On a Class of Genealogical and Interacting Metropolis Models

Pierre Del Moral[1] and Arnaud Doucet[2]

[1] CNRS-UMR 5583, Univ. P. Sabatier, 31062 Toulouse, France
e-mail: delmoral@cict.fr
[2] Engineering Department, Cambridge University, CB2 1PZ Cambridge, UK
e-mail: ad2@eng.cam.ac.uk

Summary. A genealogical tree based particle model for drawing approximate samples from the conditional path-distributions of a Markov chain with respect to its terminal values is presented. This path-particle evolution model can be interpreted as the historical process associated to a sequence of interacting Metropolis Markov chains. This novel class of interacting models can also be used to obtain approximate samples from a given target distribution which is only known up to a normalizing constant. We design an original Feynman–Kac modeling technique for studying the asymptotic analysis of these path-particle and Metropolis type simulation models. We provide precise convergence results as the time or the size of the systems tends to infinity. In contrast to the traditional Metropolis model we show that the decays to equilibrium do not depend on the nature of the desired limiting distribution.

Key words: Interacting particle systems, genetic algorithms, genealogical trees, Feynman–Kac formulae, Metropolis algorithm, Markov chain Monte Carlo.

MSC (2000): 65C05, 65C35, 65C40.

1 Introduction

The problem of generating random samples from a given distribution plays an important role in many research areas including physics, biology, statistics and engineering science such as in signal and image processing. In practice the target distribution π is often complex and it is only known up to a normalizing constant. During the last decades several strategies have been proposed in the literature and Markov chain Monte Carlo (MCMC) methods have become the most popular tools. We refer for instance the reader to the book of W. R. Gilks, S. Richardson, D. J. Spielgelhalter [13] and to the pioneering articles of W. K. Hastings [14] and N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller [17].

J. Azéma et al. (Eds.): LNM 1832, pp. 415–446, 2003.

The underlying idea in MCMC methods consists in building a Markov transition K_π admitting π as a stationary measure. After running the chain for a long time its random states are approximately distributed according to the desired measure. These techniques are widely used in practice and the study of their asymptotic behavior has been the subject of many research articles.

Another important and related question is to generate random samples from the path distributions of a Markov chain with initial distribution π and restricted to a fixed terminal value. The interacting and genealogical particle models presented in this article give a novel strategy for drawing recursively in time approximate samples according to these conditional path-distributions. In addition these path-particle evolution models can be interpreted as the historical process associated to a sequence of interacting Metropolis algorithms. The latter also provides a new strategy for drawing samples according to a given target distribution π.

This article also sheds some new lights on the connection between traditional Monte Carlo Markov chain studies and the recently developed branching and interacting particle techniques arising in physics and advanced signal processing ([6], [10]). To motivate this article let us briefly mention how these two approaches are combined:

We first design a strategy to represent the desired conditional and target distributions in terms of a non linear Feynman–Kac distribution flow model.

This original model is a combination of a non linear Feynman–Kac and Metropolis type Markov chain. The non linear nature of the latter comes from the fact that its elementary transitions depend on the distributions of the current random states. In contrast to the classical Metropolis situation the decays to equilibrium of this non homogeneous Feynman–Kac–Metropolis model does not depend on the nature of the limiting measure π. The interacting particle approximating model associated to this flow can be interpreted as a genetic type sequence of interacting Metropolis Markov chains. Another difference with the traditional Metropolis model is that the historical process associated to this scheme can be regarded as a path-particle simulation technique for sampling from the conditional path distribution of a chain restricted to its terminal values. In this article we use this original Feynman–Kac modeling technique to analyze the asymptotic behavior of this class of particle models as the population size or the time parameter tends to infinity.

This opening section is decomposed in two parts. In the first one we describe the interacting Metropolis particle model and its historical process and we compare the former with the traditional Metropolis algorithm. We also give in some details the two main results presented in this article. In the second part we close with some comments on related works on this subject and we propose new lines of research problems. We finally initiate a comparison between the traditional and this novel interacting particle Metropolis model.

Here are some standard notations to be used in all the paper. Let $\mathcal{P}(S)$ and $\mathcal{B}_b(S)$ denote respectively the set of probability measures and bounded measurable functions on a given measurable space $(S, \mathcal{S})$. We denote by δ_x the Dirac measure at a point $x \in S$.

As usual $\mathcal{B}_b(S)$ is regarded as a Banach space with the supremum norm

$$\forall f \in \mathcal{B}_b(S), \qquad \|f\| = \sup_{x \in S} |f(x)|.$$

We also slightly abuse notations and denote by 1 the unit function on S.

The total variation distance $\|\mu_1 - \mu_2\|_{\mathrm{tv}}$ between probability measures μ_1, $\mu_2 \in \mathcal{P}(S)$ is defined by

$$\|\mu_1 - \mu_2\|_{\mathrm{tv}} = \frac{1}{2} \sup\bigl\{|\mu_1(f) - \mu_2(f)| \,:\, f \in \mathcal{B}_b(S),\ \|f\| \leqslant 1\bigr\}.$$

Any Markov transition $K(u, \mathrm{d}v)$ from S into itself generates two operators. One acting on bounded $\mathcal{S}$-measurable functions $f \in \mathcal{B}_b(S)$ and taking values in $\mathcal{B}_b(S)$

$$\forall (u, f) \in \bigl(S \times \mathcal{B}_b(S)\bigr), \qquad (Kf)(u) = \int_S K(u, \mathrm{d}v)\, f(v),$$

and the other one acting on measures $\mu \in \mathcal{P}(S)$ and taking values in $\mathcal{P}(S)$

$$\forall (\mu, A) \in \bigl(\mathcal{P}(S) \times \mathcal{S}\bigr), \qquad (\mu K)(A) = \int_S \mu(\mathrm{d}u)\, K(u, A).$$

We also recall that the contraction coefficient $\beta(K)$ associated to a Markov operator K on $\mathcal{P}(S)$ is defined by

$$\beta(K) = \sup_{\mu_1, \mu_2 \in \mathcal{P}(S)} \frac{\|\mu_1 K - \mu_2 K\|_{\mathrm{tv}}}{\|\mu_1 - \mu_2\|_{\mathrm{tv}}} = \sup_{u, v \in S} \|K(u, .\,) - K(v, .\,)\|.$$

If L is another Markov transition from $(S, \mathcal{S})$ into itself then we denote by (KL) the composite operator

$$(KL)(u, \mathrm{d}w) = \int_S K(u, \mathrm{d}v)\, L(v, \mathrm{d}w).$$

We also write K^n, $n \geqslant 0$, the n-th time iterate of the operator K defined by the inductive formula

$$K^n = K^{n-1} K \quad \text{with} \quad K^0 = \mathrm{Id}.$$

For a distribution μ and a Markov transition K on S we denote by $(\mu \times K)_i$, $i = 1, 2$, the distributions on S^2 defined by

$$(\mu \times K)_1(\mathrm{d}(u, v)) = \mu(\mathrm{d}u)\, K(u, \mathrm{d}v) \quad \text{and} \quad (\mu \times K)_2(\mathrm{d}(u, v)) = \mu(\mathrm{d}v)\, K(v, \mathrm{d}u).$$

Sometimes we simplify notations and write $(\mu \times K)$ instead of $(\mu \times K)_1$.

We usually associate to a given Markov transition K on S a canonical Markov chain

$$\left(\Omega = S^{\mathbb{N}}, F = (F_n)_{n\geqslant 0}, Y = (Y_n)_{n\geqslant 0}, \left(\mathbb{P}^K_y\right)_{y\in S}\right)$$

with elementary transitions K. We recall that $(F_n)_{n\geqslant 0}$ represents the non decreasing family of σ-algebras F_n generated by the canonical variables $X_0, \ldots, X_n$. For a given distribution $\mu \in \mathcal{P}(S)$ we define the probability measures

$$\mathbb{P}^K_\mu = \int_S \mu(\mathrm{d}y)\, \mathbb{P}^K_y.$$

In this notation we have for instance

$$\mathbb{P}^K_\mu\big((Y_0, \ldots, Y_n) \in \mathrm{d}(y_0, \ldots, y_n)\big) = \mu(\mathrm{d}y_0)\, K(y_0, \mathrm{d}y_1) \ldots K(y_{n-1}, \mathrm{d}y_n).$$

We use $\mathbb{E}^K_\mu$ and $\mathbb{E}^K_y$ for the expectation with respect to $\mathbb{P}^K_\mu$ and $\mathbb{P}^K_y$. Finally we will use the traditional conventions

$$(\sup_\varnothing, \inf_\varnothing) = (-\infty, \infty) \qquad \text{and} \qquad \left(\textstyle\sum_\varnothing, \prod_\varnothing\right) = (0, 1).$$

1.1 Description of the models and statement of some results

Let π and L be a probability measure and a Markov kernel on a measurable state space $(S, \mathcal{S})$. We associate to the pair (π, L) the S-valued Markov chain $(\Omega, F, Y, \mathbb{P}^L_\pi)$ with initial distribution π and Markov transitions L.

In this article we design a genealogical tree based model for solving numerically and recursively in time (a version of) the conditional distribution flow

$$n \longrightarrow \mathbb{P}^L_\pi\big((Y_0, \ldots, Y_n) \in \mathrm{d}(y_0, \ldots, y_n) \,\big|\, Y_{n+1} = y\big). \tag{1}$$

When L is sufficiently regular, it is convenient to notice that in a sense to be defined we have that

$$\lim_{n\to\infty} \mathbb{P}^L_\pi(Y_0 \in \mathrm{d}y_0 \,|\, Y_{n+1} = y) = \pi(\mathrm{d}y_0).$$

We will combine this observation with a time inversion formula to prove that the current occupation measures of the genealogical particle model converge to the distribution π. In other words the particle approximating model presented in this work can also be used for drawing approximate samples from π as soon as the latter is known up to a normalizing constant.

To describe briefly this algorithm and its connection with the traditional Metropolis model we suppose we are given an auxiliary Markov kernel K on S such that

- The measures $(\pi \times K)_1$ and $(\pi \times L)_2$ are mutually absolutely continuous.
- The Radon–Nykodim derivative defined by

$$G = \frac{\mathrm{d}(\pi \times L)_2}{\mathrm{d}(\pi \times K)_1} \tag{2}$$

  is a bounded and strictly positive function on $E = (S \times S)$.

The standard Metropolis ratio corresponds to the case where $K = L$.

In practice the Markov kernel L is fixed and the choice of K depends on the problem at hand. Usually the state space S is endowed with a topology and $K(u, .)$ is often defined as the uniform distribution on a suitably chosen neighborhood of $u \in S$. If we are not interested in computing the conditional path-distributions (1) but only want to draw samples according to π then we have the choice of the pair (K, L). To illustrate this observation and motivate this article let us suppose that π is the Boltzmann–Gibbs measure associated to a pair measure/potential function (ν, H) with $H \geqslant 0$, $\nu(\mathrm{e}^{-H}) > 0$ and defined by

$$\pi(\mathrm{d}x) = \frac{1}{\nu(\mathrm{e}^{-H})}\, \mathrm{e}^{-H(x)}\, \nu(\mathrm{d}x). \tag{3}$$

In this situation the generalized Metropolis ratio takes the form

$$G(y, y') = \mathrm{e}^{-(H(y')-H(y))}\, \frac{\mathrm{d}(\nu \times L)_2}{\mathrm{d}(\nu \times K)_1}(y, y'). \tag{4}$$

Whenever K is reversible with respect to ν if we take $L = K$ then we find that

$$G(y, y') = \mathrm{e}^{-(H(y')-H(y))}. \tag{5}$$

This indicates that if we want to approximate simultaneously the conditional path-distributions (1) and the target distribution π using the forthcoming interacting Metropolis model then we need to consider the ratio (4). In the previous reversible case if we only want to draw samples according to π we can alternatively take the ratio (5).

To better connect our strategy with the traditional Metropolis model we have chosen to present the latter in a way which parallels the forthcoming construction.

