Annealed Feynman-Kac Models

P. Del Moral* L. Miclo*

Abstract

We analyze the concentration properties of an annealed Feynman-Kac model in distribution space. We characterize the concentration regions in terms of a variational problem involving a competition between the potential function and the mutation kernel. When the temperature parameter is evanescent with time and under appropriate hypotheses, the probability mass tends to concentrate on regions with minimal potential values. We give a precise description of these areas using non-linear semi-group contractions and large deviations techniques. We illustrate this annealed model with two physical interpretations related respectively to Markov motions in absorbing media and interacting measure valued processes.

1 Introduction

This study is concerned with the long time behavior of a Feynman-Kac model associated to a potential function and a cooling schedule. This annealed distribution flow can be interpretated as the evolution of the laws of a Markov particle in an absorbing medium conditioned to non-extinction. In this context the cooling schedule represents in some sense the temperature of the medium. The smaller it is, the more stringent becomes the obstacles.
These Feynman-Kac models can alternatively be regarded as a dynamical system in distribution space. In this context they can model the evolution of the marginal laws of a non-linear and non-homogeneous Markov process. The non-linearity comes from the fact that the elementary transitions depend on the distribution flow itself. In this connection the non-linear Markov model can be regarded as a Feynman-Kac type simulated annealing algorithm with mutation/selection transitions. As in the traditional simulated annealing model the temperature parameter is used to increase the selection pressure of the algorithm. These non-linear measure valued processes have a natural genetic type particle interpretation. They have some important applications in biology, advanced signal processing and numerical function analysis [4]. In contrast to previous studies on the convergence of genetic algorithms in global optimization problems (see for instance [2, 3] and the references given there) we underline that our study is not restricted to finite state spaces and above all that our

*Laboratoire de Statistique et Probabilités, UMR 5583, CNRS and Université Paul Sabatier, 31062 Toulouse cedex 4, France, {delmoral, miclo}@cict.fr

mutation kernel is homogeneous in time (and not related to the cooling schedule), in particular it does not force the underlying (ie unperturbed by potentials) system to be motionless for large time. Our final counter-example relative to convergence to the global minima enables to better apprehend this classical assumption and we will see that it can be advantageously replaced by permitting the mutation kernel to have loops on each point of the state space. This latter precaution is very mild from the point of view of implementation of the procedure, whose speed of convergence should benefit a lot from homogeneity in time of the mutations (and furthermore, endowed with this feature, our algorithm seems closer to true biological/genetical mechanisms than the traditional ones).

The above physical interpretations result from two ways to turn a sub-Markov and Boltzmann-Gibbs operator into a Markov kernel. One of the central questions in the study of these annealed Feynman-Kac models is of course the investigation of the long time behavior of these flows. Intuitively speaking when the temperature parameter tends to zero the probability mass of regions with high potential values decreases and the flow tends to concentrates to regions with minimal potential. The main objective of this article is to make clear this statement. First we discuss the convergence to equilibrium of the annealed models. We exhibit two different types of cooling schedules depending on the mixing parameter of the mutation transitions. Then we characterize the asymptotic regions where the flow concentrates in terms of a variational problem in distribution space. We show that the concentration properties of the annealed model are the result of a competition between the selection potential and the mutation transition. When the temperature parameter tends to zero the variation problem is solved by taking the infimum of the mean potentials over a suitably chosen collection of measures. We already mention that for sufficiently regular mutations and for judicious choices of cooling schedules the annealed model converge in probability to the global infimum of the potential function.

To our knowledge the asymptotic concentration properties of the annealed Feynman-Kac flow presented in this study have not been covered by the literature. We propose an original strategy based on non-linear semi-group contraction and large deviation techniques. The question of the stability of non-linear Feynman-Kac semi-groups arise in many research areas. To our knowledge the first studies in this field originate in non-linear filtering literature. We again refer the reader to [4] for a precise discussion and a precise list of referenced papers. To our knownledge most of these works are only concerned in proving that the flow forgets its initial condition.
The reason why these studies do not apply are two folds. First the annealed Feynman-Kac flows presented here are related to an increasing cooling schedule and the resulting functions $e^{-\beta(n)V}$ tends to the indicator function of null potential regions. The analysis of this degenerate situation is more involved and clearly differs from traditional filtering studies. On the other hand our objective is not to check that the flow forgets its initializations but we want to identify the regions on which the distributions concentrates. Our way to enter into this question has been influenced by the article [5]. This study shed some new light on the connections between the limiting distribution of the homogenous model and the Lyapuonv

exponent of Schrödinger-Feynman-Kac semi-groups. Here we develop the profound interplay between the asymptotic behavior of these exponents and the concentration properties of the limiting measures.
We also present a set of sufficient conditions on the mutation transitions under-which these limiting regions coincide with the essential infimum of the potential function with respect to the invariant measure of the mutation kernel. We end this article showing that annealed models with off-diagonal type mutations may lose the essential infimum.

1.1 Description of the models and motivations

Let E be a separable complete metric space endowed with its Borel σ-field $\mathcal{E}$ and let $V : E \to \mathbb{R}_+ = [0, \infty)$ be a non-negative bounded continuous potential function on E whose oscillation will be denoted by

$$\operatorname{osc}(V) = \sup\{V(x) - V(y) \; ; \; (x, y) \in E^2\}$$

We consider a Markov kernel $M(x, dy)$ on $(E, \mathcal{E})$ and we denote by

$$(\Omega = E^{\mathbb{N}}, F = (F_n)_{n\in\mathbb{N}}, X = (X_n)_{n\in\mathbb{N}}, (\mathbb{P}_x)_{x\in E})$$

the canonical Markov chain with elementary transition M. For a given distribution μ on E we use the notation $\mathbb{E}_\mu(.)$ for the expectation with respect to

$$\mathbb{P}_\mu(.) = \int_E \mu(dx) \; \mathbb{P}_x(.)$$

We recall that a Markov kernel M generates two integral operators, one acting on the Banach space $\mathcal{B}(E)$ of bounded measurable functions f and the other on the set $\mathcal{P}(E)$ of probability measures μ on E by

$$M(f)(x) = \int_E M(x, dy) \; f(y) \quad \text{and} \quad (\mu M)(A) = \int_E \mu(dx) \, M(x, A)$$

for any $x \in E$ and $A \in \mathcal{E}$.
The space $\mathcal{B}(E)$ is endowed with the supremum norm

$$\|f\| = \sup_{x\in E} |f(x)|$$

and the set $\mathcal{P}(E)$ with the total variation distance

$$\|\mu - \eta\|_{\mathrm{tv}} = \sup_{A\in\mathcal{E}} |\mu(A) - \eta(A)| = \frac{1}{2} \sup_{f\in\mathcal{B}(E) \,:\, \|f\|\leq 1} |\mu(f) - \eta(f)|$$

Here are collected some standard notations and conventions. If Q is a bounded positive operator on $\mathcal{B}(E)$ then we denote by Q^n, $n \in \mathbb{N}$, the semi-group defined by the inductive formula

$$Q^n = QQ^{n-1} \quad \text{with} \quad Q^0 = Id$$

Unless otherwise mentioned, x, y, z denote generic points in E, f an arbitrary bounded and measurable test-function on E and μ a given probability measure on E. We also use the conventions $\inf_\emptyset = \infty$, $\sup_\emptyset = -\infty$ and $\sum_\emptyset = 0$, $\prod_\emptyset = 1$ and $\lfloor a \rfloor$ denotes the integer part of a number $a \in \mathbb{R}_+$.

For a given inverse-freezing schedule $\widetilde{\beta} : \mathbb{N} \to \mathbb{R}_+$ we denote by $\mu_n \in \mathcal{P}(E)$ the annealed distribution flow defined by the Feynman-Kac foumulae

$$\mu_n(f) = \frac{\gamma_n(f)}{\gamma_n(1)} \quad \text{with} \quad \gamma_n(f) = \mathbb{E}_{\mu_0}[\ f(X_n)\ e^{-\sum_{p=1}^n \widetilde{\beta}(p) V(X_p)}\ ] \tag{1}$$

In this notation we notice that $\gamma_0 = \mu_0$ is the initial distribution law of the chain X. When the cooling schedule parameter $\widetilde{\beta}(n) \to \infty$ as $n \to \infty$, the distributions μ_n should intuitively concentrate on some regions with minimal potential values. To illustrate this observation and motivate this article we present next two different physical interpretations of the Feynman-Kac model. To clarify the presentation we restrict ourselves to the homogeneous situation and we suppose $\widetilde{\beta}(n) = \beta \in \mathbb{R}_+$.

Before proceeding further, we notice from the multiplicative nature of these formulae that the finite measures γ_n satisfy a linear equation of the form

$$\gamma_n = \gamma_{n-1} Q \tag{2}$$

$$\text{with} \qquad Q(x, dz) = \int_E M(x, dy)\ G(y, dz) \quad \text{and} \quad G(x, dy) = e^{-\beta V(x)} \delta_x(dy)$$

The first way to turn the sub-Markovian kernel G into the Markov case consists in adding a cemetery point $c \notin E$ to the state space E and in extending the various quantities as follows. The test functions $f \in \mathcal{B}(E)$ and respectively the Markov transition M are first extended to $E \sqcup \{c\}$ by setting $f(c) = 0$ (note that $V(c) = 0$) and

$$M^c(x, dy) = 1\!\!1_E(x)\ M(x, dy) + 1\!\!1_c(x)\ \delta_c(dy)$$

Finally the Markov extension G^c of G on $E \sqcup \{c\}$ is given by

$$G^c(x, dy) = e^{-\beta V(x)}\ \delta_x(dy) + (1 - e^{-\beta V(x)})\ \delta_c(dy) \tag{3}$$

The corresponding Markov chain $(\Omega, F, X, (\mathbb{P}_x^c)_{x \in E \sqcup \{c\}})$ with elementary transitions $M^c G^c$ can be regarded as a Markov particle evolving in an environment with absorbing obstacles related to a potential function V. When the temperature of the medium decreases the obstacles become more and more stringent and the particle is more rapidely absorbed. In this physical context the Feynman-Kac flow $(\mu_n)_{n \in \mathbb{N}}$ defined by (1) represents the conditional distributions of the particle motions given the fact it has not been absorbed. With some obvious abusive notations we have

$$\mu_n(f) = \mathbb{E}_{\mu_0}^c(f(X_n) \mid T > n)$$

where $T = \inf\{n \geq 0\ ;\ X_n = c\}$ is the life time of X and where $\mathbb{E}_{\mu_0}^c(.)$ is the expectation with respect to $\mathbb{P}_{\mu_0}^c(.)$. Intuitively speaking, when the temperature of the medium decreases a non-absorbed particle will evolve in regions with low potential obstacles.

