{\beta_n}(dx) = \frac{1}{Z_{\beta_n}} e^{-\beta_n V(x)} m(dx)$;
- $G_p(x) = e^{-\Delta_p . V(x)}$;
- and then $g_p = e^{\Delta_p . \mathrm{osc}(V)}$.

where Δ_p are the increments of temperature $\Delta_p = \beta_p - \beta_{p-1}$. At a fixed temperature β , let us consider the simulated annealing Markov kernel, designated by K_β. It involves a proposition kernel $K(x,dy)$, assumed here as being fixed, according to the following formulae (written here in the case where K is symmetric, see [3]):

$$K_\beta(x,dy) = K(x,dy).\min\left(1, e^{-\beta(V(y)-V(x))}\right) \qquad \forall y \neq x$$

$$K_\beta(x,\{x\}) = 1 - \int_{y\neq x} K(x,dy).\min\left(1, e^{-\beta(V(y)-V(x))}\right)$$

Under the assumption $K^{k_0}(x,\cdot) \geq \delta\nu(\cdot)$ for some integer k_0, some measure ν and some $\delta > 0$, one can show (see [3]) that:

$$\beta(K_\beta^{k_0}) \leq \left(1 - \delta e^{-\beta\overline{\Delta V}(k_0)}\right) \tag{8}$$

where $\overline{\Delta V}(k_0)$ is the maximum potential gap one can obtain making k_0 movements with K. This quantity is bounded by $\mathrm{osc}(V)$. To let the b_p's tuning be possible, it is out of the question to choose $M_p = K_{\beta_p}$, but $M_p = K_{\beta_p}^{k_0.m_p}$, the simulated annealing kernel iterated $k_0.m_p$ times, to obtain suitable mixing properties. The algorithm's user then has a choice to make on two parameters: the temperature schedule β_p, and the kernels $K_{\beta_p}^{k_0}$ iteration numbers m_p. Note that for all $b \in (0,1)$, condition $b_p \leq b$ is turned into $\left(1 - \delta e^{-\beta_p\overline{\Delta V}(k_0)}\right)^{m_p} \leq b$, which can also be written:

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

Then, combining the concentration inequality (7), the theorems of section 2 (taken with indicator function $f = \mathbf{1}_{\{V \geq V_{\min}+\varepsilon\}}$) , and the Dobrushin ergodic coefficient estimation (8) we obtain the following theorem:

Theorem 3. *Let us fix $a \in (0,1)$. If the temperature schedule (β_p) and the iteration numbers m_p satisfy one of these two conditions:*

1. Δ_p bounded by Δ (e.g. linear temperature schedule) and

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

2. $\Delta_p \longrightarrow 0$ *(decreasingly) and* $m_p \geq \left(osc(V).\Delta_p + log(\frac{1}{a})\right)\frac{e^{\overline{\Delta V}(k_0).\beta_p}}{\delta}$

then for all $\varepsilon > 0$*, and all* $\varepsilon' < \varepsilon$*, the proportion* $p_n^N(\varepsilon)$ *of particles* (ζ_n^i) *so that* $V(\zeta_n^i) \geq V_{\min} + \varepsilon$ *satisfies the inequality:*

$$\forall y \geq 0, \quad \mathbb{P}\left(p_n^N(\varepsilon) \geq \frac{e^{-\beta_n(\varepsilon-\varepsilon')}}{m_{\varepsilon'}} + \frac{r_i^\star N + r_j^\star y}{N^2}\right) \leq e^{-y}$$

where $(i,j) = (1,2)$ *in the case of bounded* Δ_p *(taken with* $M = e^{\Delta.osc(V)}$*) and* $(i,j) = (3,4)$ *in the second one.*

We then clearly distinguish two error terms: the first one, $\left(\frac{e^{-\beta_n(\varepsilon-\varepsilon')}}{m_{\varepsilon'}}\right)$, estimating the Boltzmann-Gibbs measure's concentration around V's minimizers, and the second one, $\left(\frac{r_i^\star N + r_j^\star y}{N^2}\right)$, estimating the occupation measure's concentration around this Boltzmann-Gibbs theoretical measure. More than providing tunings which ensure convergence, this last concentraion inequality explicits the relative impact of other parameters, such as probabilistic precision y, threshold t on the proportion of particles possibly out of the area of interest, final temperature β_n or population size N. A simple equation, deduced from this last theorem, such as $\left(\frac{e^{-\beta_n(\varepsilon-\varepsilon')}}{m_{\varepsilon'}} = \frac{r_i^\star N + r_j^\star y}{N^2} = \frac{t}{2}\right)$ may be applied to the global tuning of an Interacting Simulated Annealing algorithm, which is generally a difficult task.

Conclusion

It is instructive to compare the estimates of Theorem 3 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 3 is again a logarithmic one. In contrast to the SA model, Theorem 3 allows to quantify the performance analysis of the ISA model in terms of uniform concentration inequalities, that does not depend on a critical parameter.
Like most rigorous and non-asymptotic tuning theorems, our results may not be applied directly. They highlight important principles (as uniform accessibility of all

the state space after a given number of mutations) and the type of dependence in some parameters. Otherwise, to our knowledge, our work presently provides the most explicit non-asymptotic ISA convergence results, at least in the case $|E| = \infty$. Nevertheless, the models we studied involve a deterministic sequence β_n, while choosing the sequence of increments $\Delta_n = (\beta_n - \beta_{n-1})$ in advance can cause computational problems. In practice, 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 for instance [22, 25], [8, 19, 24]). In a forthcoming paper (see [20] for a preliminary version), we try to adapt the present work to analyze one of these adaptive tuning strategies.