The traditional Metropolis algorithm associated to the triplet (π, K, L) introduced above can be represented in terms of a Markov chain

$$Z_n = (U_n, V_n)$$

on the product space $E = S^2$ with a two-step selection/mutation elementary transition

$$\mathbf{K}_\pi = \mathbf{S}_\pi M^K \tag{6}$$

and an initial distribution of the form $\mu_0 = \delta_y \times K$ for some arbitrary point $y \in S$. The selection and mutation transitions $\mathbf{S}_\pi$ and M^K are defined by

$$\mathbf{S}_\pi\big((u,v),\,.\,\big) = \big(1 \wedge G(u,v)\big)\,\delta_{(u,v)}(\,.\,) + \big(1 - (1 \wedge G(u,v))\big)\,\delta_{(u,u)}(\,.\,)$$

and

$$M^K\big((u,v),\mathrm{d}(u',v')\big) = \delta_v(\mathrm{d}u')\,K(u',\mathrm{d}v').$$

The probabilistic interpretation of the two-step Markov chain

$$Z_n = (U_n,V_n) \xrightarrow{S_\pi} \widehat{Z}_n = \big(\widehat{U}_n,\widehat{V}_n\big) \xrightarrow{M^K} Z_{n+1} = (U_{n+1},V_{n+1})$$

is simple. At time $n=0$ we start from a random state (U_0,V_0) distributed according to $\delta_y \times K$. That is we set $U_0 = y$ and we sample V_0 according to $K(y,\,.\,)$. Starting from a point $Z_n = (U_n,V_n)$ the selection transition at time $n \geqslant 0$ consists in accepting the second component V_n with a probability $(1 \wedge G(U_n,V_n))$ and setting

$$\widehat{Z}_n = \big(\widehat{U}_n,\widehat{V}_n\big) = (U_n,V_n).$$

Otherwise we reject it and we set $\widehat{Z}_n = (\widehat{U}_n,\widehat{V}_n) = (U_n,U_n)$. The mutation transition consists in evolving randomly according to K the selected component. In other words, given the point $\widehat{Z}_n = (\widehat{U}_n,\widehat{V}_n)$ we set

$$Z_{n+1} = (U_{n+1},V_{n+1}) = \big(\widehat{V}_n,V_{n+1}\big),$$

where V_{n+1} is randomly chosen with distribution $K(\widehat{V}_n,\,.\,)$.

It is easily seen that the first component U_n of Z_n forms a Markov chain with elementary transitions

$$\mathbf{M}_\pi(u,\mathrm{d}u') = \big(1\wedge G(u,u')\big)\,K(u,\mathrm{d}u') + \left(1 - \int_S \big(1\wedge G(u,y)\big)K(u,dy)\right)\delta_u(\mathrm{d}u').$$

It is also well known that if we choose for instance $K = L$ then π is $\mathbf{M}_\pi$-invariant and in the reversible case $\mathbf{M}_\pi$ is reversible with respect to π. Thus, running the chain U_n for a long time we get a random variable approximatively distributed according to π. The convergence to the equilibrium of this chain has been studied by many authors. For instance we refer the reader to the articles of Catoni [1, 2, 3], Gaudron and Trouvé [12], S. F. Jarner and E. Hansen [15], K. L. Mengersen and R. L. Tweedie [16], Miclo [18, 19, 20], Trouvé [21], the review article P. Diaconis and L. Saloff-Coste [11] and references therein.

We would like to mention here that the decays to equilibrium depend on the target distribution π. For instance the rate of convergence to zero of the contraction coefficients $\beta(\mathbf{M}_\pi^n)$ as $n \to \infty$ strongly depends on the nature of the limiting measure π. We will illustrate this observation with a three points example in the end of section 1.2. Another main difference between our interacting particle method and this traditional scheme is that the latter cannot be used to approximate the conditional path-distributions (1).

The particle model presented in this article gives an alternative simulation technique to produce approximate samples from π. The key idea is to consider π as the fixed point of a non linear Feynman–Kac type distribution flow. More precisely instead of studying the linear evolution equation associated to the Metropolis Markov kernel (6) we consider the non linear distribution flow on $\mathcal{P}(E)$ given by

$$\mu_{n+1} = \mu_n K_{\mu_n} \quad \text{with} \quad K_\mu = S_\mu M^K. \tag{7}$$

The collection of selection transition kernels S_μ, $\mu \in \mathcal{P}(E)$, are defined by

$$S_\mu\big((u,v),\,.\,\big) = \varepsilon\, G(u,v)\, \delta_{(u,v)}(\,.\,) + \big(1 - \varepsilon\, G(u,v)\big)\, \psi(\mu)(\,.\,) \tag{8}$$

with a parameter $\varepsilon \geqslant 0$ such that $\varepsilon\, G \leqslant 1$ and the Boltzmann–Gibbs transformation $\psi : \mathcal{P}(E) \to \mathcal{P}(E)$ defined by

$$\psi(\mu)\big(\mathrm{d}(u,v)\big) = \frac{1}{\mu(G)}\, G(u,v)\, \mu\big(\mathrm{d}(u,v)\big).$$

The equation (7) can again be interpreted as the evolution in time of the distributions of a two-step non homogeneous (and non linear) E-valued Markov chain

$$Z_n = (U_n, V_n) \xrightarrow{S_{\mu_n}} \widehat{Z}_n = \big(\widehat{U}_n, \widehat{V}_n\big) \xrightarrow{M^K} Z_{n+1} = (U_{n+1}, V_{n+1})$$

with $\mu_n = \mathrm{Law}(Z_n)$. The mutation stage is the same as above but during the selection a rejected variable is replaced by sampling a new pair according to the Boltzmann–Gibbs distribution $\psi(\mu_n)$. If we denote by ϕ^n, $n \geqslant 0$, the non linear semigroup associated to the flow (7) then our first main result will be basically stated as follows.

Theorem 1. *The distribution $(\pi \times K)$ is a fixed point of ϕ and when L is sufficiently regular then we have for any $\mu_1, \mu_2 \in \mathcal{P}(E)$ and $n \geqslant 0$*

$$\|\phi^{m+n}(\mu_1) - \phi^{m+n}(\mu_2)\|_{\mathrm{tv}} \leqslant b_L\, \beta(L^n) \tag{9}$$

for some fixed $m \geqslant 1$ and finite constant $b_L < \infty$ whose values only depend on L.

The above theorem implies that the law of the first component U_n converge to the desired distribution π with a rate which does not depend on the target distribution. In contrast to the traditional Metropolis model the non homogeneous chain described above cannot be sampled perfectly and another level of approximation is needed.

One way to draw approximate samples according to the distributions μ_n is to associate to the collection of Markov transitions K_μ, $\mu \in \mathcal{P}(E)$, a sequence of interacting particle systems (cf. for instance [5, 6]). These particle approximating models consist in a Markov chain $\xi_n = (\xi_n^i)_{1 \leqslant i \leqslant N}$ on a product space E^N with initial distribution $\mu_0^{\otimes N}$ and elementary transitions

$$\mathbb{P}\big(\xi_{n+1} \in \mathrm{d}(x^1, \ldots, x^N) \mid \xi_n\big) = \prod_{i=1}^{N} K_{\frac{1}{N}\sum_{j=1}^{N} \delta_{\xi_n^j}}(\xi_n^i, \mathrm{d}x^i).$$

The asymptotic behavior as n and/or N tend to infinity depends on the nature of the transitions K_μ but, for fairly general kernels, we have for any fixed time horizon $n \geqslant 0$, as $N \to \infty$ and in a sense to be given

$$\mu_n^N = \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_n^i} \longrightarrow \mu_n.$$

Since $K_\mu = S_\mu M^K$ is the composition of a pair of selection and mutation Markov kernels we observe that the previous particle model is a two-step genetic type Markov chain in the product space E^N

$$\xi_n = (\xi_n^i)_{1 \leqslant i \leqslant N} \xrightarrow{\text{selection}} \widehat{\xi}_n = (\widehat{\xi}_n^i)_{1 \leqslant i \leqslant N} \xrightarrow{\text{mutation}} \xi_{n+1}. \tag{10}$$

During the selection stage each particle $\xi_n^i \longrightarrow \widehat{\xi}_n^i$ evolves according to the Markov transition $S_{\frac{1}{N}\sum_{j=1}^{N} \delta_{\xi_n^j}}(\xi_n^i, \,.\,)$. The mutation stage consists in evolving randomly each selected particle $\widehat{\xi}_n^i \longrightarrow \xi_{n+1}^i$ according to the Markov transition M^K.

One important feature of this genetic type particle model is that the stability properties of the limiting system (7) can be used to derive uniform estimates with respect to the time parameter for the convergence as $N \to \infty$ of the sequence of N-particle approximating measures μ_n^N. We will use this important property to calibrate the convergence of this simulation technique when the pair parameter $(n, N) \to \infty$. As we already mentioned the other impact of this class of interacting Metropolis models is that it gives a natural and simple way for drawing approximate samples from the conditional distributions

$$\mathbb{P}_\pi^L\big((Y_0, \ldots, Y_n) \in \mathrm{d}(y_0, \ldots, y_n) \;\big|\; Y_{n+1} = y\big). \tag{11}$$

To make these observations more precise let us put $\xi_n^i = (U_n^i, V_n^i)$, $1 \leqslant i \leqslant N$. If we take $\mu_0 = \delta_y \times K$ then the initial system is given by

$$\xi_0^i = (U_0^i, V_0^i) = (y, V_0^i)$$

where V_0^i are independent and identically distributed random variables with common law $K(y, \,.\,)$. From the above description we see that the particle model can be alternatively interpreted as a mutation/selection algorithm

$$(U_n^i)_{1\leqslant i\leqslant N} \xrightarrow{\text{mutation}} (V_n^i)_{1\leqslant i\leqslant N} \xrightarrow{\text{selection}} (U_{n+1}^i)_{1\leqslant i\leqslant N}.$$

During the mutation stage each particle U_n^i evolves randomly and independently according to the Markov kernel K to some new locations V_n^i. These new locations V_n^i are accepted or rejected according to a mechanism which depends on the pairs configuration

$$\{(U_n^j, V_n^j)\ ;\ 1\leqslant j\leqslant N\}.$$

With probability $\varepsilon\, G(U_n^i, V_n^i)$ we accept the i-th state V_n^i and we set $U_{n+1}^i = V_n^i$ and, with probability $1-\varepsilon\, G(U_n^i, V_n^i)$, we select randomly a state $\widetilde{V}_n^i$ with distribution

$$\sum_{j=1}^N \frac{G(U_n^j, V_n^j)}{\sum_{k=1}^N G(U_n^k, V_n^k)}\, \delta_{V_n^j}$$

and we set $U_{n+1}^i = \widetilde{V}_n^i$. Loosely speaking the selection transition intends to improve the quality of the configuration by allocating more reproductive opportunities to pair-particles (U_n^j, V_n^j) with higher Metropolis ratio. If we interpret this stage as a birth and death mechanism then we see that particles V_n^j with high ratio $G(U_n^j, V_n^j)$ have more chance to give birth to an offspring than those with poor ratio. When the i-th particle U_{n+1}^i selects a site V_n^j we can also interpret V_n^j as the parent of the individual U_{n+1}^i. Recalling that V_n^j has been sampled according to $K(U_n^j, \,.\,)$ we can interpret U_n^j as the ancestor $U_{n,n+1}^i$ of U_{n+1}^i at level n. Running back in time this construction we can trace back the complete ancestral line

$$\mathcal{U}_n^i = (U_{0,n}^i, U_{1,n}^i, \ldots, U_{n,n}^i)$$

of each current individual $U_{n,n}^i = U_n^i$. The parameter $p = 0, \ldots, n$ represents the level of the ancestors. Up to a time reversal transformation the occupation measure of the corresponding genealogical tree converges to the desired conditional path distributions (11) as N tends to infinity. We summarize and make precise this discussion in terms of the second main result of this article which can be basically stated as follows. We use the notations $\mathbb{E}_\mu^{(N)}(\,.\,)$, $\mu\in\mathcal{P}(S)$, $N\geqslant 1$, for the expectation with respect to the law of the N-particle model $\xi_n^i\in E$ starting with N independent and identically distributed random variables with common law $(\mu\times K)$. When $\mu=\delta_y$ is concentrated at a single point $y\in S$ we simplify notations and we write $\mathbb{E}_y^{(N)}(\,.\,)$ instead of $\mathbb{E}_{\delta_y}^{(N)}(\,.\,)$.