In measure valued and interacting processes literature the Feynman-Kac flow is alternatively seen as a solution of a non-linear evolution equation of the form

$$\mu_n = \mu_{n-1}\mathcal{K}^{\beta}_{\mu_{n-1}} \tag{4}$$

with a collection of Markov kernels $\mathcal{K}^{\beta}_{\mu}$ on E. The choice of $\mathcal{K}^{\beta}_{\mu}$ is not unique. By direct inspection we see from (2) that we can choose

$$\mathcal{K}^{\beta}_{\mu} = M\mathcal{S}^{\beta}_{\mu M}$$

with the Markov kernels $\mathcal{S}^{\beta}_{\mu}$ on E defined by

$$\mathcal{S}^{\beta}_{\mu}(x, dy) = e^{-\beta V(x)}\ \delta_x(dy) + (1 - e^{-\beta V(x)})\ \Psi^{\beta}(\mu)(dy)$$

Here Ψ^{β} is the Gibbs-Boltzmann mapping from $\mathcal{P}(E)$ into itself defined by

$$\Psi^{\beta}(\mu)(dy) = \frac{1}{\mu(e^{-\beta V})}\ e^{-\beta V(y)}\ \mu(dy)$$

Note that the evolution equation (4) is decomposed in two separate Markov transitions

$$\mu_{n-1} \xrightarrow{M} \eta_n = \mu_{n-1}M \xrightarrow{\mathcal{S}^{\beta}_{\eta_n}} \mu_n = \eta_n\mathcal{S}^{\beta}_{\eta_n} \tag{5}$$

In connection with the first interpretation we have turn here the sub-Markovian kernel G into the Markov case in a non-linear way by replacing the Dirac measure on the cemetery point c by the Gibbs-Boltzmann's distribution $\Psi^{\beta}(\eta_n)$. In this context the probabilistic interpretation consists in introducing a new reference probability measure $\mathbb{P}^{\beta}_{\mu_0}$ on the canonical space underwhich μ_n is the distribution law of a non-homogenous and non-linear Markov chain $(Z_n)_{n\in\mathbb{N}}$. This measure is called the McKean measure associated to the collection of kernels $\mathcal{K}^{\beta}_{\mu}$ and it is defined by its n-time marginals $\mathbb{P}^{\beta}_{\mu_0,n}$

$$\mathbb{P}^{\beta}_{\mu_0,n}(d(x_0,\ldots,x_n)) = \mu_0(dx_0)\ \mathcal{K}^{\beta}_{\eta_0}(x_0, dx_1)\ \cdots\ \mathcal{K}^{\beta}_{\eta_{n-1}}(x_{n-1}, dx_n) \tag{6}$$

In this notation we clearly have that

$$\mu_n(f) = \mathbb{E}^{\beta}_{\mu_0}(f(Z_n))$$

where $\mathbb{E}^{\beta}_{\mu_0}(\cdot)$ is the expectation with respect to $\mathbb{P}^{\beta}_{\mu_0}(\cdot)$. The non-linear model (4) has a natural interacting particle interpretation. We will not enter into more details here this would be too much disgression and we refer the interest reader to the survey paper [4]. For completeness and for the convenience of the reader we give next an intuitive feel of the particles motions. As dictated by (5) the evolution of the systems is again decomposed into two mechanisms. During the first one the particles explore the state space independently of one each other according to the mutation kernel M. After this mutation stage each particle in a site x decides with a probability $e^{-\beta V(x)}$ to stay in its location and with a probability $1-e^{-\beta V(x)}$ it selects a new particle at site y proportionally to its fitness $e^{-\beta V(y)}$. This interacting particle model can be regarded as a genetic approximating model or as a Feynman-Kac version of the traditional simulated annealing algorithm. We refer the reader to the book of Duflo [7] and references therein for a precise description of these classical stochastic algorithms. In this "engineering perspective" the objective is to find the right mutation kernel and a judicious cooling schedule so that the particle will converge as the time tends to infinity to the global minima of the potential function V.

1.2 Statement of some results

In this section we present a quick derivation of our main results. For an homogeneous cooling schedule, namely for all $n \in \mathbb{N}$, $\widetilde{\beta}(n) = \beta \in \mathbb{R}_+$, the distribution flow (4) is homogenous with respect to the time parameter. More precisely it satisfies a non-linear equation in distribution space of the form

$$\mu_n = \Phi^\beta(\mu_{n-1})$$

The mapping Φ^β from $\mathcal{P}(E)$ into itself is defined by the formula

$$\Phi^\beta(\mu)(f) = \mu M(e^{-\beta V} f)/\mu M(e^{-\beta V})$$

By Φ^β_n, $n \in \mathbb{N}$, we denote the corresponding semi-group

$$\Phi^\beta_{n+1} = \Phi^\beta \circ \Phi^\beta_n \quad \text{with} \quad \Phi^\beta_0 = Id$$

We will work with the following mixing type condition

(H) *There exists an integer $m \geq 1$ and an $\epsilon \in (0,1)$ such that for each $x, y \in E$*

$$M^m(x, \boldsymbol{.}) \geq \epsilon\ M^m(y, \boldsymbol{.})$$

This condition clearly holds true for every finite state space and irreducible Markov kernel M but it is also satisfied for Gaussian or bi-Laplace transitions with bounded drift functions (cf. [4]).

The reasons for the introduction of this mixing type condition are two fold.

- The main reason is that it guarantees the existence of an unique fixed point $\mu^\beta = \Phi^\beta(\mu^\beta) \in \mathcal{P}(E)$ of each mapping Φ^β, $\beta \in \mathbb{R}_+$. We will deduce this result from a contraction property of the semi-group Φ^β_n. More precisely we will find a contraction parameter $k \in (0,1)$ and a collection of integers $n(\beta)$, $\beta \in \mathbb{R}_+$, such that for any $\mu, \eta \in \mathcal{P}(E)$

$$\|\Phi^\beta_{n(\beta)}(\mu) - \Phi^\beta_{n(\beta)}(\eta)\|_{\mathrm{tv}} \ < \ k\ \|\mu - \eta\|_{\mathrm{tv}}$$

For judicious choices of cooling schedule $\widetilde{\beta}(n)$ one expects that the non-homogeneous Feynman-Kac flow behaves as the limiting distributions $\mu^{\widetilde{\beta}(n)}$, as it is usual for simulated annealing (see for instance [9, 10]).

- On the other side, condition (H) allows to connect the concentration properties of μ^β with the asymptotic behavior of the logarithmic Lyapunov exponent

$$\Lambda(-\beta V) = \lim_{n\to\infty} \frac{1}{n} \ln \sup_{x\in E} \mathbb{E}_x(e^{-\beta \sum_{p=1}^n V(X_p)}) \tag{7}$$

of the semi-group $Q(f) = M(e^{-\beta} f)$ (on the Banach space $\mathcal{B}(E)$).

To describe in some details the interplay between the exponents (7) and the fixed point measures μ^β we recall that under condition (H) the occupation measures

$$L_n = \frac{1}{n} \sum_{p=1}^n \delta_{X_p}$$

converge as $n \to \infty$ to an unique invariant measure $\nu = \nu M$. The exponential deviant behavior of L_n is expressed by a large deviation principle. Namely, the distributions sequence L_n satisfy as $n \to \infty$ a large deviation principle with a convex rate function $I : \mu \in \mathcal{P}(E) \to [0, \infty]$ defined by

$$I(\mu) = \inf \ \{\int_E \ \mu(dx) \ \mathrm{Ent}(K(x, .)|M(x, .))\}$$

where the infimum is taken over all Markov kernels K with invariant measure μ. For a proof of this statement we refer the reader to the book of Dupuis and Ellis [8]. Note that $I(\mu) < \infty$ implies that we can find a Markov kernel K leaving μ invariant, $\mu K = \mu$, and verifying $K(x, .) << M(x, .)$ for μ-a.s. all $x \in E$. Loosely speaking these probability measures can be interpretated as the limiting distributions of M-admissible Markov chains.

The rate function I, the fixed points μ^β and the logarithmic Lyapunov exponents $\Lambda(-\beta V)$, $\beta \in \mathbb{R}_+$, are connected by the following formula

$$-\Lambda(-\beta V) = \ln \mu^\beta(e^{\beta V}) = \inf_{\eta\in\mathcal{P}(E)} (\beta \ \eta(V) + \ I(\eta))$$

This formula expresses the concentration properties of the limiting measures μ^β in terms of a variational problem in $\mathcal{P}(E)$ with a competition between the mean potential $\eta(V)$ and the I-entropy $I(\eta)$. If we combine this expression with the exponential version of Markov inequality we can check that for any $\delta > 0$

$$\lim_{\beta\to\infty} \frac{1}{\beta} \ln \mu^\beta(V \geq V_I + \delta) \leq -\delta \quad \text{with} \quad V_I = \inf\{\eta(V) \ ; \ \eta \in \mathcal{P}(E) \, , \quad I(\eta) < \infty\}$$

Note that V_I represents the minimal mean potential value we can asymptotically obtain running all M-admissible Markov chains. This asymptotic results indicates that the fixed points μ^β concentrate as β tends to infinity to regions with potential less than V_I. We will see that V_I is always greater than the ν-essential infimum of V defined by

$$V_\nu = \sup\{v \in \mathbb{R}_+ \; ; \; V \geq v \quad \nu - \text{a.e.}\}$$

The interplay between V_I and V_ν depends on the nature of Markov kernel M. For regular Markov kernels M we will see that $V_I = V_\nu$. Neverless we will also exhibit situations where $V_I > V_\nu$ and for some $\delta > 0$, $\mu^\beta(V \leq V_\nu + \delta) \to 0$ as the parameter $\beta \to \infty$.

