The Annals of Applied Probability
2010, Vol. 20, No. 3, 841–868
DOI: 10.1214/09-AAP636

A CAUTIONARY TALE ON THE EFFICIENCY OF SOME ADAPTIVE MONTE CARLO SCHEMES

BY YVES F. ATCHADÉ

University of Michigan

There is a growing interest in the literature for adaptive Markov chain Monte Carlo methods based on sequences of random transition kernels $\{P_n\}$ where the kernel P_n is allowed to have an invariant distribution π_n not necessarily equal to the distribution of interest π (target distribution). These algorithms are designed such that as $n \to \infty$, P_n converges to P, a kernel that has the correct invariant distribution π. Typically, P is a kernel with good convergence properties, but one that cannot be directly implemented. It is then expected that the algorithm will inherit the good convergence properties of P. The equi-energy sampler of [*Ann. Statist.* **34** (2006) 1581–1619] is an example of this type of adaptive MCMC. We show in this paper that the asymptotic variance of this type of adaptive MCMC is always at least as large as the asymptotic variance of the Markov chain with transition kernel P. We also show by simulation that the difference can be substantial.

1. Introduction. Adaptive Markov chain Monte Carlo (AMCMC) is an approach to Markov chain Monte Carlo (MCMC) simulation where the transition kernel of the algorithm is allowed to change over time as an attempt to improve efficiency. It grows out of the seminal works of [11, 12]. Let π be the distribution of interest. The problem is to sample efficiently from π given a family of Markov kernels $\{P_\theta, \theta \in \Theta\}$. This can be solved adaptively using a joint process $\{(X_n, \theta_n), n \geq 0\}$ such that the conditional distribution of X_{n+1} given the information available up to time n is P_{θ_n} and where θ_n is adaptively tuned over time. Some general sufficient conditions for the convergence of such algorithms can be found in [6, 18]. It is also shown in [1] that under some regularity conditions, if a "best" limiting kernel P_{θ^*} exists, the marginal chain $\{X_n, n \geq 0\}$ in the joint adaptive process behaves in many ways like a standard Markov chain with transition kernel P_{θ^*}. In all the above-mentioned papers, the assumption that each P_θ has invariant distribution π plays an important role.

More recently, interest has emerged in building Monte Carlo algorithms where the transition kernel P_n used at time n has invariant distribution π_n not necessarily equal to π. These algorithms are designed such that as $n \to \infty$, P_n converges to a transition kernel P which is invariant with respect to π. This limiting kernel P is

Received September 2007; revised August 2009.
AMS 2000 subject classifications. Primary 60C05, 60J27, 60J35, 65C40.
Key words and phrases. Monte Carlo methods, adaptive MCMC, equi-energy sampler, martingale approximation, central limit theorems, importance resampling.

typically a very efficient kernel that would be difficult to implement otherwise. The interest of this approach is that as $n \to \infty$, P_n approaches P and one expects the algorithm to inherit the good convergence properties of P. The equi-energy (EE) sampler of [15] is an example. Another example based on importance resampling appeared independently in [3] and [5].

This paper provides a detailed analysis of the law of large numbers and central limit theorem for the EE sampler. It is also an attempt to address the question of whether such algorithms can deliver the same performance as their limiting kernel P. We give a negative answer. We show, in the case of the EE sampler, that its asymptotic variance is always at least as large as the asymptotic variance of the limiting transition kernel P. The difference can be substantial and we illustrate this with a simulation example.

On the related literature, the law of large numbers for of the EE sampler has been studied in [3] but using different techniques than those in this work. We also mention a new class of interacting MCMC algorithms proposed by [8, 10] for solving numerically some discrete-time measure-valued equations. These algorithms share the same framework with the EE sampler. In these two papers, the authors develop a number of asymptotic results for interacting MCMC including a strong law of large numbers and a central limit theorem.

The paper is organized as follows. In Section 2 we present the EE sampler and IR-MCMC in a slightly more general framework. The limit theorems are developed in Section 3 and proved in Section 4. The main ingredient of the proofs is the martingale approximation method. We present a simulation example in Section 3.5 comparing these algorithms to a Random Walk Metropolis algorithm.

2. A class of adaptive Monte Carlo algorithms.

Let $(\mathcal{X}, \mathcal{B}, \lambda)$ be a reference Polish space equipped with its Borel σ-algebra $\mathcal{B}$ and a σ-finite measure λ and $K \geq 1$ an integer. We denote by $\mathcal{M}$ the set of all probability measure on $(\mathcal{X}, \mathcal{B})$. Let $\{\pi^{(l)}, l = 0, \ldots, K\}$ be probability measures on $(\mathcal{X}, \mathcal{B})$ such that

$$\pi^{(l)}(dx) = \frac{1}{Z_l} e^{-E_l(x)} \lambda(dx) \tag{1}$$

for some measurable functions $E_l : (\mathcal{X}, \mathcal{B}) \to \mathbb{R}$. $Z_l := \int_{\mathcal{X}} e^{-E_l(x)} \lambda(dx)$ (assumed finite) is the normalizing constant. We study a class of Monte Carlo algorithms to sample from the family $\{\pi^{(l)}\}$. These algorithms will generated an ergodic random process $\{(X_n^{(0)}, \ldots, X_n^{(K)}), n \geq 0\}$ on $\mathcal{X}^{K+1}$ with limiting distribution $\pi^{(0)} \times \cdots \times \pi^{(K)}$.

We introduce some notation in order to describe the algorithm. Whenever necessary and without further notice, any subset of $\mathbb{R}^d$ will be equipped with its Borel σ-algebra. If $(\mathcal{Y}, \mathcal{E})$ and $(\mathcal{Z}, \mathcal{F})$ are two measurable spaces, a kernel from $(\mathcal{Y}, \mathcal{E})$ to $(\mathcal{Z}, \mathcal{F})$ is any function $P : \mathcal{Y} \times \mathcal{F} \to [0, 1]$ such that $P(y, \cdot)$ is a probability measure on $(\mathcal{Z}, \mathcal{F})$ for all $y \in \mathcal{Y}$ and $P(\cdot, A)$ is a measurable map for all $A \in \mathcal{F}$.

If $(\mathcal{Y}, \mathcal{E}) = (\mathcal{Z}, \mathcal{F})$, we call P a kernel on $(\mathcal{Z}, \mathcal{F})$. If P is a kernel from $(\mathcal{Y}, \mathcal{E})$ to $(\mathcal{Z}, \mathcal{F})$, $f : (\mathcal{Z}, \mathcal{F}) \to \mathbb{R}$ a measurable function and $y \in \mathcal{Y}$, we shall use the notation $P(y, f)$ or $Pf(y)$ to denote the integral $\int_{\mathcal{Z}} P(y, dz) f(z)$ whenever it is well defined.

2.1. *A general algorithm.* Let $\{P^{(l)}, l = 0, \dots, K\}$ be kernels on $(\mathcal{X}, \mathcal{B})$ such that $\pi^{(l)}$ is the invariant distribution of $P^{(l)}$. Let $\{T^{(l)}, l = 1, \dots, K\}$ be kernels from $(\mathcal{X}^2, \mathcal{B}^2)$ to $(\mathcal{X}, \mathcal{B})$, $\{\omega^{(l)}, l = 1, \dots, K\}$ positive real-valued measurable functions defined on $(\mathcal{X}^2, \mathcal{B}^2)$ and $\theta_l \in (0, 1)$ for $l = 1, \dots, K$. For $\mu \in \mathcal{M}$ and $l = 1, \dots, K$, we define the following kernel on $(\mathcal{X}, \mathcal{B})$:

$$\begin{aligned} P_\mu^{(l)}(x, A) &= \theta_l P^{(l)}(x, A) \\ &\quad + (1 - \theta_l) \frac{\int \mu(dy) \omega^{(l)}(y, x) T^{(l)}(y, x, A)}{\int \mu(dy) \omega^{(l)}(y, x)}, \qquad x \in \mathcal{X}, A \in \mathcal{B}. \end{aligned} \tag{2}$$

For $n \geq 1$, we introduce the maps $H_n : \mathcal{M} \times \mathcal{X} \to \mathcal{M}$ defined as $H_n(\mu, x) = \mu + n^{-1}(\delta_x - \mu)$, where δ_x is the Dirac measure. Let $\{(X_n^{(0)}, \dots, X_n^{(K)}, \mu_n^{(0)}, \dots, \mu_n^{(K-1)}), n \geq 0\}$ be the nonhomogeneous Markov chain on $\mathcal{X}^{K+1} \times \mathcal{M}^K$ [defined on some probability space $(\Omega, \mathcal{F})$ that can be taken as the canonical space $(\mathcal{X}^{K+1} \times \mathcal{M}^K)^\infty$] with sequence of transition kernels $\bar{P}_n$ given by

$$\begin{aligned} &\bar{P}_n\big((x^{(0)}, \dots, x^{(K)}, \mu^{(0)}, \dots, \mu^{(K-1)}); \\ &\quad \big(dy^{(0)}, \dots, dy^{(K)}, d\nu^{(0)}, \dots, d\nu^{(K-1)}\big)\big) \\ &\quad = P^{(0)}\big(x^{(0)}, dy^{(0)}\big) \prod_{l=1}^{K} P_{\mu^{(l-1)}}^{(l)}\big(x^{(l)}, dy^{(l)}\big) \prod_{l=0}^{K-1} \delta_{H_n(\mu^{(l)}, y^{(l)})}\big(d\nu^{(l)}\big). \end{aligned} \tag{3}$$

Throughout, we denote $\{\mathcal{F}_n, n \geq 0\}$ the natural filtration of the process. We will assume that the initial value of the process is fixed. For simplicity we take $\mu_0^{(l)} = 0$. Finally, we call $\mathbb{P}$ and $\mathbb{E}$ the probability distribution and expectation of the process.

Algorithmically, $\{(X_n^{(0)}, \dots, X_n^{(K)}, \mu_n^{(0)}, \dots, \mu_n^{(K-1)}), n \geq 0\}$ can be described as follows.

ALGORITHM 2.1. At time n and given $\{(X_k^{(0)}, \dots, X_k^{(K)}, \mu_k^{(0)}, \dots, \mu_k^{(K-1)}), k \leq n - 1\}$:

1. Generate $X_n^{(0)} \sim P^{(0)}(X_{n-1}^{(0)}, \cdot)$.
2. For $l = 1, \dots, K$, generate independently $X_n^{(l)}$ from $P_{\mu_{n-1}^{(l-1)}}^{(l)}(X_{n-1}^{(l)}, \cdot)$ as given by (2).
3. For $l = 0, \dots, K - 1$, set $\mu_n^{(l)} = H_n(\mu_{n-1}^{(l)}, X_n^{(l)}) = \mu_{n-1}^{(l)} + n^{-1}(\delta_{X_n^{(l)}} - \mu_{n-1}^{(l)})$.

The heuristic of the algorithm is the following. By construction, $\{X_n^{(0)}, \mathcal{F}_n\}$ is a Markov chain with kernel $P^{(0)}$ and invariant distribution $\pi^{(0)}$. If this chain is ergodic, then as $n \to \infty$, $\mathbb{P}(X_n^{(1)} \in A|\mathcal{F}_{n-1}) = P^{(1)}_{\mu_{n-1}^{(l-1)}}(X_{n-1}^{(l)}, A)$, will converge to $K^{(1)}$ where $K^{(l)}$ is given by

$$K^{(l)}(x, A) = \theta_l P^{(l)}(x, A) + (1-\theta_l)\frac{1}{z^{(l)}(x)}\int_{\mathcal{X}} \pi^{(l-1)}(dy)\omega^{(l)}(x, y)T^{(l)}(y, x, A), \tag{4}$$

where $z^{(l)}(x) = \int_{\mathcal{X}} \pi^{(l-1)}(dy)\omega^{(l)}(x, y)$. We will discuss below two ways of choosing $\omega^{(l)}$ and $T^{(l)}$ so that $K^{(l)}$ has invariant distribution $\pi^{(l)}$. With these choices we can reasonably expect $\{X_n^{(1)}\}$ to be ergodic with limiting distribution $\pi^{(1)}$. The same argument can then be repeated. In other words, with appropriate choice of $\omega^{(l)}$ and $T^{(l)}$, the marginal process $\{X_n^{(l)}, n \geq 0\}$ can be used for Monte Carlo simulation from $\pi^{(l)}$.

2.2. *Importance-resampling MCMC*. For $l = 1, \ldots, K$ define the importance function

$$r^{(l)}(x) = \exp\big(E_{l-1}(x) - E_l(x)\big).$$

In Algorithm 2.1 we can take $\omega^{(l)}(x, y) = r^{(l)}(y)$ and $T^{(l)}(y, x, A) = T_0^{(l)}(y, A)$, where $T_0^{(l)}$ is some kernel on $(\mathcal{X}, \mathcal{B})$ with invariant distribution $\pi^{(l)}$. This leads to the IR-MCMC algorithm ([3, 5]). In this case, step 2 of Algorithm 2.1 can be described as follows: with probability θ_l we sample $X_n^{(l)}$ from $P^{(l)}(X_{n-1}^{(l)}, \cdot)$ and with probability $1 - \theta_l$, we obtain $Y^{(l)}$ by resampling from $\{X_0^{(l-1)}, \ldots, X_{n-1}^{(l-1)}\}$ with weights $\{r^{(l)}(X_0^{(l-1)}), \ldots, r^{(l)}(X_{n-1}^{(l-1)})\}$ and then propose $X_n^{(l)} \sim T_0^{(l)}(Y^{(l)}, \cdot)$.