Theorem 2. *For any $n\geqslant 0$, $p\geqslant 1$, $y\in S$ and $f_n\in\mathcal{B}_b(S^{n+1})$ with $\|f_n\|\leqslant 1$ we have the estimate*

$$\mathbb{E}_y^{(N)}\left(\left|\frac{1}{N}\sum_{i=1}^N f_n(\mathcal{U}_n^i) - \mathbb{E}_\pi^L\big(f_n(Y_n, Y_{n-1}, \ldots, Y_0)\mid Y_n = y\big)\right|^p\right)^{1/p} \leqslant \frac{c_p}{\sqrt{N}}\, d(n)$$

for some finite constant $d(n) < \infty$ which only depends on the time parameter $n \geqslant 0$ and an universal constant c_p whose values only depend on the parameter $p \geqslant 1$.

In addition, when the Markov kernel L is sufficiently regular then we have for any $n \geqslant 0$, $\mu \in \mathcal{P}(S)$ and $f \in \mathcal{B}_b(E)$ with $\|f\| \leqslant 1$

$$\mathbb{E}_{\mu}^{(N)}\left(\left|\frac{1}{N}\sum_{i=1}^{N} f(U_{m+n}^{i}) - \pi(f)\right|^{p}\right)^{1/p} \leqslant c_p\,\frac{a_{\pi,L}}{\sqrt{N}} + b_L\,\beta(L^n)$$

for some finite constant $a_{\pi,L} < \infty$ which depends on the pair (π, L) and the same constants (m, b_L) as those arising in (9) of theorem 1. In this situation we also have $d(n) = O(n)$.

1.2 Notes and contents

The idea to introduce interactions for improving convergence of the models is not new. In a short Note D. Chauveau and P. Vandekerkhove [4] have presented a different strategy based on interacting proposal densities. The interaction structure of the latter strongly differs from the genetic type models presented in this article. In contrast to our study their approach is restricted to finite dimensional state spaces and the convergence of the particle approximating models depends on the dimension parameter. Furthermore their model cannot be used to approximate the conditional path distributions of a chain restricted to its terminal values.

The interacting Metropolis method presented here belongs to the class of particle approximating models of Feynman–Kac formulae. The abstract description and the analysis of the latter have been investigated in several research articles. Many asymptotic results are available in this field including empirical process convergence, central limit theorems, large deviation principles as well as increasing propagation of chaos estimates and uniform convergence estimates with respect to the time parameter. The interested reader is referred to the survey article [6] and the more recent studies [5, 7, 9].

The models presented here correspond to the particular situation where the potential functions are given by the generalized Metropolis ratio (2). This particular choice of potential function simplifies the analysis and many known estimates on the asymptotic behavior of these interacting processes can be greatly improved. For instance the asymptotic stability properties stated in theorem 1 and the convergence estimate presented in theorem 2 are here expressed in terms of the contraction coefficient $\beta(L^n)$. Up to our knowledge these results improve the ones obtained in the literature on the subject.

To the best of our knowledge this class of interacting process simulation technique is also new and it has never been covered in the MCMC literature nor on the one on Feynman–Kac particle approximating models. In contrast to the traditional Metropolis model the interaction structure of our model allows to compute recursively in time the conditional path-distribution flow of a given Markov chain restricted to its terminal values. The Feynman–Kac modeling technique and the asymptotic behavior estimates presented in this article also give a novel and solid probabilistic theoretical framework for studying the convergence of a class of interacting Metropolis models and their genealogies.

Our approach also better connects the recently developed branching particle methods for solving non linear filtering problems with the more traditional MCMC literature. In this connection we mention that the choice of the particle approximating model associated to an abstract class of Feynman–Kac flow is not unique. Many branching particle variants have been proposed in the literature including random population size particle models (cf. for instance [6] and references therein). These variants intend to improve the performance of the algorithm for a given computational cost. In this article we have restricted our attention to a generic mean-field type particle strategy. To illustrate this observation and guide the reader to construct interacting models along these lines we give next a brief discussion on one strategy to define branching variants with fixed population size. First we observe that the evolution equation (7) can be rewritten as follows

$$\mu_{n+1} = \phi(\mu_n) \quad \text{with} \quad \phi(\mu) = \psi(\mu) M^K.$$

This already shows that formula (7) holds true for any selection transition S_{μ_n} such that $\mu_n S_{\mu_n} = \psi(\mu_n)$. Instead of (8) and as soon as

$$g = \sup_{x_1, x_2} \big(G(x_1)/G(x_2) \big) < \infty$$

we can alternatively choose

$$S_\mu(x, \,.\,) = \varepsilon \, \frac{G(x)}{\mu(G)} \, \delta_x(\,.\,) + \left(1 - \varepsilon \, \frac{G(x)}{\mu(G)} \right) \psi(\mu)(\,.\,)$$

for any $\varepsilon \leqslant 1/g$. The selection transition (8) associated to the corresponding N-particle approximating model is now given by

$$S_{\frac{1}{N}\sum_{j=1}^N \delta_{\xi_n^j}}(\xi_n^i, \,.\,) = N\varepsilon \, \frac{G(\xi_n^i)}{\sum_{j=1}^N G(\xi_n^j)} \, \delta_{\xi_n^i}(\,.\,) + \left(1 - N\varepsilon \, \frac{G(\xi_n^i)}{\sum_{j=1}^N G(\xi_n^j)}\right) \psi\left(\frac{1}{N}\sum_{j=1}^N \delta_{\xi_n^j}\right)(\,.\,).$$

Notice that for any $N \geqslant g$ we can choose $\varepsilon = 1/N$. The convergence of this particle algorithm can be studied along the same line as the one associated to the choice (8). The optimal choice of the selection transitions is an important open problem. One way to tackle this problem is probably to compare the covariance functions in the central limit theorems for the N-particle approximating measures.

In another context the literature on Markov chains stability and simulated annealing algorithms abounds with precise rates of decays to equilibrium including spectral analysis, large deviation as well as log-Sobolev semigroup contractions. We refer the reader to the list of referenced articles on this subject. One new line of research is to develop the convergence of these interacting and non linear processes along the same lines as for the traditional Metropolis algorithm. In this connection we also notice that the Metropolis model is also used in engineering problems to solve global optimization problems. The idea is to introduce a suitably decreasing cooling schedule so that the corresponding annealing algorithm converge in law to to desired extremal states. The interacting version of this simulated annealing model can also be regarded as a genetic global search algorithm. One open problem is to study the asymptotic behavior of the latter. This analysis is related to the annealed properties of non linear Feynman–Kac flows and it can probably be studied along the same lines as in [8]. Nevertheless the estimates of the convergence of the N-particle approximating models presented in the present article are too crude to solve completely this question. We will hopefully analyze this problem in a forthcoming study.

A precise comparison between these different approaches is beyond the scope of this article. We initiate hereafter this subject in the situation where the Markov kernel $L = K$ is reversible with respect to some distribution ν. We also suppose that the contraction coefficient $\beta(K) \leqslant (1-\varepsilon)$, for some $\varepsilon \in (0,1)$ and the target distribution π is the Boltzmann–Gibbs measure (3) associated to an non negative Hamiltonian function H with bounded oscillations

$$\operatorname{osc}(H) = \sup_{y_1, y_2 \in S} |H(y_1) - H(y_2)| < \infty.$$

In this situation we have the well know estimates

$$\beta(M_\pi^n) \leqslant (1 - \varepsilon\, \mathrm{e}^{-\operatorname{osc}(H)})^n \qquad \text{and} \qquad \beta(K^n) \leqslant (1-\varepsilon)^n.$$

The exponential term arising in the left hand side estimate cannot be removed. For instance suppose $S = \{0, 1, 2\}$ and the Hamiltonian function is given by

$$H = \begin{pmatrix} H(0) \\ H(1) \\ H(2) \end{pmatrix} = \begin{pmatrix} 0 \\ h \\ 0 \end{pmatrix}.$$

Also suppose that (K, ν) are defined in matrix and vector notation as follows

$$K = \begin{pmatrix} 0 & 1 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 1 & 0 \end{pmatrix} \qquad \text{and} \qquad \nu = \big(\nu(0), \nu(1), \nu(2)\big) = (1/5, 3/5, 1/5).$$

Since ν is reversible with respect to K if we take $L = K$ then we have

$$G(y_1, y_2) = \mathrm{e}^{-(H(y_2) - H(y_1))} \qquad \text{and} \qquad M_\pi = \begin{pmatrix} 1 - \mathrm{e}^{-h} & \mathrm{e}^{-h} & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & \mathrm{e}^{-h} & 1 - \mathrm{e}^{-h} \end{pmatrix}.$$

Using some elementary algebraic calculations we find that M_π has three eigenvalues

$$\lambda_2 = \frac{1}{3} - \mathrm{e}^{-h} < \lambda_1 = 1 - \mathrm{e}^{-h} < \lambda_0 = 1.$$

The corresponding orthogonal and normalized eigenvector basis

$$\mathbb{L}_2(\pi) = \mathrm{Span}(\varphi_0, \varphi_1, \varphi_2)$$

is given by

$$\varphi_2 = \sqrt{\frac{3\,\mathrm{e}^{-h}}{2}} \begin{pmatrix} -1 \\ \frac{2\mathrm{e}^h}{3} \\ -1 \end{pmatrix}, \qquad \varphi_1 = \sqrt{1 + \frac{3\,\mathrm{e}^{-h}}{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \qquad \varphi_0 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Using the decomposition

$$M_\pi^n(y_1, y_2) = \pi(y_2) + \lambda_1^n\, \varphi_1(y_1)\varphi_1(y_2)\,\pi(y_2) + \lambda_2^n\, \varphi_2(y_1)\varphi_2(y_2)\pi(y_2),$$

we find after some elementary computations that, as soon as $h \geqslant 2$, one has

$$\beta(M_\pi^n) = \|M_\pi^n(0, .) - M_\pi^n(2, .)\|_{\mathrm{tv}} = (1 - \mathrm{e}^{-h})^n.$$

Similar arguments yield that $\beta(K^{n+1}) = (2/3)^n$.

By the weak law of large numbers the analog of the L_p bound presented in theorem 2 for a sequence of N independent and identical distributed Metropolis Markov chains is proportional to a term of the form

$$e_{iid}^N(n) = \frac{1}{\sqrt{N}} + \big(1 - \varepsilon\,\mathrm{e}^{-\operatorname{osc}(H)}\big)^n$$

In the case of N-interacting Metropolis models we will see that the constant $a_{\pi,K}$ arising in theorem 2 is proportional to e^{4h}. Thus, the resulting $\mathbb{L}_p$ bounds are now proportional to a term of the form

$$e^N_{ips}(n) = \frac{\mathrm{e}^{4\,\mathrm{osc}(H)}}{\sqrt{N}} + (1-\varepsilon)^n.$$

These two estimates seem to lead to the following conclusion: To get the same accuracy the interacting Metropolis algorithm need more particles but less time runs. Nevertheless the exponential term arising in $e^N_{iid}(n)$ may really slow down the decays to equilibrium. When the oscillation of Hamiltonian function are not too large, it seems more advantageous to use the interacting Metropolis model. Indeed, we observe that

$$\left(1-\varepsilon\,\mathrm{e}^{-\,\mathrm{osc}(H)}\right) = (1-\varepsilon)\left(1+\frac{\varepsilon}{1-\varepsilon}\left(1-\mathrm{e}^{-\,\mathrm{osc}(H)}\right)\right).$$

Thus, if we have $\varepsilon \geqslant \mathrm{e}^{-\,\mathrm{osc}(H)}$ and if we take $\sqrt{N} = (1-\varepsilon)^{-n}$ then we find that

$$\sqrt{N}\,e^N_{iid}(n) \geqslant \left(1+(1+\varepsilon)^n\right) \geqslant \left(1+\mathrm{e}^{4\,\mathrm{osc}(H)}\right) = \sqrt{N}\,e^N_{ips}(n)$$

as soon as $n\log(1+\varepsilon) \geqslant 4\,\mathrm{osc}(H)$.