If $(\beta(n))_{n\in\mathbb{N}}$ (respectively $(t(n))_{n\in\mathbb{N}}$) is an increasing sequence of non-negative reals (resp. non-negative integers, with $t(0) = 0$), we will consider the cooling schedule $\widetilde{\beta}$ given by

$$\forall\, n \in \mathbb{N},\ \forall\, t(n) \leq p < t(n+1), \qquad \widetilde{\beta}(p) = \beta(n)$$

But by a slight language abuse, from now on the sequence $\beta := (\beta(n))_{n\in\mathbb{N}}$ will be called the cooling schedule and $(t(n))_{n\in\mathbb{N}}$ the time mesh sequence.
Our result is basically stated as follows

Theorem 1.1 *When the mixing condition holds true for $m = 1$ we have $V_I = V_\nu$. In addition then there exists an integer parameter $\Delta \geq 1$ such that for any choice of cooling schedule of the type*

$$\beta(n) = \beta(0)\ (n+1)^a \quad , \text{with} \quad a \in (0,1) \quad \text{and} \quad \beta(0) < \infty$$

we have for the time mesh $t(n) = n\Delta$,

$$\|\mu_{t(n)} - \mu^{\beta(n)}\|_{\text{tv}} \leq \frac{C_0}{(n+1)^{1-a}}$$

for a certain constant $C_0 > 0$.
When condition (H) is only satisfied for some $m > 1$, then for a logarithmic cooling schedule of the type

$$\beta(n) = \beta(0)\ \ln(n+e) \quad , \text{with} \quad b = (m-1)\text{osc}(V)\beta(0) < 1$$

we have for some time mesh $t(n) = \mathcal{O}(n^{1+b}\ln n)$,

$$\|\mu_{t(n)} - \mu^{\beta(n)}\|_{\text{tv}} \leq C_1\ \frac{\ln(n+e)}{(n+1)^{1-b}}$$

for a certain constant $C_1 > 0$, and for each $\delta > 0$

$$\lim_{n\to\infty} \mu_{t(n)}(V \geq V_I + \delta) = 0$$

We note that in the latter case, the "true" cooling schedule $\widetilde{\beta}$ is also quasi-logarithmic: for large time $p \in \mathbb{N}$, $\widetilde{\beta}(p) = \mathcal{O}(\ln(p))$, as it is customary in simulated annealing.
In the context of Markov motions in an aborbing medium the choice of the mutation kernel M is dictated by the problem at hand. In this situation the analysis on the concentration levels discussed in this article will provide a way to predic the location

of a non-absorbed particle when the temperature of the medium decreases.
On the other side of the picture if we want to construct a Feynman-Kac simulated annealing which converge to the global infimum of a potential we have the choice of the mutation transitions. The concentration results presented here can be used to find a judicious choice of mutation kernel. In view of the above theorem it is preferable to explore the state space with a mutation transition satisfying condition (H) with a mixing parameter $m = 1$. In the genetic interpretation (5) the algorithm evolves according to a two stages mutation/selection. When the mutation transition has a mixing parameter m it seems preferable to run m mutations between each selection procedure, so that we are brought back to the situation $m = 1$. But in practice m can be very large and it may be not efficient to wait too long between mutations (quantitatively, this corresponds to a large constant C_0 in the above theorem), specially for a given number of iteration $n \in \mathbb{N}^*$.

2 Regularity properties

The asymptotic stability of Feynman-Kac type semi-groups has been started in [4]. Although this study does not discuss the contraction properties of the non-linear Feynman-Kac semi-group it designs a semi-group technique which can be easily transfered to obtain contraction properties of the mappings Φ_n^β. Next we present a slightly more general formulation which applies to the annealed Feynman-Kac models and capture the main ideas.

Suppose Q is a given bounded and positive operator on $\mathcal{B}(E)$. We associate to Q the non-linear semi-group Φ_n, $n \in \mathbb{N}$, defined on $\mathcal{P}(E)$ by the formula

$$\forall\ A \in \mathcal{E}, \qquad \Phi_n(\mu)(A) \;=\; \mu Q^n(A)/\mu Q^n(1)$$

Proposition 2.1 *Suppose there exists an integer* $m \geq 1$ *and a collection of numbers* $\epsilon_Q \in (0,1)$ *such that*

$$Q^m(x, \centerdot) \;\geq\; \epsilon_Q\; Q^m(y, \centerdot)$$

Then we have for any μ, η *and* $n \in \mathbb{N}$

$$\|\Phi_n(\mu) - \Phi_n(\eta)\|_{\mathrm{tv}} \;\leq\; 2\epsilon_Q^{-2}\ (1-\epsilon_Q^2)^{\lfloor n/m \rfloor}\ \|\mu - \eta\|_{\mathrm{tv}} \tag{8}$$

Proof:
The contraction inequality (8) is essentially proved in [4]. We give next a short proof which complete the arguments given there. By construction we notice that

$$\Phi_n(\mu)(f) = \mu Q^n(f)/\mu Q^n(1) = \Psi_n(\mu)\overline{Q}_n(f)$$

with the Markov kernel $\overline{Q}_n$ on E and the mapping Ψ_n on $\mathcal{P}(E)$ defined by

$$\overline{Q}_n(f) = \frac{Q^n(f)}{Q^n(1)} \quad \text{and} \quad \Psi_n(\mu)(f) = \frac{\mu(Q^n(1)\ f)}{\mu(Q^n(1))}$$

The sequence of kernels $\overline{Q}_n$ and the mappings Ψ_n do not satisfy the semi-group property but we have for any pair of indexes $p+q=n$

$$\overline{Q}_n = \overline{Q}_p^{(q)}\ \overline{Q}_q$$

with the Markov kernel

$$\overline{Q}_p^{(q)}(f) = \frac{Q^p(Q^q(1)\ f)}{Q^p(Q^q(1))}$$

To see this claim it suffices to use the semi-group property of Q and the definition of $\overline{Q}_n$, indeed we have

$$\overline{Q}_n(f) \quad = \quad \frac{Q^p(Q^q(f))}{Q^p(Q^q(1))} = \frac{Q^p(Q^q(1)\overline{Q}_q(f))}{Q^p(Q^q(1))} = \overline{Q}_p^{(q)}\ \overline{Q}_q$$

¿From this observation we can write

$$\overline{Q}_n \quad = \quad \overline{Q}_m^{(n-m)}\overline{Q}_m^{(n-2m)}\ldots\overline{Q}_m^{(n-\lfloor n/m\rfloor m)}\overline{Q}_{n-\lfloor n/m\rfloor m}$$

Under our assumptions we have for any $q\in\mathbb{N}$, $x,y\in E$ and $n\geq m$

$$\overline{Q}_m^{(q)}(x,\textbf{.})\ \geq\ \epsilon_Q\ \overline{Q}_m^{(q)}(y,\textbf{.}) \quad \text{and} \quad Q^n(1)(x)\ \geq\ \epsilon_Q\ Q^n(1)(y) \tag{9}$$

We recall that for any Markov kernel K such that for any $x,y\in E$

$$K(x,\textbf{.})\ \geq\ \delta\ K(y,\textbf{.})$$

for some $\delta\in(0,1)$ we have the contraction property

$$\|\mu K-\eta K\|_{\rm tv}\ \leq\ (1-\delta)\ \|\mu-\eta\|_{\rm tv}$$

¿From this and the above considerations we obtain

$$\|\mu\overline{Q}_m^{(q)}-\eta\overline{Q}_m^{(q)}\|_{\rm tv}\ \leq\ (1-\epsilon_Q^2)\ \|\mu-\eta\|_{\rm tv}$$

We end the proof of the proposition by noting that the decomposition

$$\Psi_n(\mu)(f)-\Psi_n(\eta)(f) = \frac{\eta(Q^n(1))}{\mu(Q^n(1))}\ (\mu-\eta)\left(\frac{Q^n(1)}{\eta(Q^n(1))}\ (f-\Psi_n(\eta)(f))\right)$$

together with (9) yields that

$$\|\Psi_n(\mu)-\Psi_n(\eta)\|_{\rm tv}\ \leq\ 2\epsilon_Q^{-2}\ \|\mu-\eta\|_{\rm tv}$$

This ends the proof of the proposition.

■

Next we apply proposition 2.1 to determine the degree of contraction of the semi-group Φ_n^β. When a Markov kernel M satisfies condition (H) for some $m\geq 1$ and some $\epsilon\in(0,1)$ we write $\Delta(0)$ the integer

$$\Delta(0) = (1+\lfloor\epsilon^{-2}\rfloor)(4+\lfloor\ln(2\epsilon^{-2})\rfloor)$$

and $\Delta_m(\beta)$, $\beta\in\mathbb{R}_+$, the collection of integers

$$\Delta_m(\beta) = \Delta(0)\ (1+\lfloor e^{\alpha(m)\beta}\rfloor)\ (1+\lfloor\alpha(m)\beta\rfloor) \quad \text{with} \quad \alpha(m)=(m-1)\text{osc}(V)$$

Theorem 2.2 *Suppose the Markov kernel M satisfies condition (H) for some integer $m \geq 1$ and some parameter $\epsilon \in (0,1)$. Then for any $n \geq m$ we have*

$$\|\Phi_n^\beta(\mu) - \Phi_n^\beta(\eta)\|_{\mathrm{tv}} \;<\; 2\epsilon^{-2}\; e^{2\alpha(m)\beta}\; (1-\epsilon^2\; e^{-\alpha(m)\beta})^{\lfloor n/m \rfloor}\; \|\mu-\eta\|_{\mathrm{tv}}$$

In addition we have for any $\beta \in \mathbb{R}_+$

$$\|\Phi^\beta_{\Delta_m(\beta)}(\mu) - \Phi^\beta_{\Delta_m(\beta)}(\eta)\|_{\mathrm{tv}} \;<\; \frac{1}{e}\; \|\mu-\eta\|_{\mathrm{tv}}$$

Thus $\triangle_m(\beta)$ can be seen as a relaxation time for Φ^β with respect to $\|\cdot\|_{\mathrm{tv}}$.