The lth limiting kernel here takes the form

$$K^{(l)}(x, A) = \theta_l P^{(l)}(x, A) + (1-\theta_l)\pi^{(l)}(A)$$

has invariant distribution $\pi^{(l)}$ and has better mixing than $P^{(l)}$. But direct sampling from $K^{(l)}$ is impossible as it requires that we be able to sample from $\pi^{(l)}$ which is the problem that we are trying to solve in the first place.

2.3. *The EE sampler*. Taking $\omega^{(l)}(x, y) \equiv 1$ and

$$T^{(l)}(y, x, A) = \min\left(1, \frac{r_l(y)}{r_l(x)}\right)\mathbf{1}_A(y) + \left(1 - \min\left(1, \frac{r_l(y)}{r_l(x)}\right)\right)\mathbf{1}_A(x), \tag{5}$$

in (2), we get the EE sampler ([15]). In this case the limiting kernel becomes

$$\begin{aligned} K^{(l)}(x, A) &= \theta_l P^{(l)}(x, A) + (1-\theta_l)\int_{\mathcal{X}} \pi^{(l-1)}(dy)T^{(l)}(y, x, A) \\ &= \theta_l P^{(l)}(x, A) + (1-\theta_l)R^{(l)}(x, A), \end{aligned} \tag{6}$$

where $R^{(l)}$ is the kernel of the Metropolis–Hastings algorithm with proposal $\pi^{(l-1)}$ and target distribution $\pi^{(l)}$:

$$R^{(l)}(x, A) = \int_A \min\left(1, \frac{r^{(l)}(y)}{r^{(l)}(x)}\right)\pi^{(l-1)}(dy) + \left[1 - \int_{\mathcal{X}} \min\left(1, \frac{r^{(l)}(y)}{r^{(l)}(x)}\right)\pi^{(l-1)}(dy)\right]\mathbf{1}_A(x).$$

Clearly, $K^{(l)}$ has invariant distribution $\pi^{(l)}$. In general $K^{(l)}$ will converge faster than $P^{(l)}$. For example if $E_l - E_{l-1}$ is bounded from below it is easy to show that $K^{(l)}$ is always uniformly ergodic, independently of $P^{(l)}$.

For the EE sampler, step 2 of Algorithm 2.1 can now be described as follows. With probability θ_l we sample $X_n^{(l)}$ from $P^{(l)}(X_{n-1}^{(l)}, \cdot)$ and with probability $1 - \theta_l$, we obtain $Y^{(l)}$ by resampling uniformly from $\{X_k^{(l-1)} : k \le n-1\}$. Then $Y^{(l)}$ is accepted with probability $\min(1, \frac{r^{(l)}(Y^{(l)})}{r^{(l)}(X_{n-1}^{(l)})})$ in which case we set $X_n^{(l)} = Y^{(l)}$; otherwise $Y^{(l)}$ is rejected and we set $X_n^{(l)} = X_{n-1}^{(l)}$.

Actually the EE sampler described above is a simplified version of [15]. Their original algorithm uses an idea of partitioning. Let $\{\mathcal{X}_i, i = 1, \dots, d\}$ be a partition of $\mathcal{X}$ (in [15], $E_l(x) = E(x)/t_l$ and they take $\mathcal{X}_i = \{x \in \mathcal{X} : E_{i-1} < E(x) \le E_i\}$ for some predefined valuse $E_0 < E_1 < \cdots < E_d$). Define the function $I(x) = i$ if $x \in \mathcal{X}_i$; so $\mathcal{X}_{I(x)}$ represents the component of the partition to which x belongs. Now set $\omega^{(l)}(x, y) = \mathbf{1}_{\mathcal{X}_{I(x)}}(y)$ and $T^{(l)}$ as in (5) and we get the EE sampler of [15]. In this general case, the limiting kernel has the same form as in (6) but where $R^{(l)}$ is now a Metropolis–Hastings algorithm with target distribution $\pi^{(l)}$ and proposal kernel $Q^{(l)}(x, dy) \propto \pi^{(l-1)}(y)\mathbf{1}_{\mathcal{X}_{I(x)}}(y)\lambda(dy)$. Partitioning the state space and using the proposal $Q^{(l)}(x, dy) \propto \pi^{(l-1)}(y)\mathbf{1}_{\mathcal{X}_{I(x)}}(y)\lambda(dy)$ works well in practice as it can allow large jumps in the state space to be accepted. But it does not add any significant feature to the algorithm from the theoretical standpoint. Therefore and to simplify the analysis, we only consider the case where no partitioning is used ($\mathcal{X}_{I(x)} = \mathcal{X}$ for all $x \in \mathcal{X}$).

3. Asymptotics of the EE sampler. For the remaining of the paper, we restrict our attention to the EE sampler. In other words, we consider the process defined in Section 2 with $\omega^{(l)}(x, y) \equiv 1$ and $T^{(l)}$ as defined in (5).

3.1. *Notation and assumptions.* We start with some notation. If P_1, P_2 are kernels on $(\mathcal{X}, \mathcal{B})$, the product $P_1 P_2$ is the kernel $P_1 P_2(x, A) = \int_{\mathcal{X}} P_1(x, dy) P_2(y, A)$. If μ is a signed measure on $(\mathcal{X}, \mathcal{B})$, we write $\mu(f)$ to denote the integral $\int \mu(dx) f(x)$ and we will also use μ to denote the linear functional on the space of $\mathbb{R}$-valued functions on $(\mathcal{X}, \mathcal{B})$ thus induced. Similarly, we will write $\mu P_1(A)$ for $\int \mu(dx) P_1(x, A)$. Let $V : \mathcal{X} \to [1, \infty)$ be given. For $f : (\mathcal{X}, \mathcal{B}) \to \mathbb{R}$, we define its

V-norm as $|f|_V := \sup_{x\in\mathcal{X}} \frac{|f(x)|}{V(x)}$ and we introduce the space L_V^∞ of measurable real-valued functions defined on $\mathcal{X}$ such that $|f|_V < \infty$. For a signed measure μ on $(\mathcal{X}, \mathcal{B})$ we define by $\|\mu\|_V := \sup\{|\mu(f)|, f \in L_V^\infty, |f|_V \le 1\}$. We equip $\mathcal{M}$, the set of all probability measures on $(\mathcal{X}, \mathcal{B})$, with the metric $\|\mu - \nu\|_V$ and the Borel σ-algebra $\mathcal{B}_{\mathcal{M}}(V)$ induced by $\|\cdot\|_V$. Whenever V is understood, we will write $(\mathcal{M}, \mathcal{B}_{\mathcal{M}})$ instead of $(\mathcal{M}, \mathcal{B}_{\mathcal{M}}(V))$. For a linear operator T from $(L_V^\infty, |\cdot|_V)$ into itself, we define its operator norm by $|||T|||_V := \sup\{|Tf|_V, f \in L_V^\infty, |f|_V \le 1\}$.

We assume that $\pi^{(l)}$ is of the form

$$\pi^{(l)}(dx) = \frac{1}{Z_l} e^{-E(x)/t_l} \lambda(dx) \tag{7}$$

for some continuous function $E : (\mathcal{X}, \mathcal{B}) \to \mathbb{R}$ that is bounded from below and $t_1 > \cdots > t_K = 1$ is a decreasing sequence of positive numbers (temperatures). In addition, we make the following assumption.

ASSUMPTION (A1). *For $l = 1, \ldots, K$, there exist a set $C_l \subset \mathcal{X}$, a probability measure ϕ_l such that $\phi_l(C_l) > 0$ an integer $n_0 > 0$ and constants $\lambda_l \in (0, 1)$, $b_l \in [0, \infty)$, $\varepsilon_l \in (0, 1]$ such that for $x \in \mathcal{X}$ and $A \in \mathcal{B}$,*

$$\left[P^{(l)}\right]^{n_0}(x, A) \ge \varepsilon_l \phi_l(A) \mathbf{1}_{C_l}(x) \tag{8}$$

and

$$P^{(l)} V(x) \le \lambda_l V(x) + b_l \mathbf{1}_{C_l}(x), \tag{9}$$

where $V(x) = ce^{\kappa E(x)} \ge 1$ for some finite constants $c > 0$ and $\kappa \in (0, 1)$ and $0 < \kappa < (\frac{1}{t_l} - \frac{1}{t_{l-1}})$. Moreover

$$\frac{1}{1 + (1 - \lambda_l)(\kappa^{-1}(t_l^{-1} - t_{l-1}^{-1}) - 1)} < \theta_l \le 1, \qquad l = 1, \ldots, K. \tag{10}$$

REMARK 3.1.

(1) The drift and minorization conditions (8)–(9) of Assumption (A1) can be checked for many practical examples. If each $P^{(l)}$ is a Random Walk Metropolis kernel or a Metropolis Adjusted Langevin kernel then (8) and (9) are known to hold under some regularity conditions on the energy function E (see [4, 13]). In these cases, it is always possible to choose κ small enough to satisfy $0 < \kappa < (\frac{1}{t_l} - \frac{1}{t_{l-1}})$.

(2) The condition (10) is a technical condition that quantifies the idea that the rate of resampling $1 - \theta_l$ should not be too large. It is needed to guarantee that the geometric drift condition (9) on $P^{(l)}$ transfers to kernels of the type $P_\mu^{(l)}$ that drive the EE sampler.

3.2. *Law of large numbers.* We consider an arbitrary pair $\{(X_n^{(l-1)}, X_n^{(l)}), n \geq 0\}$. We will show that under Assumption (A1), if $\{X_n^{(l-1)}, n \geq 0\}$ satisfies a strong law of large numbers, then so does $\{X_n^{(l)}, n \geq 0\}$. Then we use the fact that $\{X_n^{(0)}, n \geq 0\}$ is an ergodic Markov chain to derive a law of large numbers for any $\{X_n^{(l)}, n \geq 0\}$.

THEOREM 3.1. *Assume Assumption* (A1) *holds and let* $\beta \in [0, 1)$. *Let* $f : (\mathcal{M}, \mathcal{B}_{\mathcal{M}}) \times (\mathcal{X}, \mathcal{B}) \to \mathbb{R}$ *be a measurable function such that*

$$\sup_{\nu \in \mathcal{M}} |f_\nu|_{V^\beta} < \infty. \tag{11}$$

Suppose that there exists a finite constant C *such that for any* $\nu, \mu \in \mathcal{M}$,

$$|f_\nu - f_\mu|_{V^\beta} \leq C \|\nu - \mu\|_{V^\beta}. \tag{12}$$

Suppose also that for any $h \in L^\infty_{V^\beta}$,

$$\frac{1}{n} \sum_{k=1}^{n} h\bigl(X_k^{(l-1)}\bigr) \longrightarrow \pi^{(l-1)}(h), \qquad \mathbb{P}\text{-a.s. as } n \to \infty, \tag{13}$$

and that there exists $\mathcal{D} \in \mathcal{F}$, $\mathbb{P}(\mathcal{D}) = 1$ *such that for each sample path* $\omega \in \mathcal{D}$, $f_{\mu_n^{(l-1)}}(x)(\omega)$ *converges to* $f_{\pi^{(l-1)}}(x)$ *as* $n \to \infty$ *for all* $x \in \mathcal{X}$. *Then*

$$\frac{1}{n} \sum_{k=1}^{n} f_{\mu_{k-1}^{(l-1)}}\bigl(X_k^{(l)}\bigr) \longrightarrow \pi^{(l)}\bigl(f_{\pi^{(l-1)}}\bigr), \qquad \mathbb{P}\text{-a.s. as } n \to \infty. \tag{14}$$

PROOF. See Section 4.3. $\square$

The following corollary is then immediate.

COROLLARY 3.1. *Assume Assumption* (A1) *holds and suppose that* $\{X_n^{(0)}, n \geq 0\}$ *is a* ϕ*-irreducible aperiodic Markov chain with invariant distribution* $\pi^{(0)}$ *and* $\pi^{(0)}(V) < \infty$. *Let* $f \in L^\infty_{V^\beta}$, $\beta \in [0, 1)$. *Then for any* $l \in \{1, \ldots, K\}$,

$$\frac{1}{n} \sum_{i=1}^{n} f\bigl(X_i^{(l)}\bigr) \longrightarrow \pi^{(l)}(f), \qquad \mathbb{P}\text{-a.s. as } n \to \infty. \tag{15}$$

3.3. *Central limit with a random centering.* We now turn to central limit theorems. It can be shown that the kernel $P_\mu^{(l)}$ admits a unique invariant distribution $\pi_\mu^{(l)}$. Since the conditional distribution of $X_n^{(l)}$ given $\mathcal{F}_{n-1}$ is $P^{(l)}_{\mu_{n-1}^{(l-1)}}$, it is natural to consider a central limit theorem for $\sum_{k=1}^n f(X_k^{(l)})$ in which $f(X_k^{(l)})$ is centered around $\pi^{(l)}_{\mu_{n-1}^{(l-1)}}(f)$. This is done in the next theorem. $\Rightarrow$ denotes weak convergence and $\mathcal{N}(\mu, \sigma^2)$ denotes the Gaussian distribution on $\mathbb{R}$ with mean μ and variance σ^2.