This article has the following structure: Section 2 is concerned with modeling the conditional path-distributions of a given Markov chain restricted to its terminal values in terms of Feynman–Kac path-distributions. In section 3 we present the non linear evolution equations associated to these Feynman–Kac formulae. We also recall the construction of the mean field interacting particle approximating model associated to these flows. We connect these modeling techniques with the genealogical Metropolis particle algorithm presented earlier. Section 4 focuses on the asymptotic behavior of these measure valued models. In a first subsection 4.1 we analyze the stability properties of the non linear Feynman–Kac flow as the time parameter tends to infinity. In the final subsection 4.2, we presents essentially the proof of theorem 2.

2 Feynman–Kac models

In this section we give a Feynman–Kac functional representation of the conditional distributions of a given Markov chain restricted to its terminal values. This preliminary modeling strategy is the essential step to define and study the asymptotic behavior of the interacting Metropolis model presented in section 1.1. We begin with presenting some of the main technical assumptions on the triplet (π, K, L) used in this article. We also present a time inversion formula and a state space enlargement technique which allow to describe the conditional path-distributions (1) in terms of class Feynman–Kac path measures associated to the Metropolis potential ratio (2).

Let π, L and K be a probability measure and a pair of Markov kernels on a measurable state space $(S, \mathcal{S})$. We associate to the pairs (π, L) and (π, K) the S-valued canonical Markov chains

$$\left(\Omega, F, Y, \mathbb{P}^L_\pi\right) \qquad \text{and} \qquad \left(\Omega, F, Y, \mathbb{P}^K_\pi\right)$$

with initial distribution π and respective Markov transitions L and K. Without further mention we suppose the triplet (π, K, L) is chosen such that for some finite $g < \infty$ and for any $(x, x') \in E^2$ we have that

$$G(x) \leqslant g\, G(x').$$

This condition can alternatively be written in terms of the pairs (π, K) and (π, L) as follows: for any $x,\ x' \in E = S^2$

$$\frac{1}{g} \leqslant G(x)/G(x') = \frac{\mathrm{d}(\pi \times L)_2}{\mathrm{d}(\pi \times K)_1}(x)\, \frac{\mathrm{d}(\pi \times K)_1}{\mathrm{d}(\pi \times L)_2}(x') \leqslant g.$$

It is also instructive to note that the above condition is equivalent to the following inequalities

$$\frac{1}{\sqrt{g}}\, (\pi \times K)_1 \leqslant (\pi \times L)_2 \leqslant \sqrt{g}\, (\pi \times K)_1.$$

Using a clear induction on the time parameter $n \in \mathbb{N}$ we find that for any $n \in \mathbb{N}$ we have $\pi L^n \ll \pi$ and for any $y \in S$

$$\frac{\mathrm{d}\pi L^n}{\mathrm{d}\pi}(y) \in [g^{-n/2}, g^{n/2}].$$

Finally observe that for the Boltzmann–Gibbs measure (3) and when $K = L$ is ν-reversible the above condition is met with $\log(g) = 2\,\mathrm{osc}(H)$.

By definition of G we readily obtain the inversion formula

$$\begin{aligned} &\pi(\mathrm{d}y_0)K(y_0, \mathrm{d}y_1) \ldots K(y_n, \mathrm{d}y_{n+1}) \textstyle\prod_{p=0}^{n} G(y_p, y_{p+1}) \\ &= \pi(\mathrm{d}y_{n+1})L(y_{n+1}, \mathrm{d}y_1) \ldots L(y_1, \mathrm{d}y_0). \end{aligned} \tag{12}$$

In other words we have

$$\frac{\mathrm{d}\mathbb{P}^L_\pi((Y_{n+1}, Y_n, \ldots, Y_0) \in\ .)}{\mathrm{d}\mathbb{P}^K_\pi((Y_0, Y_1, \ldots, Y_{n+1}) \in\ .)}(y_0, \ldots, y_{n+1}) = \prod_{p=0}^{n} G(y_p, y_{p+1}),$$

and we obtain the following pivotal lemma.

Lemma 1 (inversion formula). *For any $n \geqslant 0$ and $\varphi_n \in \mathcal{B}_b(S^{n+1})$ we have*

$$\mathbb{E}^L_\pi\big(\varphi_n(Y_{n+1}, Y_n, \ldots, Y_0)\big) = \mathbb{E}^K_\pi\left(\varphi_n(Y_0, Y_1, \ldots, Y_{n+1}) \prod_{p=0}^{n} G(Y_p, Y_{p+1})\right).$$

If we put

$$X_n = (Y_n, Y_{n+1}) \qquad \text{and} \qquad \overleftarrow{X}_n = (Y_{n+1}, Y_n) \in E = S \times S$$

then under $\mathbb{P}_\pi^K$ the sequence $X = (X_n)_{n \geqslant 0}$ constitutes an E-valued Markov chain with elementary transitions

$$M^K\big((y_0, y_1), \mathrm{d}(y'_0, y'_1)\big) = \delta_{y_1}(\mathrm{d}y'_0)\, K(y'_0, \mathrm{d}y'_1).$$

Given a probability measure $\eta \in \mathcal{P}(E)$ we slightly abuse notations and denote by $\mathbb{P}_\eta^K$ instead of $\mathbb{P}_\eta^{M^K}$ the probability measure on $E^{\mathbb{N}}$ defined by

$$\mathbb{P}_\eta^K\big((X_0, \ldots, X_n) \in \mathrm{d}(x_0, \ldots, x_n)\big) = \eta(\mathrm{d}x_0) M^K(x_0, \mathrm{d}x_1) \ldots M^K(x_{n-1}, \mathrm{d}x_n).$$

In these simplified notations and if $\mathbb{E}_\eta^K$ stands for the expectation with respect to $\mathbb{P}_\eta^K$ then we have for any bounded measurable function f on E^{n+1}

$$\begin{aligned} &\mathbb{E}_\eta^K\big(f(X_0, \ldots, X_n)\big) = \mathbb{E}_\eta^K\big(f\big((Y_0, Y_1), \ldots, (Y_n, Y_{n+1})\big)\big) \\ &\quad = \int_{S^{n+1}} \eta\big(\mathrm{d}(y_0, y_1)\big)\, K(y_0, \mathrm{d}y_1) \ldots K(y_{n-1}, \mathrm{d}y_n)\, f\big((y_0, y_1), \ldots, (y_n, y_{n+1})\big). \end{aligned}$$

Using this state space enlargement technique the above formulae lead to the following Feynman–Kac representations.

Proposition 1. *For any $f_n \in \mathcal{B}_b(E^{n+1})$, $n \geqslant 0$ and $(y_0, y_1) \in E$ we have*

$$\begin{aligned} &\mathbb{E}_{(y_0,y_1)}^K\Big(f_n(X_0, X_1, \ldots, X_n) \prod_{p=1}^n G(X_p)\Big) \\ &\qquad = \frac{\mathrm{d}\pi L^n}{\mathrm{d}\pi}(y_1)\, \mathbb{E}_{(\pi \times L)_1}^L\Big(f_n\big(\overleftarrow{X}_n, \ldots, \overleftarrow{X}_0\big) \,\Big|\, \overleftarrow{X}_n = (y_0, y_1)\Big) \end{aligned}$$

and

$$\begin{aligned} &\mathbb{E}_{(y_0,y_1)}^K\Big(f_n(X_0, X_1, \ldots, X_n) \prod_{p=1}^{n-1} G(X_p)\Big) \\ &\qquad = \frac{\mathrm{d}\pi L^{n-1}}{\mathrm{d}\pi}(y_1)\, \mathbb{E}_{(\pi \times K)_2}^L\Big(f_n\big(\overleftarrow{X}_n, \ldots, \overleftarrow{X}_0\big) \;\Big|\; \overleftarrow{X}_n = (y_0, y_1)\Big). \end{aligned}$$

Proof. First we use the inversion formula presented in lemma 1 to check that for any $\varphi \in \mathcal{B}_b(E)$

$$\begin{aligned} \mathbb{E}_\pi^K\Big(\varphi(Y_0, Y_1) \prod_{p=1}^n G(Y_p, Y_{p+1})\Big) &= \mathbb{E}_\pi^K\Big(\varphi(Y_0, Y_1)\, G(Y_0, Y_1)^{-1} \prod_{p=0}^n G(Y_p, Y_{p+1})\Big) \\ &= \mathbb{E}_\pi^L\big(\varphi(Y_{n+1}, Y_n)\, G(Y_{n+1}, Y_n)^{-1}\big) \\ &= \int \frac{\mathrm{d}\pi L^n}{\mathrm{d}\pi}(y_n)\, \pi(\mathrm{d}y_{n+1})\, K(y_{n+1}, \mathrm{d}y_n)\, \varphi(y_{n+1}, y_n). \end{aligned}$$

This yields that

$$\mathbb{E}_{\pi}^{K}\Big(\varphi(Y_0,Y_1)\prod_{p=1}^{n}G(Y_p,Y_{p+1})\Big)=\mathbb{E}_{\pi}^{K}\Big(\varphi(Y_0,Y_1)\,\frac{\mathrm{d}\pi L^n}{\mathrm{d}\pi}(Y_1)\Big).$$

Since this formula holds true for any $\varphi\in\mathcal{B}_b(E)$ we conclude that for any $(y_0,y_1)\in E$

$$\mathbb{E}_{(y_0,y_1)}^{K}\Big(\prod_{p=1}^{n}G(Y_p,Y_{p+1})\Big)=\frac{\mathrm{d}\pi L^n}{\mathrm{d}\pi}(y_1). \tag{13}$$

From lemma 1 we find that for any $\varphi'\in\mathcal{B}_b(E)$

$$\begin{aligned}
&\mathbb{E}_{\pi}^{L}\big(\varphi'(Y_{n+1},Y_n)\,\varphi_n(Y_{n+1},\ldots,Y_0)\big)\\
&\quad=\mathbb{E}_{\pi}^{K}\Big(\varphi'(Y_0,Y_1)\,\varphi_n(Y_0,\ldots,Y_{n+1})\prod_{p=0}^{n}G(Y_p,Y_{p+1})\Big)\\
&\quad=\mathbb{E}_{\pi}^{K}\Big(\varphi'(Y_0,Y_1)\,\mathbb{E}_{(Y_0,Y_1)}^{K}\Big(\varphi_n(Y_0,\ldots,Y_{n+1})\prod_{p=0}^{n}G(Y_p,Y_{p+1})\Big)\Big).
\end{aligned}$$

Using (13) we also prove that

$$\begin{aligned}
&\mathbb{E}_{\pi}^{L}\big(\varphi'(Y_{n+1},Y_n)\,\varphi_n(Y_{n+1},\ldots,Y_0)\big)\\
&\quad=\mathbb{E}_{\pi}^{K}\Big(\varphi'(Y_0,Y_1)\,\mathbb{E}_{(Y_0,Y_1)}^{K}\Big(\prod_{p=0}^{n}G(Y_p,Y_{p+1})\Big)\\
&\qquad\qquad\times\frac{\mathrm{d}\pi}{\mathrm{d}\pi L^n}(Y_1)\,\mathbb{E}_{(Y_0,Y_1)}^{K}\Big(\varphi_n(Y_0,\ldots,Y_{n+1})\prod_{p=1}^{n}G(Y_p,Y_{p+1})\Big)\Big)\\
&\quad=\mathbb{E}_{\pi}^{K}\Big(\varphi'(Y_0,Y_1)\prod_{p=0}^{n}G(Y_p,Y_{p+1})\\
&\qquad\qquad\times\frac{\mathrm{d}\pi}{\mathrm{d}\pi L^n}(Y_1)\,\mathbb{E}_{(Y_0,Y_1)}^{K}\Big(\varphi_n(Y_0,\ldots,Y_{n+1})\prod_{p=1}^{n}G(Y_p,Y_{p+1})\Big)\Big).
\end{aligned}$$

Then by lemma 1 we get

$$\begin{aligned}
&\mathbb{E}_{\pi}^{L}\big(\varphi'(Y_{n+1},Y_n)\,\varphi_n(Y_{n+1},\ldots,Y_0)\big)\\
&\quad=\mathbb{E}_{\pi}^{L}\Big(\varphi'(Y_{n+1},Y_n)\,\frac{\mathrm{d}\pi}{\mathrm{d}\pi L^n}(Y_n)\\
&\qquad\qquad\times\mathbb{E}_{(Y_{n+1},Y_n)}^{K}\Big(\varphi_n(Y_0,\ldots,Y_{n+1})\prod_{p=1}^{n}G(Y_p,Y_{p+1})\Big)\Big).
\end{aligned}$$