Proof:
Let $Q = MG$ be the composition of the Markov kernel M with the multiplicative kernel $G(f) = e^{-\beta V} f$. For any positive function $f \in \mathcal{B}(E)$ we find that

$$\frac{Q^m(f)(x)}{Q^m(f)(y)} \;=\; \frac{QQ^{m-1}(f)(x)}{QQ^{m-1}(f)(y)} \geq e^{-\beta \mathrm{osc}(V)}\; \frac{MQ^{m-1}(f)(x)}{MQ^{m-1}(f)(y)}$$

and by induction

$$\frac{Q^m(f)(x)}{Q^m(f)(y)} \geq e^{-(m-1)\beta \mathrm{osc}(V)}\; \frac{M^m G(f)(x)}{M^m G(f)(y)}$$

(this can also be seen directly on normalized Feynman-Kac formulae).
Hence under our assumptions we conclude that

$$Q^m(x, .) \geq \epsilon\; e^{-\alpha(m)\beta}\; Q^m(y, .)$$

Thus Q satisfies the mixing condition stated in proposition 2.1 with

$$\epsilon_Q = \epsilon\; e^{-\alpha(m)\beta}$$

Consequently we find that

$$\|\Phi_n^\beta(\mu) - \Phi_n^\beta(\eta)\|_{\mathrm{tv}} \;<\; 2\epsilon^{-2}\; e^{2\alpha(m)\beta}\; (1-\epsilon^2\; e^{-2\alpha(m)\beta})^{\lfloor n/m \rfloor}\; \|\mu-\eta\|_{\mathrm{tv}}$$

The factor 2 in the second exponential term in the last display can be removed by using the same lines of arguments as in the proof of proposition 2.1 and noting that for any positive function $f \in \mathcal{B}(E)$

$$\begin{aligned}\overline{Q}_m^{(q)}(f)(x) = \frac{Q^m(\;Q^q(1)\;f)(x)}{Q^m(\;Q^q(1)\;)(x)} \;&\geq\; e^{-(m-1)\beta \mathrm{osc}(V)}\; \frac{M^m G(Q^q(1)f)(x)}{M^m G(Q^q(1))(x)} \\ &\geq\; \epsilon^2\; e^{-(m-1)\beta \mathrm{osc}(V)}\; \frac{M^m G(Q^q(1)f)(y)}{M^m G(Q^q(1))(y)}\end{aligned}$$

To prove the second assertion we observe that

$$(1-\epsilon^2\; e^{-\alpha(m)\beta})^{\Delta_m(\beta)} \leq e^{-\epsilon^2\;\Delta_m(\beta)\; e^{-\alpha(m)\beta}} \leq e^{-\epsilon^2\;\Delta(0)(1+\lfloor \alpha(m)\beta \rfloor)}$$

Since we have

$$\begin{aligned}\epsilon^2\Delta(0)(1+\lfloor\alpha(m)\beta\rfloor) &= \epsilon^2(1+\lfloor\epsilon^{-2}\rfloor)\ \ (4+\lfloor\ln{(2\epsilon^{-2})}\rfloor)\ \ (1+\lfloor\alpha(m)\beta\rfloor)\\ &\geq (4+\lfloor\ln{(2\epsilon^{-2})}\rfloor+2\lfloor\alpha(m)\beta\rfloor)\\ &\geq 1+\ln{(2\epsilon^{-2})}+2\alpha(m)\beta\end{aligned}$$

we conclude that

$$2\epsilon^{-2}\ e^{2\alpha(m)\beta}\ (1-\epsilon^2\ e^{-\alpha(m)\beta})^{\Delta_m(\beta)}\leq 1/e$$

The end of the proof is now straighforward.

■

The uniform contraction property stated in theorem 2.2 is an important first step in proving the convergence theorem 1.1. First, and as mentioned in section 1.2, it guarantees the existence of a unique collection of fixed point probability measures

$$\mu^\beta=\Phi^\beta(\mu^\beta)\in\mathcal{P}(E)\,,\qquad\beta\in\mathbb{R}_+$$

Furthermore it is also an important instrument tool for proving the following regularity condition.

Proposition 2.3 *When condition (H) holds true for some $m\geq 1$ and $\epsilon\in(0,1)$ we have for any $0\leq\beta_1\leq\beta_2$*

$$\|\mu^{\beta_1}-\mu^{\beta_2}\|_{\rm tv}\leq \mathrm{osc}(V)\Delta_m(\beta_1)\ (\beta_2-\beta_1)$$

Proof:
Using the fixed point property we have the decomposition

$$\mu^{\beta_1}-\mu^{\beta_2}=\Phi^{\beta_1}_{\Delta_m(\beta_1)}(\mu^{\beta_1})-\Phi^{\beta_1}_{\Delta_m(\beta_1)}(\mu^{\beta_2})+\Phi^{\beta_1}_{\Delta_m(\beta_1)}(\mu^{\beta_2})-\Phi^{\beta_2}_{\Delta_m(\beta_1)}(\mu^{\beta_2})$$

By the uniform property stated in theorem 2.2 we notice that

$$\|\mu^{\beta_1}-\mu^{\beta_2}\|_{\rm tv}\leq\frac{e}{e-1}\ \|\Phi^{\beta_1}_{\Delta_m(\beta_1)}(\mu^{\beta_2})-\Phi^{\beta_2}_{\Delta_m(\beta_1)}(\mu^{\beta_2})\|_{\rm tv}$$

Let $\mathbb{P}_{\mu,n}$ be the distribution of the sequence of random variables $(X_0,\ldots,X_n)$ with initial distribution μ, that is

$$\mathbb{P}_{\mu,n}(d(x_0,\ldots,x_n))=\mu(dx_0)\ M(x_0,dx_1)\ \ldots\ M(x_{n-1},dx_n)$$

In this notations we see that that each distribution $\Phi^\beta_n(\mu)$ is the n-time marginal of the Gibbs-Boltzmann measure on E^{n+1} defined by

$$\mathbb{P}^\beta_{\mu,n}(dx)=\frac{1}{\mathbb{P}_{\mu,n}(e^{-\beta V_n})}\ e^{-\beta V_n(x)}\ \mathbb{P}_{\mu,n}(dx)$$

with the potential V_n from E^{n+1} into $\mathbb{R}_+$ defined for any $x=(x_0,\ldots,x_n)$ by

$$V_n(x)=\sum_{p=1}^n V(x_p)$$

It is now well-known that for any $\beta_1 \leq \beta_2$

$$\|\mathbb{P}^{\beta_1}_{\mu,n} - \mathbb{P}^{\beta_2}_{\mu,n}\|_{\mathrm{tv}} \leq \frac{\beta_2 - \beta_1}{2}\mathrm{osc}(V_n) \leq n\frac{\beta_2 - \beta_1}{2}\ \mathrm{osc}(V)$$

from which we conclude that

$$\|\Phi^{\beta_1}_{\Delta_m(\beta_1)}(\mu^{\beta_2}) - \Phi^{\beta_2}_{\Delta_m(\beta_1)}(\mu^{\beta_2})\|_{\mathrm{tv}} \quad \leq \quad \frac{\Delta_m(\beta_1)}{2}(\beta_2 - \beta_1)\ \mathrm{osc}(V)$$

This ends the proof of the proposition, due to the elementary bound $e \leq 2(e-1)$.

■

3 Asymptotic behavior

In the further development of this section we assume without further mention that the Markov kernel satisfies condition (H) for some integer $m \geq 1$ and some $\epsilon \in (0,1)$. Let $\beta = (\beta(n))_{n\in\mathbb{N}}$ be a given non-decreasing inverse cooling schedule and let $t_m(n)$, $n \in \mathbb{N}$, be the associated time mesh defined by the following recursive formula

$$t_m(n+1) = t_m(n) + \Delta_m(\beta(n)) \quad \text{with} \quad t_m(0) = 0$$

We associate to the pair $(\beta(.), t_m(.))$ the annealed Feynman-Kac flow μ_p, $p \in \mathbb{N}$, defined for each $n \in \mathbb{N}$ by

$$\mu_{p+1} = \Phi^{\beta(n)}(\mu_p) \quad \text{for each} \quad t_m(n) \leq p < t_m(n+1)$$

In other words μ_p is the annealed Feynman-Kac flow with a constant inverse temperature parameter $\beta(n)$ between the dates $t_m(n)$ and $t_m(n+1)$, that is for each $0 \leq p < t_m(n+1) - t_m(n)$

$$\mu_{t_m(n)+p}(f) = \frac{\mathbb{E}_{\mu_{t_m(n)}}(f(X_p)\ e^{-\beta(n)\sum_{q=1}^{p} V(X_q)})}{\mathbb{E}_{\mu_{t_m(n)}}(e^{-\beta(n)\sum_{q=1}^{p} V(X_q)})}$$

The core idea in the study of the long time behavior of the annealed model consists in combining the regularity properties of the fixed point distributions μ^β with the contraction properties of the mappings Φ^β. To this end we introduce the decomposition

$$\mu_{t_m(n)} - \mu^{\beta(n)} = (\Phi^{\beta(n-1)}_{\Delta_m(\beta(n-1))}(\mu_{t_m(n-1)}) - \Phi^{\beta(n-1)}_{\Delta_m(\beta(n-1))}(\mu^{\beta(n-1)})) + (\mu^{\beta(n-1)} - \mu^{\beta(n)})$$

¿From this display we now apply theorem 2.2 and proposition 2.3 to get the following system of inequalities

$$\begin{aligned}
&\|\mu_{t_m(n)} - \mu^{\beta(n)}\|_{\mathrm{tv}} \\
&\leq \ \frac{1}{e}\ \|\mu_{t_m(n-1)} - \mu^{\beta(n-1)}\|_{\mathrm{tv}} + \mathrm{osc}(V)\ \ \Delta_m(\beta(n-1))\ (\beta(n) - \beta(n-1))
\end{aligned}$$

and thus it appears that

$$
\begin{aligned}
& e^n \, \|\mu_{t_m(n)} - \mu^{\beta(n)}\|_{\mathrm{tv}} \\
& \quad \leq \; \|\mu_0 - \mu^{\beta(0)}\|_{\mathrm{tv}} + \mathrm{osc}(V) \sum_{p=1}^{n} e^p \; \Delta_m(\beta(p-1)) \; (\beta(p) - \beta(p-1)) \qquad (10)
\end{aligned}
$$

We are now in position to state the main result of this section.