THEOREM 3.2. *Assume Assumption (A1) holds. Let* $f \in L^\infty_{V^\beta}$, $\beta \in [0, 1/2)$ *be such that* $\pi^{(l)}(f) = 0$. *Define*

$$\sigma_l^2(f) := \pi^{(l)}(f^2) + 2\sum_{k=1}^{\infty} \int_{\mathcal{X}} \pi^{(l)}(dx) f(x) \big[K^{(l)}\big]^k f(x), \tag{16}$$

where $K^{(l)}$ *is given by* (6). *Assume that* $\sigma_l^2(f) > 0$. *Then there exists a random sequence* $\{\pi_n^{(l)}(f)\}$, $\pi_n^{(l)}(f) \to \pi^{(l)}(f)$ (*almost surely*) *as* $n \to \infty$ *such that*

$$\frac{1}{\sqrt{n}\sigma_l(f)} \sum_{k=1}^{n} \big[f(X_k^{(l)}) - \pi_k^{(l)}(f)\big] \Rightarrow \mathcal{N}(0, 1) \qquad \text{as } n \to \infty. \tag{17}$$

PROOF. See Section 4.4. $\square$

3.4. *Central limit theorem with a deterministic centering*. We now derive a central limit theorem for $\sum_{k=1}^n f(X_k^{(l)})$ around $\pi^{(l)}(f)$ which gives more insight in the efficiency of the method as a Monte Carlo sampler from $\pi^{(l)}$. We restrict ourselves to the case where $l = 1$; that is, we only consider the pair $\{(X_n^{(0)}, X_n^{(1)}), n \geq 0\}$. Moreover, we assume in this section that $\mathcal{X}$ is a compact subset of $\mathbb{R}^d$ (equipped with its Euclidean metric). More precisely:

ASSUMPTION (A1′). $\mathcal{X}$ *is a compact subset of* $\mathbb{R}^d$. *For* $l = 0, 1$, *there exist an integer* $n_0 > 0$, *a constant* $\varepsilon_l \in (0, 1]$ *a probability measure* ϕ_l *such that for* $x \in \mathcal{X}$ *and* $A \in \mathcal{B}$,

$$\big[P^{(l)}\big]^{n_0}(x, A) \geq \varepsilon_l \phi_l(A). \tag{18}$$

Let $\mathcal{C}(\mathcal{X}, \mathbb{R})$ be the space of all continuous functions from $\mathcal{X} \to \mathbb{R}$. We endowed $\mathcal{C}(\mathcal{X}, \mathbb{R})$ with the uniform metric $|f|_\infty := \sup_{x \in \mathcal{X}} |f(x)|$ and its Borel σ-algebra. Let $\mathrm{Lip}(\mathcal{X}, \mathbb{R})$ be the subset of Lipschitz functions of $\mathcal{C}(\mathcal{X}, \mathbb{R})$ [we say that $f : \mathcal{X} \to \mathbb{R}$ is Lipschitz if there exists $C < \infty$ such that for any $x, y \in \mathcal{X}$, $|f(x) - f(y)| \leq C|x - y|$].

For $f : \mathcal{X} \to \mathbb{R}$ bounded measurable, define the function

$$U(x) = U_f(x) := \sum_{j \geq 0} \big(P_{\pi^{(0)}}^{(1)}\big)^j f(x),$$

the solution to the Poisson equation for f and $P_{\pi^{(0)}}^{(1)}$. To simplify the notations, we omit the dependence of U on f. Notice that $P_{\pi^{(0)}}^{(1)}$ is the limiting kernel in the EE sampler, denoted $K^{(1)}$ in (6). Clearly, Assumption (A1′) implies as shown in Lemma 4.1 below that the kernel $P_\mu^{(1)}$ is also uniformly ergodic, uniformly in μ.

In particular $|U|_\infty < \infty$. We assume that the function U is Lipschitz whenever f is Lipschitz:

$$f \in \operatorname{Lip}(\mathcal{X}, \mathbb{R}) \qquad \text{implies that} \quad \sum_{j\ge 0} \big(P^{(1)}_{\pi^{(0)}}\big)^j f \in \operatorname{Lip}(\mathcal{X}, \mathbb{R}). \tag{19}$$

We comment on (19) below. Let $f \in \mathcal{C}(\mathcal{X}, \mathbb{R})$ such that $\pi^{(1)}(f) = 0$. Consider the partial sum $S_n = \sum_{k=1}^n f(X_k^{(1)})$. Since U satisfies the Poisson equation $U - P^{(1)}_{\pi^{(0)}} U = f$, we can rewrite S_n as

$$\begin{aligned} S_n &= \sum_{k=1}^n U\big(X_k^{(1)}\big) - P^{(1)}_{\pi^{(0)}} U\big(X_k^{(1)}\big) \\ &= M_n + \sum_{k=1}^n P^{(1)}_{\mu_k^{(0)}} U\big(X_k^{(1)}\big) - P^{(1)}_{\pi^{(0)}} U\big(X_k^{(1)}\big) + \varepsilon_n^{(1)}, \end{aligned}$$

where $M_n = \sum_{k=1}^n U(X_k^{(1)}) - P^{(1)}_{\mu_{k-1}^{(0)}} U(X_{k-1}^{(1)})$ is a martingale and $\varepsilon_n^{(1)} = P^{(1)}_{\mu_0^{(0)}} \times U(X_0^{(1)}) - P^{(1)}_{\mu_n^{(0)}} U(X_n^{(1)})$.

We introduce the function

$$\begin{aligned} H_x(y) &:= T^{(1)}(y, x, U) - R^{(1)}(x, U) \\ &= \int T^{(1)}(y, x, dz) U(z) - \int \pi^{(0)}(dy) \int T^{(1)}(y, x, dz) U(z). \end{aligned} \tag{20}$$

Since $P^{(1)}_\mu(x, dz) = \theta_1 P^{(1)}(x, dz) + (1-\theta_1) \int \mu(dy) \int T^{(1)}(y, x, dz)$, we have

$$P^{(1)}_\mu U(x) - P^{(1)}_{\pi^{(0)}} U(x) = (1-\theta_1) \int \mu(dy) H_x(y),$$

so that we can rewrite S_n as

$$\begin{aligned} S_n &= M_n + (1-\theta_1) \sum_{k=1}^n \frac{1}{k} \sum_{j=1}^k H_{X_k^{(1)}}\big(X_j^{(0)}\big) + \varepsilon_n^{(1)} \\ &= M_n + (1-\theta_1) \sum_{k=1}^n \frac{1}{\sqrt{k}} \eta_k\big(X_k^{(1)}\big) + \varepsilon_n^{(1)}, \end{aligned}$$

where η_n is the random field

$$\eta_n(x) := n^{-1/2} \sum_{k=1}^n H_x\big(X_k^{(0)}\big).$$

We will see that η_n is a $\mathcal{C}(\mathcal{X}, \mathbb{R})$-valued random element. To describe its asymptotic behavior we introduce the function

$$U_x^{(0)}(y) = \sum_{j\ge 0} \big[P^{(0)}\big]^j H_x(y),$$

where for a kernel Q, $QH_x(y) = \int Q(y, dz)H_x(z)$ and the covariance function

$$\Gamma(x, y) = \int [U_x^{(0)}(z)U_y^{(0)}(z) - (P^{(0)}U_x^{(0)}(z))(P^{(0)}U_y^{(0)}(z))]\pi^{(0)}(dz). \tag{21}$$

If $f, g \in \mathcal{C}(\mathcal{X}, \mathbb{R})$, with an abuse of notation we will also write $\Gamma(f, g)$ for the quantity

$$\Gamma(f, g) = \int [U_f^{(0)}(z)U_g^{(0)}(z) - (P^{(0)}U_f^{(0)}(z))(P^{(0)}U_g^{(0)}(z))]\pi^{(0)}(dz),$$

where $U_f^{(0)}(x) = \sum_{j\geq 0}[P^{(0)}]^j f(x)$.

THEOREM 3.3. *Assume Assumption* (A1′) *and* (19) *hold and suppose that* $E \in \mathrm{Lip}(\mathcal{X}, \mathbb{R})$. *Let* $f \in \mathrm{Lip}(\mathcal{X}, \mathbb{R})$ *such that* $\pi^{(1)}(f) = 0$. *Then*

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n} f(X_k^{(1)}) \Rightarrow \mathcal{N}(0, \sigma_\star^2(f) + 4(1-\theta_1)^2\Gamma(\bar{g}, \bar{g})) \qquad \textit{as } n \to \infty, \tag{22}$$

where $\bar{g}(\cdot) := \int \pi^{(1)}(dx)H_x(\cdot)$ *and*

$$\sigma_\star^2(f) := \pi^{(1)}(f^2) + 2\sum_{k=1}^{\infty}\int_{\mathcal{X}} \pi^{(1)}(dx) f(x)(P_{\pi^{(0)}}^{(1)})^k f(x). \tag{23}$$

PROOF. See Section 4.5. □

Notice from (20) that $\bar{g}(\cdot) = \int \pi^{(1)}(dx)T^{(1)}(\cdot, x, U) - \int \pi^{(0)}(dz)\int \pi^{(1)}(dx) \times T^{(1)}(z, x, U)$. Thus Theorem 3.3 shows that the asymptotic variance of the EE sampler is the sum of the asymptotic variance in estimating $\pi^{(1)}(f)$ as if the limiting kernel $P_{\pi^{(0)}}^{(1)}$ is known [the term $\sigma_\star^2(f)$] plus the asymptotic in using the chain $\{X_n^{(0)}, n \geq 0\}$ to estimate the expectation under $\pi^{(0)}$ of the function $\int \pi^{(1)}(dx)T^{(1)}(\cdot, x, U)$. In their analysis [8] arrive at a similar CLT for interacting MCMC algorithms. Notice also that $U(x) = \sum_{j\geq 0}(P_{\pi^{(0)}}^{(1)})^j f(x)$. Thus in most cases, the function $\int \pi^{(1)}(dx)T^{(1)}(\cdot, x, U)$ will typically take large values and the asymptotic variance in estimating its expectation will also tend to be large particularly if the kernel $P^{(0)}$ mixes poorly. Theorem 3.3 thus suggests that for the EE sampler to be effective in practice it is important that the initial chain $\{X_n^{(0)}, n \geq 0\}$ enjoys a very fast mixing.

A remaining question is to know whether $n^{-1}\mathbb{E}[(\sum_{k=1}^n f(X_k^{(1)}))^2]$ converges to $\sigma_\star^2(f) + 4(1-\theta_1)^2\Gamma(\bar{g}, \bar{g})$. Unfortunately the answer is no in general as shown by the following example:

PROPOSITION 3.1. *Assume Assumption* (A1′) *holds. Suppose that* $P^{(0)} = P^{(1)} = P$ *and* $\pi^{(0)} = \pi^{(1)} = \pi$. *Let* $f : \mathcal{X} \to \mathbb{R}$ *be a bounded measurable function such that* $\pi(f) = 0$. *Then*

$$\lim_{n\to\infty} n^{-1}\mathbb{E}\left[\left(\sum_{k=1}^{n} f(X_k^{(1)})\right)^2\right] = \sigma_\star^2(f) + 2(1-\theta_1)^2\Gamma(\bar{g},\bar{g}).$$

In the present case $\bar{g}(x) = U(x) = \sum_{j\geq 0}\theta_1^j P^j f(x)$ *and*

$$\sigma_\star^2(f) = \pi(|f|^2) + 2\sum_{k=1}^{\infty}\theta_1^k \int \pi(dx) f(x) P^k f(x).$$

PROOF. See Section 4.6. □

REMARK 3.2. Assumption (19) can often be easily checked. Indeed, we have $U(x) = f(x) + P^{(1)}_{\pi^{(0)}} U(x)$, where $P^{(1)}_{\pi^{(0)}} = \theta_1 P^{(1)} + (1-\theta_1)R^{(1)}$, where $R^{(1)}$ is the independent Metropolis–Hastings algorithm with target $\pi^{(1)}$ and proposal $\pi^{(0)}$. Let us assume that $P^{(1)}$ is also a Metropolis–Hastings kernel with target $\pi^{(1)}$ and proposal $q(x,y)$. Denote $\alpha(x,y)$ [resp. $\bar{\alpha}(x,y)$] the acceptance probability of $P^{(1)}$ [resp. $R^{(1)}$], and denote $a(x) := \int \alpha(x,y)q(x,y)\,dy$ [resp. $a(x) := \int \alpha(x,y)\pi^{(0)}(y)\,dy$] the average acceptance probability at x for $P^{(1)}$ [resp. for $R^{(1)}$]. Then we have

$$\begin{aligned} &U(x)\big(1-\theta_1\big(1-a(x)\big) - (1-\theta_1)\big(1-\bar{a}(x)\big)\big) \\ &\qquad = f(x) + \theta_1 \int \alpha(x,y)q(x,y)U(y)\,dy + (1-\theta_1)\int \bar{\alpha}(x,y)\pi^{(0)}(y)U(y)\,dy. \end{aligned}$$

Thus if $\pi^{(0)}, \pi^{(1)}$ and q such that a and $\bar{a}$ remains bounded away from 0 and the integral operators $h \to \int \alpha(x,y)q(x,y)h(y)\,dy$ and $h \to \int \bar{\alpha}(x,y)\pi^{(0)}(y)h(y)\,dy$ transform bounded measurable functions into Lipschitz functions, then (19) hold. For example, if $\pi^{(0)}, \pi^{(1)}$ and q are all positive on $\mathcal{X}$ and of class $\mathcal{C}^1$ then (19) hold.