Since this formula is valid for any φ' we conclude that for any $(y_0,y_1)\in E$

$$\begin{aligned}
&\mathbb{E}_{(y_0,y_1)}^{K}\Big(\varphi_n(Y_0,\ldots,Y_{n+1})\prod_{p=1}^{n}G(Y_p,Y_{p+1})\Big)\\
&\qquad=\frac{\mathrm{d}\pi L^n}{\mathrm{d}\pi}(y_1)\,\mathbb{E}_{\pi}^{L}\big(\varphi_n(Y_{n+1},\ldots,Y_0)\;\big|\;(Y_{n+1},Y_n)=(y_0,y_1)\big).
\end{aligned}$$

This ends the proof of the first assertion of the proposition. To prove the second formula we first observe that

$$\mathbb{E}^L_{(\pi\times K)_2}\big(\varphi_n(Y_{n+1},\ldots,Y_0)\big) = \mathbb{E}^L_{(\pi\times L)_1}\bigg(\varphi_n(Y_{n+1},\ldots,Y_0)\,\frac{\mathrm{d}(\pi\times K)_2}{\mathrm{d}(\pi\times L)_1}(Y_0,Y_1)\bigg).$$

By lemma 1 we find that

$$\begin{aligned}
&\mathbb{E}^L_{(\pi\times K)_2}\big(\varphi_n(Y_{n+1},\ldots,Y_0)\big)\\
&\qquad= \mathbb{E}^K_{\pi}\left(\varphi_n(Y_0,\ldots,Y_{n+1})\,\frac{\mathrm{d}(\pi\times K)_2}{\mathrm{d}(\pi\times L)_1}(Y_{n+1},Y_n)\prod_{p=0}^{n} G(Y_p,Y_{p+1})\right)\\
&\qquad= \mathbb{E}^K_{\pi}\left(\varphi_n(Y_0,\ldots,Y_{n+1})\,G(Y_n,Y_{n+1})^{-1}\prod_{p=0}^{n} G(Y_p,Y_{p+1})\right)
\end{aligned}$$

and therefore

$$\mathbb{E}^L_{(\pi\times K)_2}\big(\varphi_n(Y_{n+1},\ldots,Y_0)\big) = \mathbb{E}^K_{\pi}\bigg(\varphi_n(Y_0,\ldots,Y_{n+1})\prod_{p=0}^{n-1} G(Y_p,Y_{p+1})\bigg). \tag{14}$$

From this formula we prove that for any $\varphi'\in\mathcal{B}_b(E)$

$$\begin{aligned}
&\mathbb{E}^L_{(\pi\times K)_2}\big(\varphi'(Y_{n+1},Y_n)\,\varphi_n(Y_{n+1},\ldots,Y_0)\big)\\
&\qquad= \mathbb{E}^K_{\pi}\left(\varphi'(Y_0,Y_1)\,\varphi_n(Y_0,\ldots,Y_{n+1})\prod_{p=0}^{n-1} G(Y_p,Y_{p+1})\right)\\
&\qquad= \mathbb{E}^K_{\pi}\left(\varphi'(Y_0,Y_1)\,\mathbb{E}^K_{(Y_0,Y_1)}\left(\varphi_n(Y_0,\ldots,Y_{n+1})\prod_{p=0}^{n-1} G(Y_p,Y_{p+1})\right)\right).
\end{aligned}$$

On the other hand by (13) we have for any $(y_0,y_1)\in E$ and $n\geqslant 1$

$$\mathbb{E}^K_{(y_0,y_1)}\bigg(\prod_{p=1}^{n-1} G(Y_p,Y_{p+1})\bigg) = \frac{\mathrm{d}\pi L^{n-1}}{\mathrm{d}\pi}(y_1).$$

This yields that

$$\begin{aligned}
&\mathbb{E}^L_{(\pi\times K)_2}\big(\varphi'(Y_{n+1},Y_n)\,\varphi_n(Y_{n+1},\ldots,Y_0)\big)\\
&\quad= \mathbb{E}^K_{\pi}\left(\varphi'(Y_0,Y_1)\,\mathbb{E}^K_{(Y_0,Y_1)}\bigg(\prod_{p=0}^{n-1} G(Y_p,Y_{p+1})\bigg)\right.\\
&\qquad\qquad\left.\times\frac{\mathrm{d}\pi}{\mathrm{d}\pi L^{n-1}}(Y_1)\,\mathbb{E}^K_{(Y_0,Y_1)}\left(\varphi_n(Y_0,\ldots,Y_{n+1})\prod_{p=1}^{n-1} G(Y_p,Y_{p+1})\right)\right)\\
&\quad= \mathbb{E}^K_{\pi}\left(\varphi'(Y_0,Y_1)\prod_{p=0}^{n-1} G(Y_p,Y_{p+1})\right.
\end{aligned}$$

$$\times \frac{\mathrm{d}\pi}{\mathrm{d}\pi L^{n-1}}(Y_1)\, \mathbb{E}^K_{(Y_0,Y_1)}\left(\varphi_n(Y_0,\dots,Y_{n+1}) \prod_{p=1}^{n-1} G(Y_p,Y_{p+1})\right)\Bigg).$$

Finally using (14) we arrive at

$$\begin{aligned}
&\mathbb{E}^L_{(\pi\times K)_2}\big(\varphi'(Y_{n+1},Y_n)\,\varphi_n(Y_{n+1},\dots,Y_0)\big)\\
&\quad= \mathbb{E}^L_{(\pi\times K)_2}\Bigg(\varphi'(Y_{n+1},Y_n)\,\frac{\mathrm{d}\pi}{\mathrm{d}\pi L^{n-1}}(Y_n)\\
&\qquad\qquad \times \mathbb{E}^K_{(Y_{n+1},Y_n)}\left(\varphi_n(Y_0,\dots,Y_{n+1}) \prod_{p=1}^{n-1} G(Y_p,Y_{p+1})\right)\Bigg).
\end{aligned}$$

Since this formula holds true for any φ' we conclude that for any $(y_0,y_1) \in E$

$$\begin{aligned}
&\mathbb{E}^K_{(y_0,y_1)}\left(\varphi_n(Y_0,\dots,Y_{n+1}) \prod_{p=1}^{n-1} G(Y_p,Y_{p+1})\right)\\
&\quad= \frac{\mathrm{d}\pi L^{n-1}}{\mathrm{d}\pi}(y_1)\, \mathbb{E}^L_{(\pi\times K)_2}\big(\varphi_n(Y_{n+1},\dots,Y_0) \mid (Y_{n+1},Y_n) = (y_0,y_1)\big).
\end{aligned}$$

This ends the proof of the proposition. □

For a given distribution $\eta_0 \in \mathcal{P}(E)$ we extend the above construction and define the Feynman–Kac path measures η_n on $E_n = E^{n+1}$ given for any $f_n \in \mathcal{B}_b(E_n)$ by the formulae

$$\eta_n(f_n) = \gamma_n(f_n)/\gamma_n(1) \quad \text{with} \quad \gamma_n(f_n) = \mathbb{E}^K_{\eta_0}\left(f_n(X_0,\dots,X_n) \prod_{p=0}^{n-1} G(X_p)\right).$$

It is also convenient to introduce their updated version

$$\widehat{\eta}_n(f_n) = \widehat{\gamma}_n(f_n)/\widehat{\gamma}_n(1) \quad \text{with} \quad \widehat{\gamma}_n(f_n) = \mathbb{E}^K_{\eta_0}\left(f_n(X_0,\dots,X_n) \prod_{p=0}^{n} G(X_p)\right).$$

We close this section by presenting a couple of simple formulae which show that these Feynman–Kac path-measures contain all the desired information on the conditional distributions (1).

When $\eta_0 = \delta_x$ is concentrated at the single point $x \in E$ we write sometimes $\eta_n^{(x)}$ and $\widehat{\eta}_n^{(x)}$. In these notations proposition 1 implies that for any $x \in E$

$$\begin{aligned}
\eta_n^{(x)}(f_n) &= \mathbb{E}^L_{(\pi\times K)_2}\left(f_n(\overleftarrow{X}_n,\dots,\overleftarrow{X}_0) \;\middle|\; \overleftarrow{X}_n = x\right),\\
\widehat{\eta}_n^{(x)}(f_n) &= \mathbb{E}^L_{(\pi\times L)_1}\left(f_n(\overleftarrow{X}_n,\dots,\overleftarrow{X}_0) \;\middle|\; \overleftarrow{X}_n = x\right).
\end{aligned}$$

It is not difficult to prove that if $\eta_0 = \delta_y \times K$ for some $y \in S$ then we have that

$$\widehat{\eta}_n(f_n) = \mathbb{E}^L_\pi\Big(f_n\big((Y_{n+1}, Y_n), (Y_n, Y_{n-1}), \ldots, (Y_1, Y_0)\big) \Bigm| Y_{n+1} = y\Big).$$

In this situation we can also check that

$$\gamma_n(f_n) = \frac{\mathrm{d}\pi L^n}{\mathrm{d}\pi}(y)\, \mathbb{E}^L_{(\pi\times K)_2}\Big(f_n\big((Y_{n+1}, Y_n), (Y_n, Y_{n-1}), \ldots \\ \ldots, (Y_1, Y_0)\big) \Bigm| Y_{n+1} = y\Big).$$

from which we conclude that

$$\eta_n(f_n) = \mathbb{E}^L_{(\pi\times K)_2}\Big(f_n\big((Y_{n+1}, Y_n), (Y_n, Y_{n-1}), \ldots, (Y_1, Y_0)\big) \Bigm| Y_{n+1} = y\Big). \quad (15)$$

When choosing a test function f_n of the following form

$$f_n\big((u_0, v_1), (u_1, v_2), \ldots, (u_n, v_n)\big) = h_n(u_0, \ldots, u_n)$$

for some $h_n \in \mathcal{B}_b(S^{n+1})$ the above formula readily yields that

$$\eta_n(f_n) = \mathbb{E}^L_\pi\big(h_n(Y_n, Y_{n-1} \ldots, Y_0) \bigm| Y_n = y\big). \quad (16)$$

3 Interacting processes

In this section we present the non linear measure valued equations associated to the Feynman–Kac path distributions introduced in section 2 as well as their genealogical path particle approximations. The material here is not new. These abstract path distribution models have been developed in earlier work of one of the authors with L. Miclo [6, 7] and M. Kouritzin and L. Miclo in [5]. For the convenience of the reader we briefly recall in the next two sections some essentials and we complement these studies with presenting a novel McKean interpretation which better connects these two studies to the traditional Metropolis model. We also provide a precise recursive description of the genealogical path particle model which can be readily implemented on a computer.

3.1 Measure valued models

This subsection deals with the non linear evolution equations associated to the Feynman–Kac path-measures η_n and their n-th time marginals. First we recall the path-enlargement technique presented in [7]. This strategy will simplify the description and the asymptotic analysis of the genealogical tree based algorithm associated to the interacting Metropolis model.