Theorem 3.1

- *If $m = 1$ then we have $t_1(n) = \Delta(0)n$ and we can choose*

$$\beta(t) = (t+1)^a \quad , \textit{for any fixed} \quad a \in (0,1)$$

In addition we have for some $c(a) < \infty$

$$\|\mu_{t_1(n)} - \mu^{\beta(n)}\|_{\mathrm{tv}} \leq c(a)/n^{1-a} \qquad (11)$$

- *If $m > 1$ then we can take*

$$\beta(t) = \beta(0) \ln(t+e) \quad , \text{with} \quad b = \alpha(m)\beta(0) < 1$$

In this case we have $t_m(n) = \mathcal{O}(n^{b+1} \ln n)$ and for some $c(b) < \infty$

$$\|\mu_{t_m(n)} - \mu^{\beta(n)}\|_{\mathrm{tv}} \leq c(b) \; \ln(n+e) \; /n^{1-b}$$

Proof:
When $m = 1$ we recall that $\Delta_1(\beta) = 2\Delta(0)$ does not depend on β and $t_1(n) = 2\Delta(0)n$. By direct inspection, if we choose $\beta(t) = (t+1)^a$, for any fixed $a \in (0,1)$, then we find from (10) that

$$e^n \, \|\mu_{t_1(n)} - \mu^{\beta(n)}\|_{\mathrm{tv}} \quad \leq \quad \|\mu_0 - \mu^{\beta(0)}\|_{\mathrm{tv}} + 2\,\mathrm{osc}(V) \; \Delta(0) \; \sum_{p=1}^{n} e^p \; ((p+1)^a - p^a)$$

Recalling that $x^a - y^a \leq a y^{a-1} \; (x - y)$ for any $x, y \geq 0$, we get

$$\sum_{p=1}^{n} e^p \; ((p+1)^a - p^a) \leq a \sum_{p=1}^{n} \frac{e^p}{p^{1-a}} \leq ae \left(1 + \sum_{p=2}^{n} \frac{e^{p-1}}{p^{1-a}} \right)$$

Next we observe that for any $p \geq 2$,

$$\frac{e^{p-2}}{(p-1)^{1-a}} = \frac{1}{e} \frac{e^{p-1}}{p^{1-a}} \frac{p^{1-a}}{(p-1)^{1-a}} \leq \frac{2^{1-a}}{e} \frac{e^{p-1}}{p^{1-a}} \leq \frac{2}{e} \frac{e^{p-1}}{p^{1-a}}$$

so that we have

$$
\begin{aligned}
\sum_{p=2}^{n} \frac{e^{p-1}}{p^{1-a}} \quad & \leq \quad (1 - 2/e)^{-1} \sum_{p=2}^{n} \frac{e^{p-1}}{p^{1-a}} - \frac{e^{p-2}}{(p-1)^{1-a}} \\
& \leq \quad 2\, e \frac{e^{n-1}}{n^{1-a}}
\end{aligned}
$$

It follows that

$$\|\mu_{t_1(n)} - \mu^{\beta(n)}\|_{\rm tv} \;\;\leq\;\; e^{-n} + 6\,{\rm osc}(V)\;\Delta(0)\;\left(e^{-n} + \frac{2}{n^{1-a}}\right)$$

Since $e^n \geq n^{1-a}$, this yields that

$$\|\mu_{t_1(n)} - \mu^{\beta(n)}\|_{\rm tv} \;\;\leq\;\; c(a)/n^{1-a}$$

with

$$c(a) \leq 1 + 18\,{\rm osc}(V)\;\Delta(0)$$

Now we examine the situation where condition (H) is met only for some integer $m > 1$. We begin in this case by observing that

$$\Delta_m(\beta) \;\;\leq\;\; 4\,\Delta(0)\;e^{\alpha(m)\beta}\;(2 + \alpha(m)\beta)$$

We now choose the cooling schedule $\beta(t) = \beta(0)\ln(t+e)$ with $b = \alpha(m)\beta(0) < 1$. In this notation, we find that from the above observation that

$$\begin{aligned}\Delta_m(\beta(n)) &\leq 4\;\Delta(0)\;(e+n)^b\;(2 + b\ln(n+e))\\ &\leq 16\,\Delta(0)\;(e+n)^b\;\ln(n+e)\end{aligned} \tag{12}$$

This yields the following growth estimate on the corresponding time mesh

$$\begin{aligned}t_m(n) &= \sum_{p=0}^{n-1}\Delta_m(\beta(p)) \leq 16e^b\;\Delta(0)\;\ln(n+e)\sum_{p=1}^{n}\;p^b\\ &\leq \frac{16\,e^b}{b+1}\;\Delta(0)\;(n+1)^{b+1}\;\ln(n+e)\end{aligned}$$

We deduce from (10) and (12) that

$$\begin{aligned}&e^n\;\|\mu_{t_m(n)} - \mu^{\beta(n)}\|_{\rm tv}\\ &\leq 1 + 32\,{\rm osc}(V)\beta(0)\Delta(0)\;\textstyle\sum_{p=0}^{n-1}\;\ln(p+e)\;\frac{e^{p+1}}{(p+e)^{1-b}}\end{aligned}$$

This implies that

$$e^n\;\|\mu_{t_m(n)} - \mu^{\beta(n)}\|_{\rm tv} \leq 1 + 32\,{\rm osc}(V)\beta(0)\Delta(0)\;\ln(n+e)\;\sum_{p=1}^{n}\frac{e^p}{p^{1-b}}$$

¿From the above estimates we have

$$e^{-n}\;\sum_{p=1}^{n}\frac{e^p}{p^{1-b}} \leq e^{-n}e\left(1 + 2\frac{e^n}{n^{1-b}}\right) \leq \frac{9}{n^{1-b}}$$

This finally shows that

$$\|\mu_{t_m(n)} - \mu^{\beta(n)}\|_{\rm tv} \leq c(b)\;\ln(n+e)\;\frac{1}{n^{1-b}}$$

with

$$c(b) = 1 + 288\,{\rm osc}(V)\beta(0)\Delta(0)$$

■

4 Concentration Properties

This section is decomposed in two parts. In a first sub-section we provide a variational formulation of the concentration properties of μ^β. We use a large deviation analysis to connect the Lyapunov exponent of the underlying Feynman-Kac semigroup at temperature $\beta^{-1} > 0$ with the β-exponential moment of the potential V under μ^β. The description of the concentration level V_I is discussed in the second part of this section. We give sufficient conditions on the Markov kernel M underwhich V_I coincides with the essential infimum V_ν of the potential. In the final part of the article we examine the finite state space situation. We provide an alternative formulation of V_I in terms of the minimal mean potential along M-admissible cycles in E. We also examine the two situations where $V_I = V_\nu$ and $V_I > V_\nu$.

4.1 Large deviation analysis

In this section we examine the concentration properties of the fixed point distributions μ^β as β tends to infinity. To obtain a better grasp of what is at stake it is useful to interpret these distributions as the limiting measures of the canonical Markov chain Z under the McKean distribution $\mathbb{P}^\beta_{\mu_0}$ defined in (6). We recall that this distribution describes the evolution of a non-homogeneous Markov chain with elementary transitions

$$\mathbb{P}^\beta_{\mu_0}(Z_n \in dy \mid Z_{n-1} = x) = MS^\beta_{\mu_{n-1}M}(x, dy) \quad \text{with} \quad \mathbb{P}^\beta_{\mu_0} \circ Z^{-1}_{n-1} = \mu_{n-1}$$

This chain can be regarded as a Feynman-Kac simulated annealing model. At each time n the particle Z_{n-1} first evolves according to the transition kernel M to a new location Y_n. With a probability $e^{-\beta V(Y_n)}$ we set $Z_n = Y_n$ and with a probability $1 - e^{-\beta V(Y_n)}$ it jumps to a new location Z_n randomly chosen with the Boltzmann-Gibbs distribution

$$\Psi^\beta(\eta_n)(dy) = \frac{1}{\eta_n(e^{-\beta V})}\, e^{-\beta V(y)}\, \eta_n(dy) \quad \text{with} \quad \eta_n = \mu_{n-1}M = \mathbb{P}^\beta_{\mu_0} \circ Y_n^{-1}$$

These transitions tend to favor regions with low potential values. The precise description of these areas is contained in the concentration properties of the limiting distributions μ^β. In contrast to what would be the case for traditional simulated annealing or statistical mechanics models here the limiting distributions are not defined in terms of a Boltzman-Gibbs measure. As a result the concentration analysis is more involved and we have to find a new strategy to enter into these questions. Here the interplay between μ^β and the quantities (β, M, V) is only described by the fixed point formula

$$\forall\, f \in \mathcal{B}(E), \qquad \mu^\beta(f) \;=\; \mu^\beta(Q_\beta(f))/\mu^\beta(Q_\beta(1)) \quad \text{with} \quad Q_\beta(f) = M(e^{-\beta f})$$