REMARK 3.3. The result developed above relies heavily on the Lipschitz continuity assumption. Under that assumption, we show that the stochastic process $\{\eta_n, n \geq 0\}$ lives in the Polish space $\mathcal{C}(\mathcal{X},\mathbb{R})$ which allows us to use the standard machinery of weak convergence in Polish spaces. If f is only assumed measurable the theorem above no longer hold. But a similar result can still be obtained using weak convergence techniques in nonseparable metric spaces. But we do not pursue this here.

3.5. *An illustrative example.* Consider the following example. Suppose that we want to sample from the bivariate normal distribution $\mathcal{N}(0, \Sigma)$, with covariance matrix

$$\Sigma = \begin{bmatrix} 0.96 & 2.44 \\ 2.44 & 7.04 \end{bmatrix}.$$

For this problem, we compare a Random Walk Metropolis (RWM) algorithm, the EE sampler, the MCMC algorithm based on the limiting kernel of EE sampler (call it limit EE sampler), IR-MCMC and the MCMC algorithm based on the limiting kernel of IR-MCMC (limit IR-MCMC sampler).

For the RWM sampler, the proposal kernel is $\mathcal{N}(0, I_2)$, where I_2 is the 2-dimensional identity matrix. For the adaptive chains, we use four chains with $\pi^{(0)} = \pi^{1/10}$, $\pi^{(1)} = \pi^{1/5}$, $\pi^{(2)} = \pi^{1/2}$ and $\pi^{(3)} = \pi$. We take $\theta_l = \theta = 0.5$ and $P^{(l)}$ is taken to be a RWM algorithm with target $\pi^{(l)}$ and proposal $\mathcal{N}(0, I_2)$. It can be checked that Assumption (A1) holds for this problem. We simulate each of the five samplers for $N = 10{,}000$ iterations. We compare the samplers on their mean square errors (MSE) in estimating the first two moments of the two components of the distribution π. We calculate the MSEs by repeating the simulations 100 times. The results are reported in Table 1.

From these results we see (as expected) that the limit EE sampler is 3 to 25 times more efficient than the RWM sampler, and the limit IR-MCMC sampler is 15 to 50 more efficient than the RWM sampler. But IR-MCMC itself is hardly more efficient than the RWM sampler. If we take the computation times into account, it becomes hard to make the case that any of these adaptive sampler is better than the plain RWM. Similar conclusions can be drawn for the EE sampler.

TABLE 1
Mean square error and ratios (with respect to the RWM sampler) for IR-MCMC, limit IR-MCMC, EE and limit EE. Based on 100 replications of 10,000 iterations of each sampler

		$\mathbb{E}(\mathbf{X_1})$	$\mathbb{E}(\mathbf{X_2})$	$\mathbb{E}(\mathbf{X_1^2})$	$\mathbb{E}(\mathbf{X_2^2})$
RWM	MSE	0.0099	0.0803	0.0091	0.5525
	Ratios	1.0	1.0	1.0	1.0
IR-MCMC	MSE	0.0098	0.0774	0.0047	0.2962
	Ratios	1.00	1.04	1.95	1.87
Limit IR-MCMC	MSE	0.0002	0.0017	0.0006	0.0296
	Ratios	48.43	46.20	14.18	18.66
EE	MSE	0.0057	0.0435	0.0045	0.2810
	Ratios	1.74	1.84	2.02	1.97
Limit EE	MSE	0.0004	0.0030	0.0034	0.1966
	Ratios	25.99	26.36	2.67	2.81

4. Proofs.

4.1. *Preliminary results on kernels of the form* $P_\nu^{(l)}$. For a probability measure ν and $l = 1, \dots, K$, let $P_\nu^{(l)}$ as in (2) with $\omega^{(l)} \equiv 1$ and $T^{(l)}$ as in (5). The following lemma shows that $P_\nu^{(l)}$ satisfies a drift and a minorization conditions with constant that actually do not depend on ν.

LEMMA 4.1. *Assume Assumption (A1) holds. Then there exists $\lambda_l' \in (0, 1)$ that does not depend on ν such that for $x \in \mathcal{X}$ and $A \in \mathcal{B}$:*

$$[P_\nu^{(l)}]^{n_0}(x, A) \geq \theta_l \varepsilon_l \phi_l(A) \mathbf{1}_{C_l}(x) \tag{24}$$

and

$$P_\nu^{(l)} V(x) \leq \lambda_l' V(x) + b_l \mathbf{1}_{C_l}(x), \tag{25}$$

where C_l, ϕ_l, b_l, ε_l and V are as in Assumption (A1).

PROOF. We have $P_\nu^{(l)} \geq \theta_l P^{(l)}$. Therefore (24) follows from the minorization condition (8).

Define $\delta_l = (\kappa^{-1}(t_l^{-1} - t_{l-1}^{-1}) - 1)^{-1}$. We will show that

$$\int \nu(dy) T^{(l)}(y, x, V) \leq (1 + \delta_l) V(x). \tag{26}$$

Given the drift condition (9), this will imply

$$\begin{aligned} P_\nu^{(l)} V(x) &\leq \bigl(\theta_l \lambda + (1 - \theta_l)(1 + \delta_l)\bigr) V(x) + b_l \mathbf{1}_{C_l}(x) \\ &\leq \lambda_l' V(x) + b_l \mathbf{1}_{C_l}(x), \end{aligned}$$

where $\lambda_l' = \theta_l \lambda + (1 - \theta_l)(1 + \delta_l) \in (0, 1)$ by the condition on κ in Assumption (A1).

Observe that $r^{(l)}(x) = e^{-E(x)(t_l^{-1} - t_{l-1}^{-1})}$, $t_l^{-1} - t_{l-1}^{-1} > 0$ and $V(x) = ce^{\kappa E(x)} \geq 1$, $\kappa \in (0, 1)$. This implies that $r^{(l)}(y)/r^{(l)}(x) \geq 1$ if and only if $E(y) \leq E(x)$. Denote $\mathcal{A}(x) = \{y \in \mathcal{X} : E(y) \leq E(x)\}$ and $\mathcal{R}(x) = \{y \in \mathcal{X} : E(y) > E(x)\}$. Then we have

$$\begin{aligned} &\int \nu(dy) T^{(l)}(y, x, V) \\ &\qquad = \int_{\mathcal{A}(x)} \nu(dy) T^{(l)}(y, x, V) + \int_{\mathcal{R}(x)} \nu(dy) T^{(l)}(y, x, V) \\ &\qquad = \int_{\mathcal{A}(x)} \nu(dy) V(y) + \int_{\mathcal{R}(x)} \nu(dy) \frac{r^{(l)}(y)}{r^{(l)}(x)} V(y) \\ &\qquad\quad + V(x) \int_{\mathcal{R}(x)} \nu(dy) \biggl(1 - \frac{r^{(l)}(y)}{r^{(l)}(x)}\biggr), \end{aligned}$$

$$
\begin{aligned}
&= \int_{\mathcal{A}(x)} \nu(dy) V(y) + V(x) \int_{\mathcal{R}(x)} \nu(dy) \\
&\quad + \int_{\mathcal{R}(x)} \nu(dy) \frac{r^{(l)}(y)}{r^{(l)}(x)} \bigl(V(y) - V(x)\bigr) \\
&\le V(x) + V(x) \int_{\mathcal{R}(x)} \nu(dy) \frac{r^{(l)}(y)}{r^{(l)}(x)} \biggl(\frac{V(y)}{V(x)} - 1\biggr) \\
&= V(x) \biggl[1 + \int_{\mathcal{R}(x)} e^{-(E(y)-E(x))(1/t_l - 1/t_{l-1})} \\
&\qquad\qquad \times \bigl(e^{\kappa(E(y)-E(x))} - 1\bigr) \nu(dy)\biggr] \\
&\le V(x) \frac{\kappa}{1/t_l - 1/t_{l-1} - \kappa}.
\end{aligned}
$$

In the last line we use the following inequality: for $0 < x < y$: $e^{-y}(e^x - 1) \le x/(y - x)$. $\square$

From Lemma 4.1, we deduce that for any probability measure ν, $P_\nu^{(l)}$ has an invariant distribution $\pi_\nu^{(l)}$ such that

$$\pi_\nu^{(l)}(V) \le b_l. \tag{27}$$

See [17], Theorems 15.0.1 and 14.3.7. The lemma also implies that for any $\beta \in (0, 1]$, there exist constants $C_\beta < \infty$ and $\rho_\beta \in (0, 1)$ that does not depend on ν such that

$$\bigl\| [P_\nu^{(l)}]^k(x, \cdot) - \pi_\nu^{(l)}(\cdot) \bigr\|_{V^\beta} \le C_\beta \rho_\beta^k V^\beta(x), \qquad k \ge 0, x \in \mathcal{X}. \tag{28}$$

See, for example, [7] for a proof. The following lemma holds.

LEMMA 4.2. *Fix $\beta \in [0, 1]$ and μ and ν two probability measures on $(\mathcal{X}, \mathcal{B})$*

$$\vert\kern-0.25ex\vert\kern-0.25ex\vert P_\mu^{(l)} - P_\nu^{(l)} \vert\kern-0.25ex\vert\kern-0.25ex\vert_{V^\beta} \le 2 \|\mu - \nu\|_{V^\beta}. \tag{29}$$

PROOF. For $f \in L_{V^\beta}^\infty$ such that $|f|_{V^\beta} \le 1$, we have

$$P_\mu^{(l)} f(x) - P_\nu^{(l)} f(x) = (1 - \theta_l) \int T^{(l)}(y, x, f) \bigl(\mu(dy) - \nu(dy)\bigr),$$

where $T^{(l)}(y, x, f) = \min(1, \frac{r_l(y)}{r_l(x)})(f(y) - f(x)) + f(x)$. Therefore

$$
\begin{aligned}
&\frac{P_\mu^{(l)} f(x) - P_\nu^{(l)} f(x)}{(1 - \theta_l) V^\beta(x)} \\
&\qquad = \int \frac{\min(1, r^{(l)}(y)/r^{(l)}(x))(f(y) - f(x))}{V^\beta(x) V^\beta(y)} V^\beta(y) \bigl(\mu(dy) - \nu(dy)\bigr).
\end{aligned}
$$

Now for $|f|_{V^\beta} \le 1$, $|\frac{\min(1, r^{(l)}(\cdot)/r^{(l)}(x))(f(\cdot)-f(x))}{V^\beta(x)V^\beta(\cdot)} V^\beta(\cdot)|_{V^\beta} \le 2$ for all $x \in \mathcal{X}$. Therefore

$$\begin{aligned}
&\left|\int \frac{\min(1, r^{(l)}(y)/r^{(l)}(x))(f(y)-f(x))}{V^\beta(x)V^\beta(y)} V^\beta(y)\big(\mu(dy)-\nu(dy)\big)\right| \\
&\quad \le 2 \sup_{|f|_{V^\beta}\le 1} \left|\int f(y)\big(\mu(dy)-\nu(dy)\big)\right| \\
&\quad = 2\|\mu-\nu\|_{V^\beta}. \qquad \square
\end{aligned}$$

For $l \in \{1, \dots, K\}$, define the kernel

$$N_\mu^{(l)} f(x) = \int \mu(dy) f(y) \min\left(1, \frac{r^{(l)}(y)}{r^{(l)}(x)}\right), \qquad x \in \mathcal{X}.$$

LEMMA 4.3. *Let μ be a probability measure on $(\mathcal{X}, \mathcal{B})$. For $x_1, x_2 \in \mathcal{X}$, and $f \in L^\infty_{V^\beta}$, $\beta \in [0,1]$*

$$\begin{aligned}
&\big|N_\mu^{(l)} f(x_1) - N_\mu^{(l)} f(x_2)\big| \\
&\quad \le |f|_{V^\beta} \big|e^{\tau E(x_1)} - e^{\tau E(x_2)}\big| \left|\int \mu(dy) e^{-(\tau-\kappa\beta)E(y)}\right|
\end{aligned} \tag{30}$$

with $\tau = 1/t_l - 1/t_{l-1}$ and κ as in Assumption (A1).