As in the previously referenced article we consider the sequence of path-valued random variables

$$\mathcal{X}_n = (X_0, \ldots, X_n) \in E_n = E^{n+1}$$

and the potential functions

$$\mathcal{G}_n : (x_0, \ldots, x_n) \in E_n \longrightarrow \mathcal{G}_n(x_0, \ldots, x_n) = G(x_n) \in (0, \infty).$$

It is easily verified that under the probability measure $\mathbb{P}^K_{\eta_0}$ the sequence $\mathcal{X} = (\mathcal{X}_n)_{n \geqslant 0}$ constitutes a Markov chain taking values at each time n in the product state space $E_n = (S \times S)^{n+1}$. At time $n = 0$ its initial distribution coincides with η_0 and the elementary transitions from E_{n-1} into E_n are given by

$$\begin{aligned}\mathcal{M}^K_n\big((x_0, \ldots, x_{n-1}), \mathrm{d}(x'_0, \ldots, x'_n)\big) \\ = \delta_{(x_0,\ldots,x_{n-1})}\big(\mathrm{d}(x'_0, \ldots, x'_{n-1})\big)\, M^K(x'_{n-1}, \mathrm{d}x'_n).\end{aligned}$$

Furthermore in these notations the Feynman–Kac distributions can be rewritten for any $f_n \in \mathcal{B}_b(E_n)$ as follows

$$\gamma_n(f_n) = \mathbb{E}^K_{\eta_0}\left(f_n(\mathcal{X}_n) \prod_{p=0}^{n-1} \mathcal{G}_p(\mathcal{X}_p)\right)$$

and

$$\widehat{\gamma}_n(f_n) = \mathbb{E}^K_{\eta_0}\left(f_n(\mathcal{X}_n) \prod_{p=0}^{n} \mathcal{G}_p(\mathcal{X}_p)\right) = \gamma_n(f_n \mathcal{G}_n).$$

By the multiplicative nature of the Feynman–Kac distributions and the Markov property it is not difficult to check that the bounded non negative measures γ_n satisfy the linear measure valued equation $\gamma_n = \gamma_{n-1}\mathcal{Q}_n$ with the bounded operators $\mathcal{Q}_n$ from $\mathcal{B}_b(E_n)$ into $\mathcal{B}_b(E_{n-1})$ are defined by $\mathcal{Q}_n(f_n) = \mathcal{G}_{n-1}\ \mathcal{M}^K_n(f_n)$. It follows that the sequence of probability measures η_n satisfies the non linear and measure valued equation

$$\eta_{n+1} = \Phi_{n+1}(\eta_n). \tag{17}$$

The non linear mappings $\Phi_{n+1} : \mathcal{P}(E_n) \to \mathcal{P}(E_{n+1})$ are defined for any $\eta \in \mathcal{P}(E_n)$ by $\Phi_{n+1}(\eta) = \Psi_n(\eta)\mathcal{M}^K_{n+1}$ with the Boltzmann–Gibbs transformations Ψ_n from $\mathcal{P}(E_n)$ into itself given for any $f_n \in \mathcal{B}_b(E_n)$ by $\Psi_n(\eta)(f_n) = \eta(f_n\mathcal{G}_n)/\eta(\mathcal{G}_n)$. In what follows we denote by $\mu_n \in \mathcal{P}(E)$ the time marginals of the Feynman–Kac path-measures η_n with respect to the terminal values. Thus, for any $f \in \mathcal{B}_b(E)$ and $n \geqslant 0$ we have

$$\mu_n(f) = \nu_n(f)/\nu_n(1) \quad \text{with} \quad \nu_n(f) = \mathbb{E}^K_{\eta_0}\left(f(X_n) \prod_{p=0}^{n-1} G(X_p)\right).$$

In these notations we notice that $\eta_0 = \mu_0$ and $\nu_n(1) = \gamma_n(1)$. Arguing as before we also find that the sequence of bounded measures ν_n satisfy the linear homogeneous equations $\nu_n = \nu_{n-1}Q$ with the bounded operator Q

from $\mathcal{B}_b(E)$ into itself defined by $Q(f) = G\,M^K(f)$. In addition the sequence or probability measures μ_n satisfies the non linear homogeneous equation

$$\mu_{n+1} = \phi(\mu_n) = \psi(\mu_n)M^K \tag{18}$$

with the Boltzmann–Gibbs transformation ψ on $\mathcal{P}(E)$ defined for any $\mu \in \mathcal{P}(E)$ and $f \in \mathcal{B}_b(E)$ by $\psi(\mu)(f) = \mu(fG)/\mu(G)$. Let $\varepsilon \geqslant 0$ be a non negative number such that $\varepsilon G \leqslant 1$. We notice that the Boltzmann–Gibbs transformation ψ can be written as $\psi(\mu) = \mu S_\mu$. with the collection of Markov kernels S_μ, $\mu \in \mathcal{P}(E)$, defined by

$$S_\mu(x,\,.\,) = \varepsilon\,G(x)\,\delta_x(\,.\,) + \big(1 - \varepsilon\,G(x)\big)\,\psi(\mu). \tag{19}$$

As already mentioned in the introduction the corresponding non linear equation

$$\mu_{n+1} = \mu_n S_{\mu_n} M^K$$

can be interpreted as the evolution in time of the distributions of a two-step non homogeneous Markov chain

$$Z_n \xrightarrow{S_{\mu_n}} \widehat{Z}_n \xrightarrow{M^K} Z_{n+1} \quad \text{with} \quad \mu_n = \mathrm{Law}(Z_n).$$

This chain is realized in a canonical space

$$\big(\Omega = E^{\mathbb{N}}, F = (F_n)_{n\geqslant 0}, (Z_n, \widehat{Z}_n)_{n\geqslant 0}, \overline{\mathbb{P}}_{\mu_0}\big)$$

with the McKean measure $\overline{\mathbb{P}}_{\mu_0}$ defined in an symbolic form by

$$\begin{aligned}
&\overline{\mathbb{P}}_{\mu_0}\big((Z_0, \widehat{Z}_0, \ldots, Z_n, \widehat{Z}_n) \in \mathrm{d}(z_0, \widehat{z}_0, \ldots, z_n, \widehat{z}_n)\big) \\
&\qquad = \mu_0(\mathrm{d}z_0)\,S_{\mu_0}(z_0, \mathrm{d}\widehat{z}_0)\,M^K(\widehat{z}_0, \mathrm{d}z_1)\ldots \\
&\qquad\qquad \ldots S_{\mu_{n-1}}(z_{n-1}, \mathrm{d}\widehat{z}_{n-1})\,M^K(\widehat{z}_{n-1}, \mathrm{d}z_n) S_{\mu_n}(z_n, \mathrm{d}\widehat{z}_n).
\end{aligned}$$

Using similar arguments the equation (17) associated to the path measures $\eta_n \in \mathcal{P}(E_n)$ can be rewritten as follows:

$$\eta_{n+1} = \eta_n \mathcal{S}_{n,\eta_n} \mathcal{M}_n^K \tag{20}$$

with the collection of selection transition $\mathcal{S}_{n,\eta}$, $\eta \in \mathcal{P}(E_n)$, on E_n defined by for any $x_n \in E_n$ by

$$\mathcal{S}_{n,\eta}(x_n,\,.\,) = \varepsilon\,\mathcal{G}_n(x_n)\,\delta_{x_n}(\,.\,) + \big(1 - \varepsilon\,\mathcal{G}_n(x_n)\big)\,\Psi_n(\eta).$$

3.2 Genealogical and interacting particle models

The interacting path-particle model associated to the non linear and path-distribution model (20) is again a two-step genetic type Markov chain taking values in the product spaces E_n^N

$$\zeta_n = (\zeta_n^i)_{1\leqslant i\leqslant N} \in E_n^N \xrightarrow{\text{selection}} \widehat{\zeta}_n = \left(\widehat{\zeta}_n^i\right)_{1\leqslant i\leqslant N} \in E_n^N \xrightarrow{\text{mutation}} \zeta_{n+1} \in E_{n+1}^N.$$

During the selection stage each path-particle $\zeta_n^i = (\xi_{0,n}^i, \xi_{1,n}^i, \ldots, \xi_{n,n}^i) \in E_n$ evolves according to the Markov transition

$$\mathcal{S}_{n,\frac{1}{N}\sum_{j=1}^N \delta_{\zeta_n^j}}(\zeta_n^i, \,.\,) = \varepsilon\, \mathcal{G}_n(\zeta_n^i)\, \delta_{\zeta_n^i}(\,.\,) + \left(1 - \varepsilon\, \mathcal{G}_n(\zeta_n^i)\right) \Psi\left(\frac{1}{N}\sum_{j=1}^N \delta_{\zeta_n^j}\right)(\,.\,).$$

Since $\mathcal{G}_n(\xi_{0,n}^i, \ldots, \xi_{n,n}^i) = G(\xi_{n,n}^i)$ only depends on the fitness terminal value $\xi_{n,n}^i$ the distribution in the previous equation can be rewritten as follows:

$$\varepsilon\, G(\xi_{n,n}^i)\, \delta_{(\xi_{0,n}^i,\ldots,\xi_{n,n}^i)}(\,.\,) + \left(1 - \varepsilon\, G(\xi_{n,n}^i)\right) \sum_{j=1}^N \frac{G(\xi_{n,n}^j)}{\sum_{k=1}^N G(\xi_{n,n}^k)}\, \delta_{(\xi_{0,n}^i,\ldots,\xi_{n,n}^i)}(\,.\,).$$

The mutation mechanism simply consists in extending each selected path

$$\widehat{\zeta}_n^i = \left(\widehat{\xi}_{0,n}^i, \widehat{\xi}_{1,n}^i, \ldots, \widehat{\xi}_{n,n}^i\right) \in E_n$$

with an elementary M^K-transition, that is

$$\begin{aligned}
\zeta_{n+1}^i &= \left((\xi_{0,n+1}^i, \ldots, \xi_{n,n+1}^i), \xi_{n+1,n+1}^i\right) \\
&= \left((\widehat{\xi}_{0,n}^i, \ldots, \widehat{\xi}_{n,n}^i), \xi_{n+1,n+1}^i\right) \in E_{n+1} = E_n \times E
\end{aligned}$$

where $\xi_{n+1,n+1}^i$ is a random variable with law $M^K(\widehat{\xi}_{n,n}^i, \,.\,)$. From the above construction we also observe that the terminal values

$$\xi_n^i = \xi_{n,n}^i \qquad \text{and} \qquad \widehat{\xi}_n^i = \widehat{\xi}_{n,n}^i \tag{21}$$

of the above path-particles model coincide with the E-valued selection/mutation Markov chain defined in (10). At closer inspection we observe that the branch of ancestors only changes during the selection stage. These observations indicate that the path-valued particle model describes the evolution in time of the genealogical ancestral lines associated to the marginal model (21). It is convenient at this point to discuss the information contained in the ancestral lines. Recalling that each particle is itself a pair particle

$$\xi_n^i = (U_n^i, V_n^i) \in E = S^2,$$

it is natural to define the two components of each ancestors $\xi_{p,n}^i$ at level p in the ancestral line at time n

$$\xi_n^i = (\xi_{0,n}^i, \xi_{1,n}^i, \ldots, \xi_{n,n}^i) \in E_n = E^{n+1} = (S \times S)^{n+1}$$

as a pair particle $\xi_{p,n}^i = (U_{p,n}^i, V_{p,n}^i) \in S^2$. In an earlier study [7] we have presented several strategies to analyze the asymptotic behavior of the occupation measures of the genealogical tree

$$\eta_n^N = \frac{1}{N} \sum_{i=1}^N \delta_{(\xi_{0,n}^i, \xi_{1,n}^i, \ldots, \xi_{n,n}^i)} \longrightarrow \eta_n$$

as the size of the population $N \to \infty$. More precisely, from these results and by (15), we have for any $y \in S$ and $f_n \in \mathcal{B}_b((S \times S)^{n+1})$ the almost sure convergence

$$\begin{aligned}
&\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N f_n(\xi_{0,n}^i, \xi_{1,n}^i, \ldots, \xi_{n,n}^i) \\
&\qquad = \mathbb{E}_{(\pi\times K)_2}^L \big(f_n((Y_{n+1}, Y_n), (Y_n, Y_{n-1}), \ldots, (Y_1, Y_0)) \bigm| Y_{n+1} = y\big)
\end{aligned}$$

as soon as $\eta_0 = \delta_y \times K$. Notice that in this situation the initial configuration consists of N independent and identically distributed random pairs

$$\zeta_0^i = \xi_0^i = (U_0^i, V_0^i) = (y, V_0^i) \in S$$

with common law $\delta_y \times K$. This means that $(V_0^i)_{1\leqslant i\leqslant N}$ are N independent and identically distributed random variables with common law $K(y, .)$. From this observation we conclude that the first component $U_{0,n}^i$ of all ancestors coincides with y, that is we have that

$$\xi_{0,n}^i = (U_{0,n}^i, V_{0,n}^i) = (y, V_{0,n}^i).$$

Furthermore by (16) we have for any $h_n \in \mathcal{B}_b(S^{n+1})$ the almost sure convergence

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N h_n(U_{0,n}^i, U_{1,n}^i, \ldots, U_{n,n}^i) = \mathbb{E}_\pi^L \big(h_n(Y_n, Y_{n-1}, \ldots, Y_0) \bigm| Y_n = y\big).$$

4 Asymptotic analysis

This section focuses on the asymptotic properties of the interacting processes defined in section 3.1 and section 3.2 as the time horizon and the size of the systems tend to infinity. The first subsection is concerned with the stability properties of the non linear measure valued process (18). We provide some precise estimates of the decays to equilibrium in terms of the contraction coefficient associated to the Markov kernel L. In subsection 4.2 we analyze the asymptotic behavior of the interacting Metropolis particle model and the corresponding genealogical tree-based algorithm as the population size tends to infinity. We connect this study with the structural and the stability properties and non linear semigroup ϕ. We propose some useful uniform estimates with respect to the time parameter for the convergence of the Metropolis model towards the desired target distribution.