As we already mentioned in the introduction under the uniform mixing condition (H) the Markov kernel M has a unique invariant measure

$$\nu = \nu M \in \mathcal{P}(E)$$

and the sequence of occupation measures $L_n = \frac{1}{n}\sum_{p=1}^n \delta_{X_p}$ of the chain X under $\mathbb{P}_{\mu_0}$ satisfies as $n \to \infty$ a large deviation principle with good rate function

$$I(\mu) = \inf \ \{\int_E \ \mu(dx)\ \mathrm{Ent}(K(x,\centerdot)|M(x,\centerdot))\} \tag{13}$$

where the infimum is taken over all Markov kernels K with invariant measure μ (cf [8]). In the most naive view we could think that the Feynman-Kac simulated annealing model converges in probability to the ν-essential infimum V_ν of the potential V (since under (H) for any $x_0 \in E$, $M^m(x_0,\cdot)$ is equivalent to ν, it is not really necessary to know the latter probability to compute V_ν). This intuitive idea appears to be true for regular Markov transitions M with a diagonal term $M(x,x) > 0$ but it is false in more general situations. To better introduce our strategy to study the concentration properties of μ^β we need a more physical interpretation of the Feynman-Kac models. If we interpret the potential V as the absorption rate for a Markov particle with transition M evolving in an medium with obstacles, the normalizing constants

$$\mathbb{E}_\mu(e^{-\beta\sum_{p=1}^n V(X_p)})$$

represent the probabilities of a killed Markov particle starting with distribution μ and conditioned to perform a long crossing of length n without being trapped. The cost attached to performing long crossings is measured in terms of the logarithmic Lyapunov exponents of the semi-group Q_β on the Banach space $\mathcal{B}(E)$

$$\Lambda(-\beta V) \ = \ \lim_{n\to\infty}\frac{1}{n}\ln ||Q_\beta^n(1)|| = \lim_{n\to\infty}\frac{1}{n}\ln\sup_x \mathbb{E}_x(e^{-\beta\sum_{p=1}^n V(X_p)})$$

Next lemma shows that these Lyapunov exponents coincide with the logarithmic rate of the β-exponential moment of the fixed point measures μ^β. It also makes a link between the large deviation rate I and the concentration properties of μ^β. Informally it shows that

$$\mu^\beta(e^{\beta V}) \simeq e^{\beta V_I}$$

where V_I is the value of the variational problem

$$V_I \ := \ \inf\{\mu(V) \quad \mu \in \mathcal{P}(E) \ : \ I(\mu) < \infty\} \tag{14}$$

Loosely speaking the concentration properties of the limiting measures μ^β as β tends to infinity are related to a competition in $\mathcal{P}(E)$ between the mean potential $\mu(V)$ and the I-entropy $I(\mu)$. The next lemma also shows that the concentration of μ^β is related to a variational problem in which the competition with the entropy I becomes less and less severe as β tends to infinity.

Lemma 4.1 *For any $\beta \in \mathbb{R}_+$ we have the formulae*

$$-\frac{\Lambda(-\beta V)}{\beta} = \frac{1}{\beta}\ln\mu^\beta(e^{\beta V}) = \inf_{\eta\in\mathcal{P}(E)}\left(\eta(V) + \frac{1}{\beta}I(\eta)\right) \xrightarrow[\beta\to\infty]{} V_I \ \geq V_\nu$$

Proof:
If we take $f = Q_\beta^n(1)$ in the fixed point equality we readily find the recursive formula

$$\mu^\beta(Q_\beta^{n+1}(1)) = \mu^\beta(Q_\beta^n(1))\ \mu^\beta(Q_\beta(1))$$

Thus we have for each $n \geq 0$

$$\mu^\beta(Q_\beta^n(1)) = (\mu^\beta(Q_\beta(1)))^n = \mathbb{E}_{\mu^\beta}(e^{-\beta \sum_{p=1}^n V(X_p)}) \tag{15}$$

Now if we take $f = e^{\beta V}$ in the fixed point equation we get

$$\mu^\beta(e^{\beta V})\ \mu^\beta(Q_\beta(1)) = 1 \tag{16}$$

Recalling that under condition (H) the Laplace transformation

$$\Lambda(-\beta V) = \lim_{n\to\infty} \frac{1}{n} \ln \mathbb{E}_\mu(e^{nL_n(-\beta V)})$$

does not depend on the choice of the initial distribution μ we deduce that

$$-\Lambda(-\beta V) = -\ln \mu^\beta(Q_\beta(1)) = \ln \mu^\beta(e^{\beta V})$$

Since $\Lambda(-\beta V)$ is also given as the Fenchel transformation of I

$$\Lambda(-\beta V) \quad = \sup_{\eta\in\mathcal{P}(E)} (\eta(-\beta V) - I(\eta)) \tag{17}$$

the end of the proof of the first assertion is clear. To end the proof we notice that

$$V_I \leq \inf_{\eta\in\mathcal{P}(E)} (\eta(V) + \frac{1}{\beta} I(\eta)) \leq \eta_0(V) + \frac{1}{\beta} I(\eta_0)$$

for each distribution η_0 such that $I(\eta_0) < \infty$. Letting $\beta \to \infty$ we find that

$$V_I \leq \limsup_{\beta\to\infty} \inf_{\eta\in\mathcal{P}(E)} (\eta(V) + \frac{1}{\beta} I(\eta)) \leq \eta_0(V)$$

Taking the infimum over all distributions η_0 such that $I(\eta_0) < \infty$ we obtain

$$\lim_{\beta\to\infty} \frac{1}{\beta} \ln \mu^\beta(e^{\beta V}) = V_I$$

To see that $V_I \geq V_\nu$, it is clearly sufficient to show that for any probability μ, $I(\mu) < +\infty$ implies that $\mu \ll \nu$. One easy way to obtain this assertion in our context is to note that if $I(\mu) < +\infty$, then there exists a kernel K verifying $\mu = \mu K$ and $K(x,\cdot) \ll M(x,\cdot)$ for μ-a.s. all $x \in E$. But since for all $x \in E$, $M^m(x,\cdot)$ is equivalent to ν, due to the hypothesis (H), we get that $\mu = \mu K^m \ll \mu M^m \sim \nu$. This ends the proof of the lemma.

$\blacksquare$

By using the exponential version of Markov's inequality, lemma 4.1 provides a concentration property of μ^β in the level sets $(V < V_I + \delta)$, $\delta > 0$. More precisely we have for any $\delta > 0$

$$\begin{aligned} \mu^\beta(V \geq V_I + \delta) &= \mu^\beta(e^{\beta(V-V_I)} \geq e^{\beta\delta}) \\ &\leq e^{-\beta\delta}\ \mu^\beta(e^{\beta(V-V_I)}) \end{aligned}$$

¿From which one concludes that

$$\lim_{\beta\to\infty} \frac{1}{\beta} \ln \mu^\beta(V \geq V_I + \delta) \leq -\delta$$

Combining this concentration property with theorem 3.1 we prove the following asymptotic convergence result.

Proposition 4.2 *Suppose condition (H) holds true for some $m \geq 1$ and let $t_m(n)$ and $\beta(n)$ be respectively the time mesh sequence and the cooling schedule described in theorem 3.1. Then the corresponding annealed Feynman-Kac distribution flow $\mu_{t_m(n)}$ concentrates as $n \to \infty$ to regions with potential less than V_I, that is for each $\delta > 0$ we have that*

$$\lim_{n\to\infty} \mu_{t_m(n)}(V \geq V_I + \delta) = 0$$

Remark 4.3: The topological hypotheses that E is Polish and that V is continuous are only necessary to obtain (17), see for instance [6]. So except for the definition (14), the rest of the paper is true under the assumption that $(E, \mathcal{E})$ is a measurable space and that V is a non-negative bounded and measurable potential. In particular under this extended setting we can consider

$$V_* \ := \ -\lim_{\beta\to+\infty} \frac{1}{\beta} \lim_{n\to\infty} \frac{1}{n} \ln\left(\mathbb{E}_x[\exp(-\beta V(X_1) - \cdots - \beta V(X_n))]\right)$$

which always exists and does depend on the initial condition $x \in E$.
Indeed, if we denote

$$\forall\, n \in \mathbb{N}, \forall\, \beta \in \mathbb{R}_+, \qquad \lambda_n(\beta) \ = \ \inf_{x\in E} \ln(\mathbb{E}_x[\exp(-\beta V(X_1) - \cdots - \beta V(X_n))])$$

then it is quite clear via the Markov property that $(\lambda_n(\beta))_{n\in\mathbb{N}}$ is superadditive, so that the following limit exists

$$\lambda(\beta) \ := \ \lim_{n\to\infty} \frac{1}{n} \lambda_n(\beta)$$

(this is just a rewriting of the traditional existence of the Lyapunov exponent of the underlying unnormalized Feynman-Kac operator). Now taking into account condition (H), it appears that for any $n \geq m$ and $x, y \in E$,

$$\begin{aligned} \epsilon^2 \exp(-(m-1)\beta\, \mathrm{osc}(V)) &\leq \frac{\mathbb{E}_y[\exp(-\beta V(X_1) - \cdots - \beta V(X_n))]}{\mathbb{E}_x[\exp(-\beta V(X_1) - \cdots - \beta V(X_n))]} \\ &\leq \epsilon^{-2} \exp((m-1)\beta\, \mathrm{osc}(V)) \end{aligned}$$

thus we see that

$$\lim_{n\to\infty}\frac{1}{n}\ln\left(\frac{\mathbb{E}_y[\exp(-\beta V(X_1)-\cdots-\beta V(X_n))]}{\mathbb{E}_x[\exp(-\beta V(X_1)-\cdots-\beta V(X_n))]}\right)=0$$

and in particular for any initial distribution μ_0, we have

$$\lambda(\beta) \quad = \quad \lim_{n\to\infty}\frac{1}{n}\ln(\mathbb{E}_{\mu_0}[\exp(-\beta V(X_1)-\cdots-\beta V(X_n))])$$

Besides the lhs is convex in β, as a limit of convex functions, so we are assured of the existence of

$$-\lim_{\beta\to+\infty}\frac{\lambda(\beta)}{\beta} \quad = \quad -\lim_{\beta\to+\infty}\frac{\lambda(\beta)-\lambda(0)}{\beta} \quad = \quad -\sup_{\beta>0}\frac{\lambda(\beta)-\lambda(0)}{\beta}$$

a priori in $\mathbb{R}\sqcup\{-\infty\}$, but as V is non-negative and bounded, we conclude that $V_*\in\mathbb{R}_+$.
In this context, lemma 4.1 can be rewritten as saying that under the topological hypotheses that E is Polish and that V is continuous, we have $V_* = V_I \geq V_\nu$.