PROOF. Fix x_1 and x_2 and define $\Delta(y) = V^\beta(y)|\min(1, \frac{r^{(l)}(y)}{r^{(l)}(x_1)}) - \min(1, \frac{r^{(l)}(y)}{r^{(l)}(x_2)})|$. On $r^{(l)}(y) \ge \max(r^{(l)}(x_1), r^{(l)}(x_2))$, $\Delta(y) = 0$. On $r^{(l)}(x_1) \le r^{(l)}(y) \le r^{(l)}(x_2)$,

$$\begin{aligned}
\Delta(y) &= V^\beta(y)\left(1 - \frac{r^{(l)}(y)}{r^{(l)}(x_2)}\right) \\
&= e^{\kappa\beta E(y)}\big(1 - e^{-\tau(E(y)-E(x_2))}\big) \\
&= e^{-(\tau-\kappa\beta)E(y)}\big(e^{\tau E(y)} - e^{\tau E(x_2)}\big) \\
&\le \big(e^{\tau(E(x_1)} - e^{\tau E(x_2))}\big) e^{-(\tau-\kappa\beta)E(y)}.
\end{aligned}$$

Similarly, on $r^{(l)}(y) \le \min(r^{(l)}(x_1), r^{(l)}(x_2))$,

$$\begin{aligned}
\Delta(y) &\le \big|e^{\tau E(x_1)} - e^{\tau E(x_2)}\big| V^\beta(y) r^{(l)}(y) \\
&= \big|e^{\tau E(x_1)} - e^{\tau E(x_2)}\big| e^{-(\tau-\kappa\beta)E(y)}.
\end{aligned}$$

Putting the three parts together yields the lemma. $\square$

REMARK 4.1. Lemma 4.3 will be useful in deriving a uniform law of large numbers for $\{X_n^{(l)}\}$. Actually, this lemma shows that if the function E is continuous then the kernel $N_\mu^{(l)}$ is a strong Feller kernel that transforms a bounded function f into a continuous bounded function $N_\mu^{(l)}$ (uniformly in μ). We will use this later.

4.2. *Poisson equation.* A straightforward consequence of Section 4.1 is that for any $f \in L_{V^\beta}^\infty$, $\beta \in (0, 1]$ the function

$$U_\nu^{(l)} f(x) := \sum_{k=0}^{\infty} [P_\nu^{(l)} - \pi_\nu^{(l)}]^k f(x) \tag{31}$$

is well defined and

$$|U_\nu^{(l)} f|_{V^\beta} + |P_\nu^{(l)} U_\nu^{(l)} f|_{V^\beta} \le C |f|_{V^\beta}, \tag{32}$$

where C is finite and does not depend on ν nor f. $U_\nu^{(l)} f$ satisfies the (Poisson) equation

$$U_\nu^{(l)} f(x) - P_\nu U_\nu^{(l)} f(x) = f(x) - \pi_\nu^{(l)}(f), \qquad x \in \mathcal{X}. \tag{33}$$

Lemmas 4.1 and 4.2 implie that for all $\beta \in (0, 1]$, and μ, ν probability measures on $(\mathcal{X}, \mathcal{B})$:

$$\|\pi_\mu^{(l)} - \pi_\nu^{(l)}\|_{V^\beta} \le C \|\mu - \nu\|_{V^\beta}; \tag{34}$$

for $f \in L_{V^\beta}^\infty$,

$$|U_\mu^{(l)} f - U_\nu^{(l)} f|_{V^\beta} \le C |f|_{V^\beta} \|\mu - \nu\|_{V^\beta} \tag{35}$$

and

$$|P_\mu^{(l)} U_\mu^{(l)} f - P_\nu^{(l)} U_\nu^{(l)} f|_{V^\beta} \le C |f|_{V^\beta} \|\mu - \nu\|_{V^\beta}. \tag{36}$$

The inequalities (34), (35) and (36) can be derived, for example, by adapting the proofs of Proposition 3 of [2]. We omit the details. An important point is the fact that the constant C (whose actual value can change from one equation to the other) does not depend on f nor ν, μ.

4.3. *Proof of Theorem* 3.1. Let $f : (\mathcal{M}, \mathcal{B}_\mathcal{M}) \times (\mathcal{X}, \mathcal{B}) \to \mathbb{R}$ be a measurable function. We will use the notation $f_\mu(x)$ when evaluating f. We introduce the partial sum associated to $\{X_n^{(l)}, n \ge 0\}$:

$$\begin{aligned} S_n^{(l)}(f) &:= \sum_{k=1}^{n} f_{\mu_{k-1}^{(l-1)}}(X_k^{(l)}) \\ &= \sum_{k=1}^{n} \pi_{\mu_{k-1}^{(l-1)}}^{(l)}(f_{\mu_{k-1}^{(l-1)}}) \\ &\quad + \sum_{k=1}^{n} (f_{\mu_{k-1}^{(l-1)}}(X_k^{(l)}) - \pi_{\mu_{k-1}^{(l-1)}}^{(l)}(f_{\mu_{k-1}^{(l-1)}})). \end{aligned}$$

Using the Poisson equation (33), we have the decomposition

$$
\begin{aligned}
S_n^{(l)}(f) &= \sum_{k=1}^{n} \pi^{(l)}_{\mu^{(l-1)}_{n-1}}\big(f_{\mu^{(l-1)}_{k-1}}\big) + M_n^{(l)}(f) + R_{n,1}^{(l)}(f) + R_{n,2}^{(l)}(f), \\
M_n^{(l)}(f) &= \sum_{k=1}^{n} D_k^{(l)}(f),
\end{aligned}
\tag{37}
$$

where

$$
\begin{aligned}
D_k^{(l)}(f) &= U^{(l)}_{\mu^{(l-1)}_{k-1}} f_{\mu^{(l-1)}_{k-1}}\big(X_k^{(l)}\big) - P^{(l)}_{\mu^{(l-1)}_{k-1}} U^{(l)}_{\mu^{(l-1)}_{k-1}} f_{\mu^{(l-1)}_{k-1}}\big(X_{k-1}^{(l)}\big), \\
R_{n,1}^{(l)}(f) &= P^{(l)} U_0^{(l)} f_0\big(X_0^{(l)}\big) - P^{(l)}_{\mu^{(l-1)}_{n}} U^{(l)}_{\mu^{(l-1)}_{n}} f_{\mu^{(l-1)}_{n}}\big(X_n^{(l)}\big)
\end{aligned}
$$

and

$$
R_{n,2}^{(l)}(f) = \sum_{k=1}^{n} P^{(l)}_{\mu^{(l-1)}_{k}} U^{(l)}_{\mu^{(l-1)}_{k}} f_{\mu^{(l-1)}_{k}}\big(X_k^{(l)}\big) - P^{(l)}_{\mu^{(l-1)}_{k-1}} U^{(l)}_{\mu^{(l-1)}_{k-1}} f_{\mu^{(l-1)}_{k-1}}\big(X_k^{(l)}\big).
$$

LEMMA 4.4.

$$
\sup_{1\le l\le K} \sup_{k,k'\ge 0} \mathbb{E}\big(V\big(X_{k'}^{(l-1)}\big)V\big(X_k^{(l)}\big)\big) < \infty.
$$

PROOF. This is a straightforward consequence of the (uniform in ν) drift condition on $P_\nu^{(l)}$. $\square$

LEMMA 4.5. *Let $p > 1$ such that $p\beta \le 1$. There exists a finite constant C such that*

$$
\mathbb{E}\big[\big|R_{n,2}^{(l)}(f)\big|^p\big] \le C(\log n)^p.
$$

Moreover $n^{-1}R_{n,2}^{(l)}(f)$ converges $\mathbb{P}$-almost surely to 0.

PROOF. We use (36), (32) and (11) to obtain

$$
\begin{aligned}
&\big|P^{(l)}_{\mu^{(l-1)}_{k}} U^{(l)}_{\mu^{(l-1)}_{k}} f_{\mu^{(l-1)}_{k}}\big(X_k^{(l)}\big) - P^{(l)}_{\mu^{(l-1)}_{k-1}} U^{(l)}_{\mu^{(l-1)}_{k-1}} f_{\mu^{(l-1)}_{k-1}}\big(X_k^{(l)}\big)\big|^p \\
&\qquad \le C \sup_{\nu\in\mathcal{M}} |f_\nu|^p_{V^\beta} \big\|\mu_n^{(l-1)} - \mu_{n-1}^{(l-1)}\big\|^p_{V^\beta} V^{\beta p}\big(X_k^{(l)}\big).
\end{aligned}
\tag{38}
$$

But $\mu_n^{(l-1)} = \mu_{n-1}^{(l-1)} + n^{-1}(\delta_{X_n^{(l-1)}} - \mu_{n-1}^{(l-1)})$ and we get

$$
\begin{aligned}
\big\|\mu_n^{(l-1)} - \mu_{n-1}^{(l-1)}\big\|_{V^\beta} &= \sup_{|f|_{V^\beta}\le 1} \big|\big(\mu_n^{(l-1)} - \mu_{n-1}^{(l-1)}\big)(f)\big| \\
&\le \frac{1}{n+1}\left(V^\beta\big(X_n^{(l-1)}\big) + \frac{1}{n}\sum_{k=0}^{n-1} V^\beta\big(X_k^{(l-1)}\big)\right).
\end{aligned}
$$

In view of Lemma 4.4 and since $p\beta \le 1$, $\mathbb{E}[V^{p\beta}(X_k^{(l)})(V^{\beta}(X_n^{(l-1)}) + \frac{1}{n} \times \sum_{k=0}^{n-1} V^{\beta}(X_k^{(l-1)}))^p] \le C$ for some finite constant C that does not depend on n. Therefore, given (38) and (11), we can use Minkowski's inequality to conclude the first part of the lemma.

For the second part, by Kronecker's lemma, it is enough to show that the series

$$\sum_{k\ge 1} k^{-1}\big(P^{(l)}_{\mu_k^{(l-1)}} U^{(l)}_{\mu_k^{(l-1)}} f_{\mu_k^{(l-1)}}(X_k^{(l)}) - P^{(l)}_{\mu_{k-1}^{(l-1)}} U^{(l)}_{\mu_{k-1}^{(l-1)}} f_{\mu_{k-1}^{(l-1)}}(X_k^{(l)})\big)$$

converges almost surely. This will follow if we show that

$$\sum_{k\ge 1} k^{-1}\mathbb{E}\big(\big|P^{(l)}_{\mu_k^{(l-1)}} U^{(l)}_{\mu_k^{(l-1)}} f_{\mu_k^{(l-1)}}(X_k^{(l)}) - P^{(l)}_{\mu_{k-1}^{(l-1)}} U^{(l)}_{\mu_{k-1}^{(l-1)}} f_{\mu_{k-1}^{(l-1)}}(X_k^{(l)})\big|\big)$$

is finite. But from the above calculations, we have seen that

$$\mathbb{E}\big(\big|P^{(l)}_{\mu_k^{(l-1)}} U^{(l)}_{\mu_k^{(l-1)}} f_{\mu_k^{(l-1)}}(X_k^{(l)}) - P^{(l)}_{\mu_{k-1}^{(l-1)}} U^{(l)}_{\mu_{k-1}^{(l-1)}} f_{\mu_{k-1}^{(l-1)}}(X_k^{(l)})\big|\big) \le Ck^{-1}.$$

The lemma thus follows. $\square$

LEMMA 4.6. *Let $p > 1$ such that $\beta p \le 1$. Then*

$$\sup_n \mathbb{E}\big[\big|R_{n,1}^{(l)}(f)\big|^p\big] < \infty.$$

Moreover for any $\delta > 0$,

$$\Pr\Big[\sup_{m\ge n} \big|m^{-1}R_{m,1}^{(l)}(f)\big| > \delta\Big] \to 0 \qquad \text{as } n \to \infty.$$

PROOF. The first part is a direct consequence of (11) and (32). For the second part, by Markov's inequality, we see that

$$\begin{aligned}\Pr\Big[\sup_{m\ge n} \big|m^{-1}R_{m,1}^{(l)}(f)\big| > \delta\Big] &\le \delta^{-p}\mathbb{E}\Big[\sum_{m\ge n} m^{-p}\big|R_{m,1}^{(l)}(f)\big|^p\Big]\\ &\le C\delta^{-p}\sum_{m\ge n} m^{-p} \to 0 \qquad \text{as } n\to\infty.\end{aligned}$$

$\square$

LEMMA 4.7. *Let $p > 1$ such that $p\beta \le 1$. There exists a finite constant C such that*

$$\mathbb{E}\big[\big|M_n^{(l)}(f)\big|^p\big] \le Cn^{\max(1,p/2)}.$$

PROOF. By Burkeholder's inequality applied to the martingale $\{M_n^{(l)}(f)\}$, we get

$$\mathbb{E}\big[\big|M_n^{(l)}(f)\big|^p\big] \le C\mathbb{E}\Bigg[\bigg(\sum_{k=1}^{n} \big|D_{k-1}^{(l)}(f)\big|^2\bigg)^{p/2}\Bigg].$$

If $p \geq 2$, we apply Minkowski's inequality and use (32) to conclude that

$$\mathbb{E}[|M_n^{(l)}(f)|^p] \leq C\left\{\mathbb{E}\left[\sum_{k=1}^{n} \mathbb{E}^{2/p}\big(V^{p\beta}(X_{k-1}^{(l)})\big)\right]\right\}^{p/2} \leq Cn^{p/2}.$$

If $1 < p \leq 2$, we use the inequality $(a+b)^\alpha \leq a^\alpha + b^\alpha$ valid for all $a, b \geq 0$, $\alpha \in [0, 1]$ to write

$$\begin{aligned}\mathbb{E}[|M_n^{(l)}(f)|^p] &\leq C\mathbb{E}\left(\sum_{k=1}^{n} |D_k^{(l)}(f)|^p\right)\\ &\leq C\sum_{k=1}^{n} \mathbb{E}\big(V^{p\beta}(X_{k-1}^{(l)})\big) \leq Cn.\end{aligned}$$

□

To deal with the remaining term, we will rely on the following result which is also of some independent interest.