4.1 Stability properties

In this section we analyze the long time behavior of the Feynman–Kac distribution flow $\mu_n \in \mathcal{P}(E)$. Our main result is to describe more precisely and prove the first theorem 1 stated in the introduction. We start with two technical lemmas. The first one states that the target distribution is a fixed point of this flow. The second lemma is concerned with the dynamical structure of the semigroup ϕ. We describe the n-th iterates ϕ^n in terms of a pair of non homogeneous potential/Markov kernels (G_n, M_n). We will use these representation formulae in section 4.2 to prove uniform convergence theorems for interacting Metropolis models.

Lemma 2. *The distribution $(\pi \times K)$ is a fixed point of the flow ϕ, that is*

$$\phi(\pi \times K) = (\pi \times K).$$

Proof. By definition of ϕ we have for any $f \in \mathcal{B}_b(E)$

$$\phi(\pi \times K)(f) = \big(\pi \times K)(G\, M^K(f)\big)/(\pi \times K)(G).$$

We complete the proof by noting that

$$\begin{aligned}(\pi \times K)\big(G\, M^K(f)\big) &= \int \pi(\mathrm{d}v)\, L(v, \mathrm{d}u)\, K(v, \mathrm{d}w)\, f(v, w) \\ &= \int \pi(\mathrm{d}v)\, K(v, \mathrm{d}w)\, f(v, w) = (\pi \times K)(f). \end{aligned}$$ □

Lemma 3. *For any $\mu \in \mathcal{P}(E)$, $n \geqslant 0$ and $f \in \mathcal{B}_b(E)$ we have that*

$$\phi^n(\mu)(f) = \mu\big(G_n\, M_n(f)\big)/\mu(G_n)$$

with the potential functions G_n and the Markov kernels on E defined by

$$G_n = Q^n(1) \qquad \textit{and} \qquad M_n(f) = Q^n(f)/Q^n(1)$$

with the convention $(G_0, M_0) = (1, \mathrm{Id})$ and $(G_1, M_1) = (G, M^K)$ for $n = 0$ and $n = 1$. In addition for any $n \geqslant 1$ and $x = (u, v)$ we have

$$G_{n+1}(u, v) = G(u, v)\, \frac{\mathrm{d}\pi L^n}{\mathrm{d}\pi}(v)$$

and

$$M_n(f)(u, v) = \mathbb{E}^L_{(\pi\times K)_2}\big(f(Y_1, Y_0) \mid (Y_{n+1}, Y_n) = (u, v)\big).$$

Proof. Since we have for any $n \geqslant 0$

$$\phi^n(\mu)(f) = \mu\big(Q^n(f)\big)/\mu(Q^n(1))$$

the first assertion is clear. To check that G_n and M_n have the form indicated above we first observe that

$$Q^n(f)(u,v) = \mathbb{E}^K_{(u,v)}\bigg(f(X_n)\prod_{p=0}^{n-1} G(X_p)\bigg).$$

By proposition 1 this yields that for any $n \geqslant 1$

$$Q^n(f)(u,v) = G(u,v)\,\frac{\mathrm{d}\pi L^{n-1}}{\mathrm{d}\pi}(v)\,\mathbb{E}^L_{(\pi\times K)_2}\big(f(Y_1,Y_0)\;\big|\;(Y_{n+1},Y_n)=(u,v)\big)$$

from which we find that

$$G_n(u,v) = Q^n(1)(u,v) = G(u,v)\,\frac{\mathrm{d}\pi L^{n-1}}{\mathrm{d}\pi}(v)$$

and for $n \geqslant 1$

$$M_n(f)(u,v) = \frac{Q^n(f)(u,v)}{Q^n(1)(u,v)} = \mathbb{E}^L_{(\pi\times K)_2}\big(f(Y_1,Y_0)\;\big|\;(Y_{n+1},Y_n)=(u,v)\big).$$

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

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

Theorem 3. *Suppose there exists an integer $m \geqslant 1$ and a finite $l_m < \infty$ such that*

$$L^m(u,\,.\,) \leqslant l_m\, L^m(v,\,.\,). \tag{22}$$

Then, for any $n \geqslant 0$ we have the uniform estimate

$$\begin{aligned}\beta(M_{m+n+1}) &= \sup_{\mu_1,\mu_2}\|\phi^{m+n+1}(\mu_1)-\phi^{m+n+1}(\mu_2)\|_{\mathrm{tv}}\\ &\leqslant \sup_v \int \pi(\mathrm{d}u)\left|\frac{\mathrm{d}L^{m+n}(u,\,.\,)}{\mathrm{d}\pi L^{m+n}}(v)-1\right| \leqslant 2l_m\,\beta(L^n).\end{aligned}$$

Proof. It is immediate from lemma 3 to check that $\phi^n(\delta_x) = M_n(x,\,.\,)$ for any $x \in E$ and $n \geqslant 1$. This yields that

$$\sup_{\mu_1,\mu_2}\|\phi^n(\mu_1)-\phi^n(\mu_2)\|_{\mathrm{tv}} \geqslant \sup_{x_1,x_2\in S}\|M_n(x_1,\,.\,)-M_n(x_2,\,.\,)\|_{\mathrm{tv}} = \beta(M_n).$$

To check the reverse inequality, one uses the decomposition

$$\begin{aligned}&\phi^n(\mu_1)(f)-\phi^n(\mu_2)(f)\\ &\qquad= \int\big(M_n(f)(x_1)-M_n(f)(x_2)\big)\,\frac{G_n(x_1)}{\mu_1(G_n)}\,\frac{G_n(x_2)}{\mu_2(G_n)}\,\mu_1(\mathrm{d}x_1)\,\mu_2(\mathrm{d}x_2).\end{aligned}$$

We conclude that for any $n \geqslant 1$

$$\beta(M_n) = \sup_{\mu_1,\mu_2} \|\phi^n(\mu_1) - \phi^n(\mu_2)\|_{\mathrm{tv}}.$$

Under our assumptions we also observe that for any $y \in S$ and $n \geqslant m$ we have $L^n(y, \,.\,) \ll \pi L^n$ and for any $y' \in S$

$$\frac{\mathrm{d}L^n(y, \,.\,)}{\mathrm{d}\pi L^n}(y') \in [l_m^{-1}, l_m].$$

Using lemma 3 we prove that for any pair $(u,v) \in E = S^2$, $f \in \mathcal{B}_b(E)$ and $n \geqslant 0$

$$\begin{aligned} M_{m+n+1}(f)(u,v) &- (\pi \times K)(f) \\ &= \int \pi(\mathrm{d}y)\, K(y, \mathrm{d}z)\Big(\frac{\mathrm{d}L^{m+n}(y, \,.\,)}{\mathrm{d}\pi L^{m+n}}(v) - 1\Big) f(y,z). \end{aligned}$$

This yields the estimate

$$\big\|M_{m+n+1}\big((u,v), \,.\,\big) - (\pi \times K)\big\|_{\mathrm{tv}} \leqslant \frac{1}{2}\int \pi(\mathrm{d}y)\left|\frac{\mathrm{d}L^{m+n}(y, \,.\,)}{\mathrm{d}\pi L^{m+n}}(v) - 1\right|$$

and the end of the proof of the first inequality is now clear. To prove the final part of the theorem we use the easily checked inequality

$$\begin{aligned} &\left|\frac{\mathrm{d}L^{n+m}(u_1, \,.\,)}{\mathrm{d}\pi L^{m+n}}(v) - 1\right| \\ &\qquad \leqslant \int \pi(\mathrm{d}u_2)\left|\int \big(L^n(u_1, \mathrm{d}u) - L^n(u_2, \mathrm{d}u)\big)\, \frac{\mathrm{d}L^m(u, \,.\,)}{\mathrm{d}\pi L^{m+n}}(v)\right|. \end{aligned}$$

From this equation and under our assumptions we get

$$\begin{aligned} &\int \pi(\mathrm{d}u_1)\left|\frac{\mathrm{d}L^{n+m}(u_1, \,.\,)}{\mathrm{d}\pi L^{n+m}}(v) - 1\right| \\ &\qquad \leqslant 2\beta(L^n)\sup_{u\in S}\left|\frac{\mathrm{d}L^m(u, \,.\,)}{\mathrm{d}\pi L^{n+m}}(v)\right| \leqslant 2l_m\, \beta(L^n). \end{aligned}$$

This ends the proof of the theorem. $\square$

We end this section with some comments on the stability properties presented in theorem 3. When an m-iterate L^m satisfies the regularity condition (22) for some finite $l_m < \infty$ we first observe that

$$\beta(L^m) = \sup_{u,v\in S} \|L^m(u, \,.\,) - L^m(v, \,.\,)\|_{\mathrm{tv}} \leqslant (1 - 1/l_m).$$

From this observation we conclude that

$$\beta(L^n) \leqslant \prod_{p=1}^{[n/m]} \beta(L^m) \leqslant (1 - 1/l_m)^{[n/m]}$$

where $[a]$ stands for the integer part of a number $a \in \mathbb{R}$. This yields the uniform estimate

$$\beta(M_{n+1}) = \sup_{\mu_1,\mu_2} \|\phi^{n+1}(\mu_1) - \phi^{n+1}(\mu_2)\|_{\mathrm{tv}} \leqslant 2l_m\, \beta(L^n) \leqslant 2l_m\, (1{-}1/l_m)^{[n/m]}.$$

This estimate gives a way to calibrate in practice the convergence of the non linear Feynman–Kac model to the desired limiting distribution. Nevertheless this bound is rather crude and the regularity condition we have made on L is often related to a compactness assumption on the state space S. For these two reasons it may be more judicious in some instances to use the estimates presented in theorem 3.

4.2 A Convergence theorem

In this final section we analyze the convergence of the interacting Metropolis model and its genealogical tree structure when the size of the systems tends to infinity. We recall that the path-particle approximating measures associated to the genealogical tree based algorithm

$$\zeta_n^i = (\xi_{0,n}^i, \xi_{1,n}^i, \ldots, \xi_{n,n}^i)$$

described in section 3.2 are defined by

$$\eta_n^N = \frac{1}{N} \sum_{i=1}^N \delta_{(\xi_{0,n}^i, \xi_{1,n}^i, \ldots, \xi_{n,n}^i)}.$$

We also recall that their n-time marginals correspond to the particle approximating measures associated to the interacting particle Metropolis model ξ_n^i. They are given by

$$\mu_n^N = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_n^i}.$$

These particle approximating distributions can be regarded as the occupation measures associated to an interacting particle approximation of a class of Feynman–Kac distribution flow. Next we quote without proof a preliminary technical lemma which holds for general particle approximating models of this type. The $\mathbb{L}_p$ mean error estimates presented here have been proved originally in [6] for general Feynman–Kac evolution equations of the form (18) with a selection transition kernel of the form (19) with $\varepsilon = 0$. The extension to a more general class of transitions has been developed in [5]. The application of these $\mathbb{L}_p$-estimates for path-valued particles and related genealogical tree based models can be found in [7]. In our context and using the dynamical structure properties of the semigroup ϕ presented in lemma 3 these estimates take the following form. We slight abuse notations and we write $\mathbb{E}^{(N)}(\,.\,)$, for the expectation with respect to the law of the particle model ξ_n^i with initial distribution $\mu_0^{\otimes N}$.