4.2 Concentration levels

This section is devoted to a discussion on the concentration regions of μ^β as β tends to infinity.
In a first subsection we examine Feynman-Kac models where the Markov kernel M satisfies condition (H) with $m=1$ or has a regular diagonal term. We show that in this case the concentration level V_I coincide with the essential infimum of the potential with respect to the invariant measure of M.
The second part of this section focuses on Feynman-Kac models on finite state spaces. We relate the exponential concentration of μ^β with a collection of Bellman's fixed point equations. We propose an alternative characterization of the concentration level V_I. We show that V_I can be seen as the minimal mean potential value over all closed cycles on E. Thanks to this representation we will check that $V_I = V_\nu$ iff there exists a closed cycle on $V^{-1}(V_\nu)$. For more general off-diagonal mutation transitions we have $V_I > V_\nu$. We illustrate this assertion with a simple three point example, showing furthermore that μ^β is not concentrating on “neighborhoods” of $V^{-1}(V_\nu)$.

4.2.1 Diagonal mutations

Certainly the easiest way to insure that $V_I = V_\nu$ is to impose loops on every point of E for M. This assertion is based on the following simple upper bound.

Proposition 4.4 *Let $x_0\in E$ be such that $M(x_0,x_0)>0$, then we have $V_I\leq V(x_0)$.*

Proof:
By definition of the Markov chain X, we have that for any $\beta\in\mathbb{R}$ and any $n\in\mathbb{N}$,

$$\begin{aligned}
&\mathbb{E}_{x_0}[\exp(-\beta V(X_1)-\cdots-\beta V(X_n))]\\
&\quad\geq\ \mathbb{E}_{x_0}[\mathbb{1}_{X_1=x_0,\,X_2=x_0,\,\cdots,\,X_n=x_0}\exp(-\beta V(X_1)-\cdots-\beta V(X_n))]\\
&\quad=\ \{M(x_0,x_0)\}^n\exp(-n\beta V(x_0))
\end{aligned}$$

thus

$$\begin{aligned}\Lambda(-\beta V) &\geq \lim_{n\to\infty}\frac{1}{n}\ln\left(\mathbb{E}_{x_0}[\exp(-\beta\sum_{p=1}^{n}V(X_p))]\right)\\ &\geq \ln(M(x_0,x_0))-\beta V(x_0)\end{aligned}$$

so that

$$\begin{aligned}V_I &= -\lim_{\beta\to+\infty}\frac{\Lambda(-\beta V)}{\beta}\\ &\leq V(x_0)\end{aligned}$$

■

As a corollary, we get that

$$V_I \leq \inf_{x\in E_M} V(x)$$

where $E_M := \{x\in E\,:\,M(x,x)>0\}$.
In particular, to be sure that $V_I = V_\nu$, it is sufficient that there exists a sequence $(x_n)_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty} V(x_n) = V_\nu$ and verifying $M(x_n,x_n)>0$ for all $n\in\mathbb{N}$. In practice, this can be insured by imposing on the underlying exploration kernel M that for all $x\in E$, $M(x,x)>0$.
But let us also remark that when condition (H) is satisfied for $m=1$, this slight precaution is even useless, since we have automatically $V_I = V_\nu$. Indeed, this is an immediate consequence of the next ordering of measures, for any $\beta\in\mathbb{R}_+$,

$$\epsilon^2\,\frac{\nu(e^{-\beta V}\cdot)}{\nu(e^{-\beta V})} \leq \mu^\beta(\cdot) \leq \frac{1}{\epsilon^2}\,\frac{\nu(e^{-\beta V}\cdot)}{\nu(e^{-\beta V})}$$

itself deduced from the fixed point formula.
As already mentioned at the end of section 1, this suggests that in implementing genetic algorithms, it is better to wait a certain number of mutation steps (here m) before proceeding to a selection step. We note the same is true for simulated annealing algorithms (cf for instance [1]).

4.2.2 Finite state space

We consider here the simpler case of a finite state space E endowed with an irreducible Markov kernel M. Then the unique associated invariant probability ν charges all points of E. We note that in this context the condition (H) is equivalent to the aperiodicity of M, but we won't require this property for our first result. More precisely, our next objective is to give an explicit representation of V_I, which in this setting is just given by

$$V_I = -\lim_{\beta\to+\infty}\frac{1}{\beta}\lim_{n\to\infty}\frac{1}{n}\ln\left(\mathbb{E}_x[\exp(-\beta V(X_1)-\cdots-\beta V(X_n))]\right) \qquad (18)$$

We have already seen in remark 4.3 that this definition does not depend on the starting point $x \in E$, which could be replaced by any initial distribution.
We will say that a finite sequence of elements of E, $C = (x_1, \cdots, x_n)$, where the length $n \in \mathbb{N}^*$ will be denoted in what follows by $l(C)$, is a proper cycle (relatively to M), if for any $1 \le i \le n$, $M(x_i, x_{i+1}) > 0$, with the convention that $x_{n+1} = x_1$, and if all the x_i, $1 \le i \le n$, are distinct. We denote $\mathcal{C}$ the finite set of all such admissible proper cycles. To any $C = (x_1, \cdots, x_n) \in \mathcal{C}$, we associate its mean potential

$$V(C) \ := \ \frac{1}{n} \sum_{1 \le i \le n} V(x_i)$$

Proposition 4.5 *We have*

$$V_I \ = \ \min_{C \in \mathcal{C}} V(C)$$

In particular it appears that the equality $V_I = V_\nu$ is equivalent to the existence of a proper cycle inside $V^{-1}(V_\nu)$.

Proof:
Denoting by $V_{\mathcal{C}}$ the rhs in the above proposition, we begin by showing that $V_I \ge V_{\mathcal{C}}$. A finite collection $P = (y_1, ..., y_n)$ of elements of E is called a path (relatively to M) if for any $1 \le i < n$, $M(y_i, y_{i+1}) > 0$ and as before we associate to this object its length $l(P) = n \in \mathbb{N}$ ($n = 0$ corresponds to the empty path, note that this length was not permitted for cycles) and its mean potential $V(P) = \sum_{1 \le i \le n} V(y_i)/n$.
If such a path is given, we can find k proper cycles C_1, ..., C_k, and a subpath R of P (this does not mean that it is a subsegment, ie R is not necessarily of the form $(y_r, y_{r+1}, ..., y_{r+l(R)})$) of length less than card(E) such that

$$l(P)V(P) \ = \ \sum_{1 \le i \le k} l(C_i)V(P_i) + l(R)V(R) \tag{19}$$

To be convinced of the existence of such a decomposition, we look for the first return of the path P on itself: let $s = \min\{t \ge 2 \,:\, y_t \in \{y_1, ..., y_{t-1}\}\}$ and $1 \le r < s$ be such that $y_s = y_r$. Then we define $C_1 := (y_r, y_{r+1}, ..., y_{s-1})$ and we consider the new path $P' := (y_1, ..., y_{r-1}, y_s, y_{s+1}, ..., y_n)$ (one would have noted that $M(y_{r-1}, y_s) > 0$). Applying next recursively the previous procedure, we construct/remove the proper cycles C_2, ... , C_k and we end up with a path R whose elements are all different.
¿From formula (19), we deduce that

$$\begin{aligned} l(P)V(P) \ &\ge \ \sum_{1 \le i \le k} l(C_i)V_{\mathcal{C}} - \mathrm{card}(E) \left\| V \right\|_\infty \\ &\ge \ l(P)V_{\mathcal{C}} - 2\mathrm{card}(E) \left\| V \right\|_\infty \end{aligned}$$

Thus for any $x \in E$ and $n \in \mathbb{N}^*$, we have

$$\mathbb{E}_x[\exp(-\beta V(X_1) - \cdots - \beta V(X_n))] \ \le \ \exp(n\beta V_{\mathcal{C}} - 2\mathrm{card}(E)\beta \left\| V \right\|_\infty)$$

and the announced bound follows at once.

To see the reciprocal inequality, let us consider $C \in \mathcal{C}$ such that $V(C) = V_{\mathcal{C}}$. If an initial point x and a large enough length n are given, we construct a path P_n by first going from x to a point of C by a self-avoiding path (whose existence is insured by irreducibility) and next always following C (in the direction included in its definition and jumping from its last element to the first one). Then it is quite clear that $\lim_{n\to\infty} V(P_n) = V(C)$, thus denoting $q = \min_{x,y\in E\,:\,M(x,y)>0} M(x,y)$ and taking into account the bound

$$\mathbb{E}_x[\exp(-\beta V(X_1) - \cdots - \beta V(X_n))] \quad \geq \quad q^n \exp(n\beta V(P_n))$$

we conclude by a similar argument to the one given in the proof of proposition 4.5.

■

In fact the equality of the previous proposition is also true in the case of a Markov kernel M admitting a unique recurrence class (but in this situation ν does not necessarily charges all points of E). In the most general case, the initial point x in (18) plays a role: $V_I(x)$ is the minimal mean potential of proper cycles included in the recurrence classes which can be reach from x.