LEMMA 4.8. *Let $\mu, \mu_1, \ldots$ be a sequence of probability measures on a measurable space $(\mathcal{X}, \mathcal{B})$ such that $\mu_n(A) \to \mu(A)$ for all $A \in \mathcal{B}$ and let $f, f_1, \ldots$ be a sequence of measurable real-valued functions defined on $(\mathcal{X}, \mathcal{B})$ such that $\sup_n |f_n|_V < \infty$ and $f_n(x) \to f(x)$ for all $x \in \mathcal{X}$ for some measurable function $V : (\mathcal{X}, \mathcal{B}) \to (0, \infty)$ such that $\mu(V) < \infty$ and $\sup_n \mu_n(V^\alpha) < \infty$ for some $\alpha > 1$. Then*

$$\lim_{n\to\infty} \mu_n(f_n) = \mu(f).$$

PROOF. By [19], Chapter 11, Proposition 18, we only need to prove that $\mu_n(V) \to \mu(V)$. By [19], Chapter 11, Proposition 17, we already have $\mu(V) \leq \liminf_{n\to\infty} \mu_n(V)$. Now we show that $\limsup_{n\to\infty}\mu_n(V) \leq \mu(V)$ which will prove the lemma.

Since $V > 0$, there exists a sequence of nonnegative simple measurable functions $\{V_n\}$ that converges increasingly to V μ-a.s. For $k \geq 1$, $N \geq 1$, define $E_{k,N} = \{x \in \mathcal{X} : V(x) - V_p(x) \geq \frac{1}{k}, \text{ for some } p \geq N\}$. Clearly, $E_{k,N} \in \mathcal{B}$ and $\mu(E_{k,N}) \to 0$ as $N \to \infty$ for any $k \geq 1$. Fix $k, N \geq 1$. Then for any $n \geq 1$ and any $p \geq N$, we have

$$\begin{aligned}\mu_n(V) &= \mu_n(V_p) + \mu_n(V - V_p)\\ &= \mu_n(V_p) + \int_{E_{k,N}} \mu_n(dx)\big(V(x) - V_p(x)\big)\\ &\qquad + \int_{E_{k,N}^c} \mu_n(dx)\big(V(x) - V_p(x)\big)\\ &\leq \mu_n(V_p) + \int_{E_{k,N}} \mu_n(dx)V(x) + \frac{1}{k}\\ &\leq \mu_n(V_p) + C(\mu_n(E_{k,N}))^q + \frac{1}{k},\end{aligned} \tag{39}$$

with $q = 1 - 1/\alpha$ for some finite constant C. The last inequality uses the inequality of Holder and the assumption that $\sup_n \mu_n(V^\alpha) < \infty$ for some $\alpha > 1$. Since V_k is simple, $\mu_n(V_k) \to \mu(V_k)$. Also $\mu_n(E_{k,N}) \to \mu(E_{k,N})$. With these and letting $n \to \infty$ and $p \to \infty$ in (39), we have by monotone convergence

$$\limsup_{n\to\infty} \mu_n(V) \le \mu(V) + C(\mu(E_{k,N}))^q + \frac{1}{k}.$$

Letting $N \to \infty$ and then $k \to \infty$, we get $\limsup_{n\to\infty} \mu_n(V) \le \mu(V)$. $\square$

LEMMA 4.9. *$\pi^{(l)}_{\mu_n^{(l-1)}}(f_{\mu_n^{(l-1)}}) \to 0$ as $n \to \infty$ with $\mathbb{P}$ probability one.*

PROOF. To simplify the notations, we write $\pi_n^{(l)}$, $P_n^{(l)}$ and f_n instead of $\pi^{(l)}_{\mu_n^{(l-1)}}$, $P^{(l)}_{\mu_n^{(l-1)}}$ and $f_{\mu_n^{(l-1)}}$ respectively. For $x \in \mathcal{X}$, and $n, m \ge 1$, we have

$$\begin{aligned}
\left|\pi_n^{(l)}(f_n) - \pi^{(l)}\left(f_{\pi^{(l-1)}}\right)\right| &\le \left|\pi_n^{(l)}(f_n) - \left(P_n^{(l)}\right)^m f_n(x)\right| \\
&\quad + \left|\left(P_n^{(l)}\right)^m f_n(x) - \left(K^{(l)}\right)^m f_{\pi^{(l-1)}}(x)\right| \\
&\quad + \left|\left(K^{(l)}\right)^m f_{\pi^{(l-1)}}(x) - \pi^{(l)}\left(f_{\pi^{(l-1)}}\right)\right| \\
&\le 2 \sup_{\nu \in \mathcal{M}} |f_\nu|_{V^\beta} C_\beta V^\beta(x) \rho_\beta^m \\
&\quad + \left|\left(P_n^{(l)}\right)^m f_n(x) - \left(K^{(l)}\right)^m f_{\pi^{(l-1)}}(x)\right|,
\end{aligned} \tag{40}$$

using (28). We will show next that there exists $\mathcal{D}_0 \in \mathcal{F}$, with $\Pr(\mathcal{D}_0) = 1$ such that for each path $\omega \in \mathcal{D}_0$, $(P_n^{(l)})^m f_n(x)(\omega)$ converges to $(K^{(l)})^m f_{\pi^{(l-1)}}(x)$ as $n \to \infty$ for all $x \in \mathcal{X}$, all $m \ge 0$. Then, going back to (40), we can conclude that for each $\omega \in \mathcal{D}_0$,

$$\limsup_{n\to\infty} \left|\pi_n^{(l)}(f_n) - \pi^{(l)}\left(f_{\pi^{(l-1)}}\right)\right|(\omega) \le 2C_\beta V^\beta(x) \rho_\beta^m$$

and the proof will be finished by letting $m \to \infty$.

We can rewrite $P_n^{(l)}(x, A)$ as

$$P_n^{(l)}(x, A) = \theta_l P^{(l)}(x, A) + (1 - \theta_l) N_n^{(l)}(x, A) + (1 - \theta_l) \mathbf{1}_A(x)\left(1 - N_n^{(l)}(x, \mathbf{I})\right),$$

where $N_n^{(l)}(x, A) = \int \mu_n(dy) \mathbf{1}_A(y) \min(1, \frac{r^{(l)}(y)}{r^{(l)}(x)})$ and $N_n^{(l)}(x, \mathbf{I}) = \int \mu_n(dy) \times \min(1, \frac{r^{(l)}(y)}{r^{(l)}(x)})$.

By the law of large numbers assumed for $\{X_n^{(l-1)}, n \ge 0\}$, and since $(\mathcal{X}, \mathcal{B})$ is Polish, there exists a dense countable subset $\mathcal{C}$ in $\mathcal{X}$, a countable generating algebra $\mathcal{B}_0$ of $\mathcal{B}$ and $\mathcal{D} \in \mathcal{F}$, $\mathbb{P}(\mathcal{D}) = 1$ such that for all $x \in \mathcal{C}$ and all $A \in \mathcal{B}_0$:

$$N_n^{(l)}(x, A) \to N^{(l)}(x, A) \qquad \text{as } n \to \infty, \tag{41}$$

$$N_n^{(l)}(x, \mathbf{I}) \to N^{(l)}(x, \mathbf{I}) \qquad \text{as } n \to \infty. \tag{42}$$

We can also choose $\mathcal{D}$ such that the convergence of $f_n(x)(\omega)$ to $f_{\pi^{(l-1)}}(x)$ for all $x \in \mathcal{X}$ which is assumed in the theorem hold for all $\omega \in \mathcal{D}$. If we fix a sample path $\omega \in \mathcal{D}$, and we fix $x \in \mathcal{C}$, the convergence in (41) can actually be extended to all $A \in \mathcal{B}$ by a classical measure theory argument. Also, again for $\omega \in \mathcal{D}$ and $A \in \mathcal{B}$ fixed, we can extend the convergence in (41)–(42) to hold for all $x \in \mathcal{X}$. To see why, take $x \in \mathcal{X}$ arbitrary. Lemma 4.3 and the continuity of E implies that $N_\mu(x, A)$ is a continuous function of x uniformly in μ. Since $\mathcal{C}$ is dense, for all $k \geq 1$, there is $x_k \in \mathcal{C}$ such that

$$\left|N_\mu^{(l)}(x, A) - N_\mu^{(l)}(x_k, A)\right| \leq \frac{1}{k}$$

for all μ. In particular, $N_n^{(l)}(x, A) \geq N_n^{(l)}(x_k, A) - 1/k$ for all $n \geq 1$. As $n \to \infty$, it follows that $\liminf_{n\to\infty} N_n^{(l)}(x, A) \geq N_{\pi^{(l-1)}}^{(l)}(x_k, A) - 1/k$. As $k \to \infty$, by the continuity of $N_{\pi^{(l-1)}}^{(l)} f(\cdot)$ (Lemma 4.3), we see that $\liminf_{n\to\infty} N_n^{(l)}(x, A) \geq N_{\pi^{(l-1)}}^{(l)}(x, A)$. Similarly, we obtain $\limsup_{n\to\infty} N_n^{(l)}(x, A) \leq N_{\pi^{(l-1)}}^{(l)}(x, A)$. So that $\lim_{n\to\infty} N_n^{(l)}(x, A) = N_{\pi^{(l-1)}}^{(l)}(x, A)$. Similarly, $\lim_{n\to\infty} N_n^{(l)}(x, I) = N_{\pi^{(l-1)}}^{(l)}(x, I)$.

This shows that for each sample path $\omega \in \mathcal{D}$, $P_n^{(l)}(x, A)$ converges to $K^{(l)}(x, A)$ for all $x \in \mathcal{X}$ all $A \in \mathcal{B}$. By a successive application of Lemma 4.8 (with $V \equiv 1$), we can therefore conclude that for each sample path $\omega \in \mathcal{D}$

$$\begin{gathered}(P_n^{(l)})^m(x, A) \to (K^{(l)})^m(x, A), \\ \text{as } n \to \infty \text{ for all } x \in \mathcal{X}, A \in \mathcal{B}, m \geq 0.\end{gathered} \tag{43}$$

Since $\sup_n |f_n|_{V^\beta} < \infty$ ($\beta \in [0, 1)$) and $(P_\mu^{(l)})^m V(x)$ is uniformly bounded in μ and m, we can apply Lemma 4.8 again to conclude that for each $\omega \in \mathcal{D}$, $(P_n^{(l)})^m f_n(x)$ converges to $(K^{(l)})^m f_{\pi^{(l-1)}}(x)$ for all $x \in \mathcal{X}$, all $m \geq 0$, which ends the proof. $\square$

PROOF OF THEOREM 3.1. We are now in position to prove Theorem 3.1. Since $\beta \in [0, 1)$, we can take $p = 1/\beta$ in Lemmas 4.5 and 4.6 to conclude that $R_{i,n}^{(l)}(f)/n \to 0$, $\mathbb{P}$-a.s. for $i = 1, 2$ and by the strong law of large numbers for martingales [9], we conclude that $M_n^{(l)}(f)/n \to 0$, $\mathbb{P}$-a.s. We finish the proof using Lemma 4.9. $\square$

4.4. *Proof of Theorem* 3.2. Take $p = 1/\beta > 2$ (since $\beta \in [0, 1/2)$). By the martingale approximation (37),

$$S_n^{(l)}(f) - \sum_{k=1}^n \pi_{\mu_{k-1}^{(l-1)}}^{(l)}\left(f_{\mu_{k-1}^{(l-1)}}\right) = M_n^{(l)}(f) + R_n^{(l)}(f).$$

As above, we will simplify the notations by writing $\pi_n^{(l)}(f_n)$ instead of $\pi^{(l)}_{\mu_{k-1}^{(l-1)}}(f_{\mu_{k-1}^{(l-1)}})$ and similarly for $U_n^{(l)}$, $P_n^{(l)}$, etc.