Lemma 4. *For any $p \geqslant 1$, $n \geqslant 0$ and for any $f \in \mathcal{B}_b(E)$ with $\|f\| \leqslant 1$ we have*

$$\sqrt{N}\ \mathbb{E}^{(N)}\big(|\mu_n^N(f) - \mu_n(f)|^p\big)^{1/p} \leqslant c_p \sum_{p=0}^{n} \sup_{x_1,x_2} \big(G_p(x_1)/G_p(x_2)\big)^2 \beta(M_p)$$

for some universal constant c_p whose values only depend on the parameter $p \geqslant 1$. In addition for any $f_n \in \mathcal{B}_b(E_n)$ with $\|f_n\| \leqslant 1$ we have

$$\sqrt{N}\ \mathbb{E}^{(N)}\big(|\eta_n^N(f_n) - \eta_n(f_n)|^p\big)^{1/p} \leqslant c_p \sum_{p=0}^{n} \sup_{x_1,x_2} \big(G_p(x_1)/G_p(x_2)\big)^2.$$

This lemma connects the convergence of the particle approximating measures η_n^N and μ_n^N with the oscillations of the potential functions G_n and the contraction properties of the Markov kernels M_n.

In theorem 3 we have presented two estimates of $\beta(M_n)$ in terms of the regularity properties of the pair (π, L). In particular when the m-iterate L^m satisfies the inequality (22) for some finite constant $l_m < \infty$ then we have seen that

$$\beta(M_{n+1}) \leqslant 2l_m\,(1 - 1/l_m)^{[n/m]}.$$

Our next objective is to analyze the oscillations of the non homogeneous potential functions G_n. To this end we need to strengthen the regularity condition we made on L. When (22) is met we recall that L has a unique invariant measure $\nu = \nu L \in \mathcal{P}(E)$.

Lemma 5. *For any $n \geqslant 0$ and for any $x, x' \in E$ we have*

$$G_n(x) \leqslant g^n\, G_n(x').$$

In addition suppose there exists some integer $m \geqslant 1$ and a pair of finite constants l_m, $k_\pi < \infty$ such that

$$L^m(u,\,.\,) \leqslant l_m\, L^m(v,\,.\,) \qquad \text{and} \qquad \frac{d\nu}{d\pi}(u)\,\frac{d\pi}{d\nu}(v) \leqslant k_\pi \tag{23}$$

where $\nu = \nu L$ is the invariant distribution of L. In this situation we have for any $n \geqslant (m+1)$ the uniform estimate

$$G_n(x)/G_n(x') \leqslant g\, l_m^2\, k_\pi.$$

Proof. Using lemma 3 we have for any $x, x' \in E$

$$\frac{G_n(x)}{G_n(x')} = \frac{Q^n(1)(x)}{Q^n(1)(x')} = \frac{\mathbb{E}_x^K\left(\prod_{p=0}^{n-1} G(X_p)\right)}{\mathbb{E}_{x'}^K\left(\prod_{p=0}^{n-1} G(X_p)\right)} \leqslant g^n$$

This ends the proof of the first assertion. To prove the uniform estimate we observe that for any $u, v \in S$

$$\frac{1}{l_m}\,\frac{\mathrm{d}L^m(v,\,.\,)}{\mathrm{d}\pi}(u) \leqslant \frac{\mathrm{d}\nu}{\mathrm{d}\pi}(u) = \int \nu(\mathrm{d}y)\,\frac{\mathrm{d}L^m(y,\,.\,)}{\mathrm{d}\pi}(u) \leqslant l_m \frac{\mathrm{d}L^m(v,\,.\,)}{\mathrm{d}\pi}(u).$$

Therefore for any $n \geqslant 0$ and $u \in S$ we find that

$$l_m^{-1}\,\frac{\mathrm{d}\nu}{\mathrm{d}\pi}(u) \leqslant \frac{\mathrm{d}\pi L^{n+m}}{\mathrm{d}\pi}(u) = \int \pi L^n(\mathrm{d}v)\,\frac{\mathrm{d}L^m(v,\,.\,)}{\mathrm{d}\pi}(u) \leqslant l_m\,\frac{\mathrm{d}\nu}{\mathrm{d}\pi}(u).$$

We conclude that for any $n \geqslant 0$ and $u, v, u', v' \in S$

$$\begin{aligned}\frac{G_{n+m+1}(u,v)}{G_{n+m+1}(u',v')} &= \frac{G(u,v)}{G(u',v')}\,\frac{\mathrm{d}\pi L^{n+m}}{\mathrm{d}\pi}(v)\,\frac{\mathrm{d}\pi}{\mathrm{d}\pi L^{n+m}}(v') \\ &\leqslant g\,l_m^2\,\frac{\mathrm{d}\nu}{\mathrm{d}\pi}(v)\,\frac{\mathrm{d}\pi}{\mathrm{d}\nu}(v') \leqslant g\,l_m^2\,k_\pi.\end{aligned}$$

This completes the proof of the lemma. □

Theorem 4. *Suppose the pair (π, L) satisfies the regularity conditions (23) for some $m \geqslant 1$ and some pair of finite constant (l_m, k_π). Then for any $n \geqslant 0$ and $f \in \mathcal{B}_b(E)$ with $\|f\| \leqslant 1$ we have the $\mathbb{L}_p$-mean error estimates*

$$\mathbb{E}^{(N)}\left(|\mu_{n+m+1}^N(f) - (\pi \times K)(f)|^p\right)^{1/p} \leqslant c_p\,\frac{a_{\pi,L}}{\sqrt{N}} + b_L\,\beta(L^n)$$

for some universal constant $c_p < \infty$ which only depends on the parameter p and

$$a_{\pi,L} \leqslant (m+1)g^{2m} + 2m(gl_m^3 k_\pi)^2 \qquad \text{and} \qquad b_L \leqslant 2l_m.$$

In addition for any $f_n \in \mathcal{B}_b(E_n)$ with $\|f_n\| \leqslant 1$ we have that

$$\sqrt{N}\;\mathbb{E}^{(N)}\left(|\eta_n^N(f_n) - \eta_n(f_n)|^p\right)^{1/p} \leqslant c_p\,d(n)$$

with

$$d(n) \leqslant (m+1)g^{2m} + 2\big((n-m) \vee 0\big)\,(gl_m^3 k_\pi)^2.$$

Proof. From the potential oscillation estimates presented in lemma 5 we obtain the upper bound

$$\begin{aligned}&\sum_{p=0}^{n+m+1}\ \sup_{x_1,x_2}\left(G_p(x_1)/G_p(x_2)\right)^2\beta(M_p) \\ &\qquad\qquad \leqslant \sum_{p=0}^{m}\sup_{x_1,x_2}\left(G_p(x_1)/G_p(x_2)\right)^2 + (gl_m^2 k_\pi)^2\sum_{p=0}^{n}\beta(M_{m+p+1}).\end{aligned}$$

Since $G_0 = 1$ and $G_1 = G$ we have

$$\sum_{p=0}^{m} \sup_{x_1,x_2} \big(G_p(x_1)/G_p(x_2)\big)^2 \leqslant (m+1)g^{2m}.$$

Using the contraction estimates stated in theorem 3 this implies that

$$\sum_{p=0}^{n+m+1} \sup_{x_1,x_2} \big(G_p(x_1)/G_p(x_2)\big)^2 \beta(M_p) \leqslant (m+1)^{2m} + 2l_m(gl_m^2 k_\pi)^2 \sum_{p\geqslant 0} \beta(L^p).$$

Noting that

$$\sum_{p\geqslant 0} \beta(L^p) = \sum_{q\geqslant 0} \sum_{p=qm}^{(q+1)m-1} \beta(L^p) \leqslant m \sum_{q\geqslant 0} \beta(L^m)^q = \frac{m}{1-\beta(L^m)} \leqslant ml_m$$

it is now evident that

$$\sum_{p=0}^{n+m+1} \sup_{x_1,x_2} \big(G_p(x_1)/G_p(x_2)\big)^2 \beta(M_p) \leqslant (m+1)g^{2m} + 2ml_m^2(gl_m^2 k_\pi)^2.$$

The end of the proof of the first estimate is now a simple application of lemma 4 and theorem 3. The second estimate is proved along the same lines of arguments and the end of the proof is now straightforward. □

Acknowledgement. We are grateful to Laurent Miclo for his careful reading of the original manuscript. His comments and suggestions have greatly improved the presentation of the paper.

References

1. O. Catoni, Simulated annealing algorithms and Markov chains with rare transitions. *Séminaire de Probabilités, XXXIII, Lecture Notes in Math., 1709*, S pringer, Berlin, pp. 69–119 (1999).
2. O. Catoni, Metropolis, simulated annealing, and iterated energy transformation algorithms: theory and experiments. *Special issue for the Foundations of Computational Mathematics Conference, J. Complexity*, vol. 12, no. 4, pp. 595–623, (1996).
3. O. Catoni, Sharp large deviations estimates for simulated annealing algorithms. *Ann. Inst. H. Poincaré Probab. Statist.*, vol. 27, no. 3, pp. 291–383 (1991).
4. D. Chauveau, P. Vandekerkhove, Interacting Hastings–Metropolis algorithms, *C. R. Acad. Sci. Paris Sér. I Math.* vol. 333, no. 9, 881–884 (2001).
5. P. Del Moral and M.A. Kouritzin and L. Miclo, On a class of discrete generation interacting particle systems, *Electronic Journal of Probability*, vol. 6, no. 16, pp. 1-26, (2001).

6. P. Del Moral et L. Miclo, Branching and Interacting Particle Systems Approximations of Feynman–Kac Formulae with Applications to Non-Linear Filtering, *Séminaire de Probabilités XXXIV. Lecture Notes in Mathematics No. 1729*, J. Azéma, M. Émery, M. Ledoux, M. Yor (Eds.), pp. 1-145, 2000.
7. P. Del Moral and L. Miclo, Genealogies and Increasing Propagations of Chaos for Feynman–Kac and Genetic Models. *Annals of Applied Probability,*, vol. 11, No. 4, pp. 1166-1198 (2001).
8. P. Del Moral and L. Miclo, Annealing Properties of Feynman–Kac Models, *Publications du Laboratoire de Statistiques et Probabilités, Toulouse III (2002).*
9. P. Del Moral and T. Zajic, A note on Laplace-Varadhan's integral lemma. Publication n°LSP-CNRS, 2001-09, Univ. Paul Sabatier, Toulouse. To appear in *Bernoulli* (2003).
10. A. Doucet, N. de Freitas and N.J. Gordon, *Sequential Monte Carlo Methods in Practice,* New York: Springer-Verlag, (2001).
11. P. Diaconis, L. Saloff-Coste, What do we know about the Metropolis algorithm? *27th Annual ACM Symposium on the Theory of Computing (STOC'95) (Las Vegas, NV). J. Comput. System Sci.*, vol. 57, no. 1, pp. 20–36 (1998).
12. I. Gaudron, A. Trouvé, Fluctuations of empirical means at low temperature for finite Markov chains with rare transitions in the general case., *Probab. Theory Related Fields*, vol. 111, no. 2, pp. 215–251 (1998).
13. W.R. Gilks, S. Richardson, D.J. Spielgelhalter, *Monte Carlo Markov chain in Practice*, Chapman and Hall, London (1996).
14. W.K. Hastings, Monte Carlo Sampling Methods using Markov Chains and Their Applications, *Biometrika*, vol. 57, pp. 97-109, (1970).
15. S.F. Jarner E. Hansen, Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.*, vol.85, no. 2, pp. 341–361, (2000).
16. K.L. Mengersen, R.L. Tweedie, Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.*, vol. 24, no. 1, pp. 101–121 (1996).
17. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E.Teller, Equation of State Calculations by Fast Computing Machines, *Journal of Chemical Physics*, vol. 90, pp. 233-241, (1953).
18. L. Miclo, On inhomogeneous discrete exit problems, *Ann. Appl. Probab.* vol. 6, no. 4, pp. 1112–1156 (1996).
19. L. Miclo, Partial simulated annealing, *Stochastic Process. Appl.*, vol. 65, no. 2, pp. 281–298 (1996).
20. L. Miclo, Occupation times of low-temperature nonhomogeneous finite Markov processes, *Stochastics Stochastics Rep.*, vol. 63, no. 1-2, pp. 65–137 (1998).
21. A. Trouvé, Rough large deviation estimates for the optimal convergence speed exponent of generalized simulated annealing algorithms, *Ann. Inst. H. Poincaré Probab. Statist.*, vol. 32, no. 3, pp. 299–348 (1996).