Remark 4.6: In view of the above result, it appears that if we note $\mathcal{A}_{\mathcal{C}}$ the set of positive functions f defined on E which are of the form $f = \sum_{C\in\mathcal{C}} a_C \mathbb{1}_C$, where we have identified proper cycles C with the subset of their elements and where $(a_C)_{C\in\mathcal{C}}$ is a family of non-negative reals, we have

$$V_I \quad = \quad \inf\left\{\frac{\nu(fV)}{\nu(f)}\ ;\ f \in \mathcal{A}_{\mathcal{C}}\right\}$$

This expression should be compared with the general formula for V_ν:

$$V_\nu \quad = \quad \inf\left\{\frac{\nu(fV)}{\nu(f)}\ ;\ f \in \mathcal{A}_+\right\}$$

where $\mathcal{A}_+$ is the set of positive bounded measurable functions defined on $(E, \mathcal{E})$.

To understand precisely concentration phenomenon for μ^β, it would be very interesting to obtain a large deviation principle: there exists a function $U\ :\ E \to \mathbb{R}_+$ such that

$$\forall\ x \in E, \qquad U(x) \quad := \quad -\lim_{\beta\to+\infty} \ln(\mu^\beta(x))/\beta$$

(necessarily $\min_E U = 0$, in analogy with generalized simulated annealing, U could be called theunderlying virtual energy).
Unfortunatly we have not been able to prove such a convergence, even under the condition (H), but we are still trying to get this result. Nevertheless, we note that under the latter hypothesis the family of mappings $(\ln(\mu^\beta(\cdot))/\beta)_{\beta\geq 1}$ is compact; indeed, let $m \geq 1$ and $\epsilon > 0$ be as in (H), we have for any $\beta > 0$ and $x \in E$,

$$\mu^\beta(x) = \frac{\mu^\beta(Q_\beta^m(\mathbb{1}_{\{x\}}))}{\mu^\beta(Q_\beta^m(1))} \geq \epsilon^2 e^{-(m-1)\beta\mathrm{osc}(V)} \frac{\nu(e^{-\beta V}\mathbb{1}_{\{x\}})}{\nu(e^{-\beta V})} \geq \epsilon^2 e^{-m\beta\mathrm{osc}(V)}\nu(x)$$

thus

$$0 \le -\frac{1}{\beta}\ln\mu^\beta(x) \le m\ \mathrm{osc}(V) - \frac{1}{\beta}\ln(\epsilon^2 \min_{x\in E}\nu(x))$$

So at least under (H), we can consider accumulation functions U of $-\ln(\mu^\beta(x))/\beta$ for β large.
In order to derive Bellman's equations verified by this kind of objects, let us introduce for $n \in \mathbb{N}^*$ and $x, y \in E$, the n-communication cost function

$$V^{(n)}(x,y) \quad := \quad \min_{P\in\mathcal{P}^{(n)}_{x,y}} V(P)$$

where $\mathcal{P}^{(n)}_{x,y}$ is the set of paths of length n going from x to y (ie $P = (P_1, \ldots, P_n)$ with $M(x, P_1) > 0$ and $P_n = y$). In particular for any $x, y \in E$, $V^{(1)}(x, y) = V(y)$. As in the proof of proposition 4.4, we show without difficulties that for any $x, y \in E$, $\liminf_{n\to\infty} V^{(n)}(x, y) = V_\mathcal{C}$ (and this is a true limit if M is aperiodic, the difference of the two terms being at most of order $1/n$).
For a subset $A \subset E$ we also define the M-boundary of A as the subset of all possibles sites which are accessible from A, that is

$$\partial_M(A) = \{y \in E - A\ ; \quad \exists x \in A \quad M(x,y) > 0\}$$

Now we can state

Proposition 4.7 *Let $U \in \mathbb{R}^E_+$ be any accumulation point as above, then it satisfies the Bellman's fixed point equations*

$$U(y) = \inf_{y\in E}\left(U(x) + nV^{(n)}(x,y)\right) - nV_I \tag{20}$$

for any $n \in \mathbb{N}^*$ and $nV_I = \inf_{x,y\in E}(U(x) + nV(x,y))$. Furthermore We have the inclusions

$$U^{-1}(0) \subset (V \le V_I) \quad \text{and} \quad \partial_M U^{-1}(0) \subset (V > V_I) \tag{21}$$

Before getting into the proof of this proposition, let us pause for a while and give some comments on the consequence of these results. The inclusions (21) show that a point $x \in \{V \le V_I\}$ with energy $U(x) > 0$ cannot be reached from $U^{-1}(0)$ (the reverse being in general true). This shows that when all pairs of points $x, y \in \{V \le V_I\}$ can be joined by a path in this level set then $U^{-1}(0) = \{V \le V_I\}$.

Proof of proposition 4.7:

The Bellman's equations are immediate consequences of the equalities, seen in the proof of lemma 4.1,

$\forall\, n \in \mathbb{N}^*$, $\forall\, x \in E$, $\forall\, \beta > 0$,

$$\mu^\beta(x) \quad = \quad (\mu^\beta[\exp(\beta V)])^n \sum_{y\in E}\mu^\beta(y)\mathbb{E}_y[\mathbb{1}_{\{x\}}(X_n)\exp(-\beta V(X_1) - \cdots - \beta V(X_n))]$$

by taking the logarithm and dividing by β that we let go to infinity. To prove the inclusions (21), we suppose on the contrary that we can find a pair $(x, y) \in E^2$ such that

$$U(x) = 0\,, \quad M(x,y) > 0 \quad U(y) > 0 \quad \text{and} \quad V(y) \le V_I$$

¿From the Bellman's equation this will give that

$$\begin{aligned} U(y) &= \inf\{U(z)+V(y)-V_I\ ;\ z\in E,\quad M(z,y)>0\} \\ &\le \inf\{U(z)\ ;\quad M(z,y)>0\} \le U(x)=0 \end{aligned}$$

and we obtain a contradiction with the fact that $U(y)>0$.

■

We end this article with a simple three points example in which $V_I > V_\nu$ and $V^{-1}(V_\nu) \not\subset U^{-1}(0)$. So we take for state space $E=\{0,1,2\}$ and we consider the Markov kernel defined by

$$M = \begin{pmatrix} p & 1-p & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \quad \text{, with} \quad p\in(0,1)$$

It is clear that M is irreducible and aperiodic and we check that its unique invariant probability ν is given by

$$\nu(0) = \frac{1}{3-2p} \quad \text{and} \quad \nu(1)=\nu(2)=\frac{1-p}{3-2p}$$

Let $V : E \to \mathbb{R}_+$ be a potential function such that

$$V(0) > \frac{V(0)+V(1)+V(2)}{3} > V(2) > V(1) = 0 \tag{22}$$

So the ν-essential infimum V_ν is given by $V_\nu = 0 = V(1)$ and by proposition 4.5, we have

$$V_I = \frac{V(0)+V(1)+V(2)}{3}$$

This could also be deduced from the fact that here the rate function I satisfies

$$I(\mu)<\infty \iff \exists \in [0,1]\ :\ \mu = r(\delta_0+\delta_1+\delta_2)/3 + (1-r)\delta_0$$

property which reflects that trajectories of X are concatenations of the words [0] and [1,2,0] (except for a possible start with [2]). Our next objective is to solve explicitely the Bellman's fixed point equation (20) for $n=1$:

$$\left\{ \begin{array}{rcl} U(0) &=& \min\{U(0),U(2)\} + V(0) - V_I \\ U(1) &=& U(0) + V(1) - V_I \\ U(2) &=& U(1) + V(2) - V_I \end{array} \right.$$

By (22), we see that in the first equality the min cannot be $U(0)$ (otherwise $V(0) = V_I$), so $U(0) = U(2)+V(0)-V_I$ and this shows that $U(2)<U(0)$. The last equation also implies that $U(2)<U(1)$ and necessarily $U(2)=0$, from which we obtain that U is unique and that it is given by

$$\left\{ \begin{array}{rcl} U(0) &=& V(0) - V_I \\ U(1) &=& V_I - V(2) \\ U(2) &=& 0 \end{array} \right.$$

One concludes that $\lim_{\beta\to\infty} \mu^\beta(2) = 1$ and that this convergence is exponentially fast. In particular μ^β does not concentrate for large β on the unique point 1 where the "essential" infimum is achieved (this latter assertion could also be deduced directly from the observation (21)).

References

[1] C. Bélisle. Convergence theorems for a class of simulated annealing algorithms on $\mathbb{R}^d$. *J. Appl. Probab.*, 29(4):885–895, 1992.

[2] R. Cerf. The dynamics of mutation-selection algorithms with large population sizes. *Ann. Inst. H. Poincaré Probab. Statist.*, 32(4):455–508, 1996.

[3] P. Del Moral and L. Miclo. On the convergence and the applications of the generalized simulated annealing. *SIAM Journal on Control and Optimization*, 37(4):1222–1250, 1999.

[4] P. Del Moral and L. Miclo. Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non-linear filtering. In *Séminaire de Probabilités, XXXIV*, pages 1–145. Springer, Berlin, 2000.

[5] P. Del Moral and L. Miclo. Particle approximations of Lyapunov exponents connected to Schrödinger operators and Feynman-Kac semigroups. Preprint, publications du Laboratoire de Statistique et Probabilités, no 2001-08, 2001.

[6] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications.* Jones and Bartlett, 1993.

[7] M. Duflo. *Random iterative models.* Springer-Verlag, Berlin, 1997. Translated from the 1990 French original by Stephen S. Wilson and revised by the author.

[8] P. Dupuis and R. Ellis. *A weak convergence approach to the theory of large deviations.* John Wiley & Sons Inc., New York, 1997. A Wiley-Interscience Publication.

[9] R. Holley and D. Stroock. Simulated annealing via Sobolev inequalities. *Communications in Mathematical Physics*, 115:553–569, 1988.

[10] L. Miclo. Recuit simulé sans potentiel sur un ensemble fini. In J. Azéma, P.A. Meyer, and M. Yor, editors, *Séminaire de Probabilités XXVI*, Lecture Notes in Mathematics 1526, pages 47–60. Springer-Verlag, 1992.