By Lemmas 4.5–4.6, $\mathbb{E}[|R_n^{(l)}(f)|^p] = O((\log(n))^p)$. We then deduce that $R_n^{(l)}(f)/\sqrt{n} \xrightarrow{P} 0$ and it remains to show that a central limit theorem hold for the martingale $\{M_n^{(l)}(f), \mathcal{F}_n\}$. We need to show that the Lindeberg condition holds:

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\big[(D_k^{(l)})^2(f)\mathbf{1}_{\{|D_k^{(l)}(f)|>\varepsilon\sqrt{n}\}}\big] \xrightarrow{P} 0 \qquad \text{for all } \varepsilon > 0 \text{ as } n \to \infty \tag{44}$$

and that

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\big[(D_k^{(l)})^2(f)|\mathcal{F}_{k-1}\big] \xrightarrow{P} \sigma^2(f), \tag{45}$$

where $\sigma^2(f) = \pi(f^2) + 2\sum_{i=1}^{\infty} \pi^{(l)}[f(K^{(l)})^i f]$. Since $\sup_n \mathbb{E}(|D_n^{(l)}(f)|^p) < \infty$ for $p > 2$, it follows that the Lindeberg condition (44) holds.

For the law of large numbers, we need some notations. Let $U^{(l)}$ denote the fundamental kernel of the limiting kernel $K^{(l)}$ and define the functions $\Delta_n^{(1)}(x) = P_n^{(l)}(U_n^{(l)})^2 f(x)$ and $\Delta_n^{(2)}(x) = [P_n^{(l)} U_n^{(l)} f(x)]^2$. Simularly, define $\Delta^{(1)}(x) = K^{(l)}(U^{(l)})^2 f(x)$ and $\Delta^{(2)}(x) = [K^{(l)} U^{(l)} f(x)]^2$. Then we can rewrite

$$\begin{aligned}
&\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\big((D_k^{(l)})^2(f)|\mathcal{F}_{k-1}\big) \\
&\qquad = \frac{1}{n}\sum_{k=1}^{n} P_{k-1}^{(l)}(U_{k-1}^{(l)})^2 f(X_{k-1}^{(l)}) - [P_{k-1}^{(l)} U_{k-1}^{(l)} f(X_{k-1}^{(l)})]^2 \\
&\qquad = \frac{1}{n}\sum_{k=1}^{n} \Delta_{k-1}^{(1)}(X_{k-1}^{(l)}) + \Delta_{k-1}^{(2)}(X_{k-1}^{(l)}).
\end{aligned}$$

Fix $f \in L_{V^\beta}^{\infty}$. We have seen in the proof of Theorem 3.1 that $\pi_n^{(l)}(f)$ converges almost surely to $\pi^{(l)}(f)$. Combined with (43) and using dominated convergence it follows that there is $\mathcal{D} \in \mathcal{F}$, $\Pr(\mathcal{D}) = 1$ such that for all sample path $\omega \in \mathcal{D}$, $U_n^{(l)} f(x)$ converges to $U^{(l)} f(x)$ for all $x \in \mathcal{X}$. By virtue of Lemma 4.8, it follows that for all $\omega \in \mathcal{D}$, $\Delta_n^{(j)}(x)$ converges to $\Delta^{(j)}(x)$ for all $x \in \mathcal{X}$, $j = 1, 2$. Then the strong law of large numbers (Theorem 3.1), implies that $\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}((D_k^{(l)})^2(f)|\mathcal{F}_{k-1})$ converges almost surely to $\pi^{(l)}(K^{(l)}(U^{(l)})^2 f - [K^{(l)} U^{(l)} f]^2)$ which is equal to $\sigma^2(f) = \pi^{(l)}(f^2) + 2\sum_{i=1}^{\infty} \pi^{(l)}[f(K^{(l)})^i f]$.

4.5. *Proof of Theorem* 3.3. We continue with the notations of Section 3.4.

LEMMA 4.10. *Under the assumptions of Theorem* 3.3, *there exists a finite* *constant* c_0 *such that*

$$|\Gamma(x_1, x) - \Gamma(x, x)| \le c_0|x_1 - x| \qquad \textit{for all } x, x_1 \in \mathcal{X}.$$

PROOF. Given the expression of Γ in (21), it is enough to show that $|U_x^{(0)}(y) - U_{x_1}^{(0)}(y)| \le c_0|x - x_1|$. But since

$$\left|U_x^{(0)}(y) - U_{x_1}^{(0)}(y)\right| = \left|\sum_{j\ge 0}[\bar{P}^{(0)}]^j\left(H_x(y) - H_{x_1}(y)\right)\right| \le C|H_x - H_{x_1}|_\infty$$

(where for a kernel P with invariant distribution π, $\bar{P} = P - \pi$), the lemma follows if we show that there exists a finite constant c_0 such that for any $x_1, x_2, y \in \mathcal{X}$,

$$|H_{x_1}(y) - H_{x_2}(y)| \le c_0|x_1 - x_2|.$$

It is easy to check as in Lemma 4.3 that for any $x_1, x_2, y \in \mathcal{X}$,

$$\begin{aligned}|H_{x_1}(y) - H_{x_2}(y)| &\le 2|U(x_1) - U(x_2)| \\ &\quad + |U|_\infty\left(e^{-\tau E(y)} + \int e^{-\tau E(y)}\pi^{(0)}(dy)\right)\left|e^{\tau E(x_1)} - e^{\tau E(x_2)}\right|.\end{aligned}$$

Now the result follow from (19), the Lipschitz assumption on E and the compactness of $\mathcal{X}$. □

PROPOSITION 4.1. *Under the assumptions of Theorem* 3.3, η_n *converges weakly in* $\mathcal{C}(\mathcal{X}, \mathbb{R})$ *to a mean zero Gaussian process* G *with covariance function* Γ *and sample paths in* $\mathcal{C}(\mathcal{X}, \mathbb{R})$ *and*

$$\mathbb{E}\left(\sup_{x\in\mathcal{X}} |G(x)|\right) < \infty. \tag{46}$$

PROOF. The existence of G and the bound (46) follows from Lemma 4.10 and Dudley's Theorem on the existence of Gaussian processes with continuous sample paths (see, e.g., [16], Theorem 6.1.2). Indeed, if $d_\Gamma(x, y) := (\Gamma(x, x) + \Gamma(y, y) - 2\Gamma(x, y))^{1/2}$ denotes the pseudo-metric associated to Γ, Lemma 4.10 implies that $d_\Gamma(x, y) \le \sqrt{2c_0}|x - y|^{1/2}$ and since $\mathcal{X}$ is compact, this in turn implies that $\mathcal{N}(\mathcal{X}, d_\Gamma, \epsilon) \le (K\epsilon^{-1})^{d/2}$ for some finite constant K, where $\mathcal{N}(\mathcal{X}, d_\Gamma, \cdot)$ is the metric entropy of $\mathcal{X}$ under d_Γ.

We now show that η_n converges weakly in $\mathcal{C}(\mathcal{X}, \mathbb{R})$ to a mean zero Gaussian process with continuous sample path and covariance function Γ. Indeed, the convergence of the finite-dimensional distribution is given by the standard central limit for uniformly ergodic Markov chains. We use a moment criterion to check that the family $\{\eta_n, n \ge 0\}$ is tight ([14], Corollary 16.9). It suffices to check that:

(i) For some $x_0 \in \mathcal{X}$, $\{\eta_n(x_0), n \ge 0\}$ is tight.

(ii) For some positive finite constant a, b, c_0,

$$\mathbb{E}[|\eta_n(x_1) - \eta_n(x_2)|^a] \le c_0|x_1 - x_2|^{d+b} \qquad \text{for all } x_1, x_2 \in \mathcal{X}, n \ge 0.$$

The condition (i) is trivially true. To check (ii), we use the resolvent $U_x^{(0)}$ to write $H_{x_1}(y) - H_{x_2}(y) = (U_{x_1}^{(0)}(y) - U_{x_2}^{(0)}(y)) - (P^{(0)}U_{x_1}^{(0)}(y) - P^{(0)}U_{x_2}^{(0)}(y))$. It follows that

$$\eta_n(x_1) - \eta_n(x_2) = M_n(x_1, x_2) + \epsilon_n(x_1, x_2),$$

where $M_n(x_1, x_2) = \sum_{k=1}^n (U_{x_1}^{(0)}(X_k^{(0)}) - U_{x_2}^{(0)}(X_k^{(0)})) - (P^{(0)}U_{x_1}^{(0)}(X_{k-1}^{(0)}) - P^{(0)}U_{x_2}^{(0)}(X_{k-1}^{(0)}))$ and $\epsilon_n(x_1, x_2) = P^{(0)}U_{x_1}^{(0)}(X_0^{(0)}) - P^{(0)}U_{x_2}^{(0)}(X_0^{(0)}) - P^{(0)}U_{x_1}^{(0)}(X_n^{(0)}) - P^{(0)}U_{x_2}^{(0)}(X_n^{(0)})$.

The term $M_n(x_1, x_2)$ is a martingale and $\epsilon_n(x_1, x_2)$ is bounded in n by a constant. By Burkholder's inequality and some additional straightforward arguments it follows that for any $a \ge 2$

$$E[|\eta_n(x_1) - \eta_n(x_2)|^a] \le C\big|U_{x_1}^{(0)} - U_{x_2}^{(0)}\big|_\infty^a \le C|x_1 - x_2|^a.$$

Then it suffices to take $a > d$. $\square$

We will also need the following simple result.

LEMMA 4.11. *If $\{x_k\}$ is a sequence of real numbers such that $x_n \to 0$ as $n \to \infty$ then $n^{-1/2}\sum_{k=1}^n k^{-1/2}x_k \to 0$ as $\to \infty$.*

PROOF. Take $\varepsilon > 0$. Let $n_0 \ge 1$ s.t. $n \ge n_0$ implies $|x_n| \le \varepsilon$. Then for $n \ge n_0$, $n^{-1/2}|\sum_{k=1}^n k^{-1/2}x_k| \le n^{-1/2}\sum_{k=1}^{n_0} k^{-1/2}|x_k| + n^{-1/2}\sum_{k=n_0+1}^n k^{-1/2}\varepsilon \le n^{-1/2}\sum_{k=1}^{n_0} k^{-1/2}|x_k| + 2\varepsilon$. Letting $n \to \infty$ and $\varepsilon \to 0$ yields the result. $\square$

PROOF OF THEOREM 3.3. For the rest of the proof, let G be a mean zero Gaussian process on $\mathcal{X}$ with covariance function Γ and almost surely continuous sample paths. We take G independent from the process $\{(X_n^{(0)}, X_n^{(1)}), n \ge 0\}$. From the Gaussian process G, we define $\pi(G) := \int G(x)\pi^{(1)}(dx)$ as follows. For each sample path $\omega \in \Omega$, if $G_\omega(\cdot)$ is continuous then $\pi(G)(\omega) = \int \pi^{(1)}(dx)G_\omega(x)$. Otherwise, we set $\pi(G)(\omega) = 0$. Since $f \to \pi^{(1)}(f)$ is a continuous map from $\mathcal{C}(\mathcal{X}, \mathbb{R}) \to \mathbb{R}$, $\pi^{(1)}(G)$ is a well-defined random variable.