References

1. Assaraf,R, Caffarel, M.: A Pedagogical Introduction to Quantum Monte Carlo. Mathematical models and methods for ab initio Quantum Chemistry, 45-73 Springer (2000) doi: 10.1007/978-3-642-57237-1_3
2. Assaraf, R., Caffarel, M., Khelif, A.: Diffusion Monte Carlo with a fixed number of walkers. Physical Review E, vol. 61, no 4, p. 4566 (2000)
3. Bartoli, N., Del Moral, P.: Simulation & Algorithmes Stochastiques. Cépaduès (2001)
4. Cappé,O., Moulines, E., Rydén, T.: Inference in Hidden Markov Models. Springer, New York, (2005)
5. Cérou, F., Del Moral, P., Furon, T., Guyader, A.: Sequential Monte Carlo for Rare Event Estimation. Statistics and Computing, vol 22, no 3, 795–808 (2012) doi: 10.1007/s11222-011-9231-6
6. Cérou, F., Del Moral, P., Guyader, A.: A nonasymptotic variance theorem for unnormalized Feynman-Kac particle models. Annales de l'Institut Henri Poincaré, Probabilités et Statistiques, 47, 629–649 (2011)
7. Chopin, N.: Central Limit Theorem for Sequential Monte Carlo methods and its application to Bayesian inference. Annals of Statistics, 32, 2385-2411 (2004)
8. Clapp, T.: Statistical Methods in the Processing of Communications Data. Ph.D. thesis, Cambridge University Engineering Department. (2000)
9. Dawson D.A., Del Moral P.: Large deviations for interacting processes in the strong topology. Statistical Modeling and Analysis for Complex Data Problem P. Duchesne and B. Rmillard Editors, pp. 179–209, Springer (2005)
10. Del Moral, P.: Feynman-Kac formulae. Genealogical and interacting particle approximations Springer New York, [575p.] Series: Probability and Applications (2004).
11. Del Moral, P.: Nonlinear filtering: Interacting Particle Solution. Markov Processes and Related Fields, 2, 555–579 (1996).
12. Del Moral, P., Doucet, A., Jasra, A.: Sequential Monte Carlo Samplers. Journal of the Royal Statistical Society B, 68, 411–436 (2006)
13. Del Moral, P., Guionnet, A.: On the stability of interacting processes with applications to filtering and genetic algorithms. Annales de l'Institut Henri Poincaré, vol 37, No. 2, 155-194 (2001).
14. Del Moral P., Guionnet A. A Central Limit Theorem for Non Linear Filtering using Interacting Particle Systems. Annals of Applied Probability, Vol. 9, No. 2, 275-297 (1999).
15. Del Moral P., Guionnet A. Large Deviations for Interacting Particle Systems. Applications to Non Linear Filtering Problems. Stochastic Processes and their Applications, vol. 78, pp. 69-95 (1998).

16. Del Moral P., Ledoux M. On the Convergence and the Applications of Empirical Processes for Interacting Particle Systems and Nonlinear Filtering. Journal of Theoret. Probability, Vol. 13, No. 1, 225-257 (2000).
17. Del Moral, P., Miclo, L.: Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to nonlinear filtering. Séminaire de Probabilités XXXIV, Lecture Notes in Mathematics, vol 1729, Springer, Berlin, pp. 1-145 (2000)
18. Del Moral, P., Rio, E.: Concentration Inequalities for Mean Field Particle Models. Annals of Applied Probabilities, vol 21, no 3, 1017-1052 (2011)
19. Deutscher, J., Blake, A., Reid, I.: Articulated body motion capture by annealed particle filtering. IEEE Conference on Computer Vision and Pattern Recognition. Vol. 2, pp.126-133. (2000)
20. Giraud, F., Del Moral, P.: Non-asymptotic analysis of adaptive and annealed Feynman-Kac particle models. arXiv math.PR/12095654 (2012)
21. Hetherington, J. H.: Observations on the statistical iteration of matrices. Physical Review A, vol 30, 2713-2719 (1984)
22. Jasra, A. Stephens, D., Doucet, A., Tsagaris, T.: Inference for Lévy Driven stochastic volatility models via adaptive sequential Monte Carlo. Scand. J. Stat. (2008)
23. Künsch, H. R.: Recursive Monte-Carlo filters: algorithms and theoretical analysis. Annals of Statistics, 33, 1983–2021 (2005)
24. Minvielle, P., Doucet, A., Marrs, A., Maskell, S.: A Bayesian approach to joint tracking and identification of geometric shapes in video squences. Image and Visaion Computing. Vol. 28, pp. 111-123. (2010)
25. Schäfer, C., Chopin, N.: Sequential Monte Carlo on large binary sampling spaces. Statistics and Computing. (2011)
26. Schweizer, N.: Non-asymptotic error bounds for sequential MCMC and stability of Feynman-Kac propagators. Working Paper, University of Bonn (2012)
27. Whiteley, N.: Sequential Monte Carlo samplers: error bounds and insensitivity to initial conditions. Working Paper, University of Bristol (2011)