Back to the partial sum S_n, we have seen that

$$S_n = M_n + (1 - \theta_1)\sum_{k=1}^n k^{-1/2}\eta_k(X_k^{(1)}) + \epsilon_n^{(1)},$$

where $M_n := \sum_{k=1}^n U(X_k^{(1)}) - P_{\mu_{k-1}^{(0)}}U(X_{k-1}^{(1)})$ and $\epsilon_n^{(1)} = (P_{\mu_0^{(0)}}U(X_0^{(1)}) - P_{\mu_n^{(0)}} \times U(X_n^{(1)}))$. Clearly

$$\sup_{n\ge1}\big|(P_{\mu_0^{(0)}}U(X_0^{(1)}) - P_{\mu_n^{(0)}}U(X_n^{(1)}))\big| \le C,$$

thus the term $\epsilon_n^{(1)}$ is negligible. That is,

$$\begin{aligned}S_n &= M_n + (1-\theta_1)\sum_{k=1}^n k^{-1/2}\eta_k\big(X_k^{(1)}\big) + o_P(\sqrt{n}),\\ &= M_n + (1-\theta_1)\sum_{k=1}^n \frac{1}{\sqrt{k}} G\big(X_k^{(1)}\big)\\ &\quad + (1-\theta_1)\sum_{k=1}^n k^{-1/2}\big(\eta_k\big(X_k^{(1)}\big) - G\big(X_k^{(1)}\big)\big) + o_P(\sqrt{n}).\end{aligned}$$

In the above, we denote $o_P(n^r)$ any random variable X_n such that $n^{-r}X_n$ converges in probability to zero. To deal with the term $\sum_{k=1}^n k^{-1/2}(\eta_n(X_k^{(1)}) - G(X_k^{(1)}))$, we use the Skorohod representation of weak convergence. First note that

$$\begin{aligned}&\left|n^{-1/2}\sum_{k=1}^n k^{-1/2}\big(\eta_n\big(X_k^{(1)}\big) - G\big(X_k^{(1)}\big)\big)\right|\\ &\qquad \le n^{-1/2}\sum_{k=1}^n k^{-1/2}\sup_{x\in\mathcal{X}}|\eta_n(x) - G(x)|.\end{aligned}$$

By the Skorohod representation theorem, there exists a version $\tilde{G}$ of G and a version $\{\tilde{\eta}_n, n \ge 0\}$ of the random process $\{\eta_n, n \ge 0\}$ such that $\sup_{x\in\mathcal{X}}|\tilde{\eta}_n(x) - \tilde{G}(x)| \to 0$ a.s. Therefore, by Lemma 4.11, $n^{-1/2}\sum_{k=1}^n k^{-1/2}\sup_{x\in\mathcal{X}}|\tilde{\eta}_n(x) - \tilde{G}(x)|$ converges almost surely and thus in probability to zero. It follows that $n^{-1/2}\sum_{k=1}^n k^{-1/2}(\eta_n(X_k^{(1)}) - G(X_k^{(1)}))$ converges also in probability to zero. We thus arrive at

$$S_n = M_n + (1-\theta_1)\sum_{k=1}^n \frac{1}{\sqrt{k}} G\big(X_k^{(1)}\big) + o_P(\sqrt{n}).$$

To deal with the term $\sum_{k=1}^n \frac{1}{\sqrt{k}} G(X_k^{(1)})$, we introduce $V_0 = 0$ and $V_k = \sum_{j=1}^k (G(X_j^{(1)}) - \pi^{(1)}(G))$:

$$\begin{aligned}&\sum_{k=1}^n \frac{1}{\sqrt{k}}\big(G\big(X_k^{(1)}\big) - \pi^{(1)}(G)\big)\\ &\qquad = \sum_{k=1}^n \frac{1}{\sqrt{k}}(V_k - V_{k-1})\\ &\qquad = \sum_{k=1}^n \frac{1}{\sqrt{k}} V_k - \sum_{k=2}^n \left(\frac{1}{\sqrt{k}} - \frac{1}{\sqrt{k-1}}\right) V_{k-1} - \sum_{k=2}^n \frac{1}{\sqrt{k-1}} V_{k-1}\end{aligned}$$

$$= \frac{1}{\sqrt{n}} V_n + \sum_{k=2}^{n} \frac{1}{\sqrt{k(k-1)}(\sqrt{k}+\sqrt{k-1})} V_{k-1}$$

$$= \frac{1}{\sqrt{n}} V_n + \sum_{k=2}^{n} \left(\frac{1}{\sqrt{k}(1+\sqrt{1+1/(k-1)})} \right) \frac{1}{k-1} V_{k-1}.$$

We deduce that

$$n^{-1/2} S_n = n^{-1/2} M_n + (1-\theta_1)\pi^{(1)}(G) n^{-1/2} \sum_{k=1}^{n} k^{-1/2} + n^{-1} V_n$$
$$+ \frac{1}{\sqrt{n}} \sum_{k=2}^{n} \left(\frac{1}{\sqrt{k}(1+\sqrt{1+1/(k-1)})} \right) \frac{1}{k-1} V_{k-1} + o_P(1).$$

For almost every path $\omega \in \Omega$, $G_\omega(\cdot)$ is a continuous function from $\mathcal{X} \to \mathbb{R}$. Therefore, by the independence assumption and the law of large numbers of Theorem 3.1, $n^{-1}\sum_{j=1}^{k} G(X_j^{(1)}) - \pi^{(1)}(G)$ converges in L^1 to zero. Using Lemma 4.11 again, we conclude that $\frac{1}{\sqrt{n}}\sum_{k=2}^{n}(\frac{1}{\sqrt{k}(1+\sqrt{1+1/(k-1)})})\frac{1}{k-1}V_{k-1}$ converges also in L^1 to zero. The term $\frac{1}{\sqrt{n}}\sum_{k=1}^{n} k^{-1/2}$ converges to 2. We thus arrive at

$$n^{-1/2} S_n = n^{-1/2} M_n + 2(1-\theta_1)\pi^{(1)}(G) + o_P(1).$$

Proceeding as in the proof of Theorem 3.2, we see that $\frac{1}{\sqrt{n}} M_n$ converges weakly to Z, where $Z \sim N(0, \sigma_\star^2(f))$ and is independent from G. We thus conclude that $n^{-1/2} S_n$ converges weakly to $Z + 2(1-\theta_1)\int \pi^{(1)}(dx) G(x)$, where Z and $\int \pi^{(1)}(dx)G(x)$ are independent.

Since $f \to \pi^{(1)}(f)$ is a continuous bounded function from $\mathcal{C}(\mathcal{X}, \mathbb{R}) \to \mathbb{R}$, it follows from the above that $\pi^{(1)}(\eta_n)$ converges weakly to $\pi^{(1)}(G)$. But $\pi^{(1)}(\eta_n) = n^{-1/2}\sum_{k=1}^{n}\int \pi^{(1)}(dx) H_x(X_k^{(0)})$. By the central limit theorem for the uniformly ergodic chain $\{X_n^{(0)}, n \ge 0\}$, the latter term $n^{-1/2}\sum_{k=1}^{n}\int \pi^{(1)}(dx)H_x(X_k^{(0)})$ converges weakly to $N(0, \Gamma(\bar{g}, \bar{g}))$, where $\bar{g}(\cdot) = \int \pi^{(1)}(dx) H_x(\cdot)$ and we are finished. $\square$

4.6. *Proof of Proposition* 3.1. In the present case, one can check that $U(x) = \sum_{j\ge 0}(P_{\pi^{(0)}}^{(1)})^j f(x) = \sum_{j\ge 0}\theta_1^j P^j f(x)$ and $H_x(y) = U(y)$. Then the resolvent function $U_x^{(0)}$ becomes $U_x^{(0)}(y) = U^{(0)}(y) = \sum_{j\ge 0} P^j U(y)$ which allows use to write $\sum_{j=1}^{k} H_x(X_j^{(0)}) = M_k^{(0)} + \epsilon_k^{(0)}$, where $M_k^{(0)} = \sum_{j=1}^{k} U^{(0)}(X_k^{(0)}) - PU^{(0)}(X_{k-1}^{(0)})$ and $\epsilon_k^{(0)} = PU^{(0)}(X_0^{(0)}) - PU^{(0)}(X_k^{(0)})$. Thus we have

$$S_n = M_n + (1-\theta_1)\sum_{k=1}^{n} k^{-1} M_k^{(0)} + \epsilon_n,$$

where $\epsilon_n = \epsilon_n^{(1)} + \sum_{k=1}^n k^{-1}\epsilon_k^{(0)}$. The term ϵ_n is negligible and is suffices to study the limit of

$$\mathbb{E}\left[\left(M_n + (1-\theta_1)\sum_{k=1}^n k^{-1}M_k^{(0)}\right)^2\right] = \mathbb{E}(M_n^2) + (1-\theta_1)^2\mathbb{E}\left[\left(\sum_{k=1}^n k^{-1}M_k^{(0)}\right)^2\right] + 2(1-\theta_1)\mathbb{E}\left[M_n\sum_{k=1}^n k^{-1}M_k^{(0)}\right].$$

Define $D^{(0)}(x,y) = U^{(0)}(y) - PU^{(0)}(x)$ and $D^{(1)}(x,y) = U(y) - PU(x)$. It is easy to see that for any $i, j \geq 1$, $\mathbb{E}(D^{(0)}(X_{i-1}^{(0)}, X_i^{(0)})D^{(1)}(X_{j-1}^{(1)}, X_j^{(1)})) = 0$. From which we deduce that $\mathbb{E}[M_n\sum_{k=1}^n k^{-1}M_k^{(0)}] = 0$.

We write $\sum_{k=1}^n k^{-1}M_k^{(0)} = \sum_{j=1}^n\sum_{k=j}^n k^{-1}D^{(0)}(X_{j-1}^{(0)}, X_j^{(0)})$ and since the terms $D^{(0)}(X_{j-1}^{(0)}, X_j^{(0)})$ are martingale differences, we get

$$\begin{aligned}
&\mathbb{E}\left[\left(\sum_{k=1}^n k^{-1}M_k^{(0)}\right)^2\right] \\
&\quad = \mathbb{E}\left[\left(\sum_{j=1}^n\left(\sum_{k=j}^n k^{-1}\right)D^{(0)}(X_{j-1}^{(0)}, X_j^{(0)})\right)^2\right] \\
&\quad = \sum_{j=1}^n\left(\sum_{k=j}^n k^{-1}\right)^2\mathbb{E}[(D^{(0)}(X_{j-1}^{(0)}, X_j^{(0)}))^2] \\
&\quad = \int\pi(dx)\int P(x,dy)(D^{(0)}(x,y))^2\sum_{j=1}^n\left(\sum_{k=j}^n k^{-1}\right)^2 \\
&\qquad + \sum_{j=1}^n\left(\sum_{k=j}^n k^{-1}\right)^2\Big(\mathbb{E}[(D^{(0)}(X_{j-1}^{(0)}, X_j^{(0)}))^2] \\
&\qquad\qquad - \int\pi(dx)\int P(x,dy)(D^{(0)}(x,y))^2\Big).
\end{aligned}$$

Since $D^{(0)}$ is a bounded continuous function and $\{X_n^{(0)}\}$ is uniformly ergodic, the second term on the r.h.s. divided by n converges to zero. Then we notice that $\lim_{n\to\infty} n^{-1}\sum_{i=1}^n(\sum_{k=j}^n k^{-1})^2 = 2$ and we conclude that

$$\lim_{n\to\infty}\mathbb{E}(n^{-1}S_n^2) = \int\pi(dx)\int P(x,dy)\{(D^{(1)}(x,y))^2 + 2(1-\theta_1)^2(D^{(0)}(x,y))^2\}.$$

Acknowledgments. The author is grateful to Eric Moulines and Gersende Fort for helpful discussions and to an anonymous referee for helping improve the quality of this work.

REFERENCES

[1] ANDRIEU, C. and ATCHADÉ, Y. F. (2007). On the efficiency of adaptive MCMC algorithms. *Electron. Comm. Probab.* **12** 336–349 (electronic). MR2350572

[2] ANDRIEU, C. and MOULINES, É. (2006). On the ergodicity properties of some adaptive MCMC algorithms. *Ann. Appl. Probab.* **16** 1462–1505. MR2260070

[3] ANDRIEU, C., JASRA, A., DOUCET, A. and DEL MORAL, P. (2007). On non-linear Markov chain Monte Carlo via self-interacting approximations. Technical report.

[4] ATCHADÉ, Y. F. (2006). An adaptive version for the Metropolis adjusted Langevin algorithm with a truncated drift. *Methodol. Comput. Appl. Probab.* **8** 235–254. MR2324873

[5] ATCHADÉ, Y. F. (2009). Resampling from the past to improve on Monte Carlo samplers. *Far East J. Theor. Stat.* **27** 81–99.

[6] ATCHADÉ, Y. F. and FORT, G. (2008). Limit theorems for some adaptive MCMC algorithms with sub-geometric kernels. Technical report, available at arxiv.0807.2952.

[7] BAXENDALE, P. H. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.* **15** 700–738. MR2114987

[8] BERCU, B., DEL MORAL, P. and DOUCET, A. (2008). Fluctuations of interacting Markov Chain Monte Carlo models. Technical Report 6438, INRIA.

[9] CHOW, Y. S. (1967). On a strong law of large numbers for martingales. *Ann. Math. Statist.* **38** 610. MR0208648

[10] DEL MORAL, P. and DOUCET, A. (2008). Interacting Markov chain Monte Carlo methods for solving nonlinear measure-valued equations. Technical Report 6435, INRIA.

[11] GILKS, W. R., ROBERTS, G. O. and SAHU, S. K. (1998). Adaptive Markov chain Monte Carlo through regeneration. *J. Amer. Statist. Assoc.* **93** 1045–1054. MR1649199

[12] HAARIO, H., SAKSMAN, E. and TAMMINEN, J. (2001). An adaptive Metropolis algorithm. *Bernoulli* **7** 223–242. MR1828504

[13] JARNER, S. F. and HANSEN, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.* **85** 341–361. MR1731030

[14] KALLENBERG, O. (2002). *Foundations of Modern Probability*, 2nd ed. Springer, New York. MR1876169

[15] KOU, S. C., ZHOU, Q. and WONG, W. H. (2006). Equi-energy sampler with applications in statistical inference and statistical mechanics. *Ann. Statist.* **34** 1581–1652. MR2283711

[16] MARCUS, M. B. and ROSEN, J. (2006). *Markov Processes, Gaussian Processes, and Local Times. Cambridge Studies in Advanced Mathematics* **100**. Cambridge Univ. Press, Cambridge. MR2250510

[17] MEYN, S. P. and TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer, London. MR1287609

[18] ROBERTS, G. O. and ROSENTHAL, J. S. (2007). Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. *J. Appl. Probab.* **44** 458–475. MR2340211

[19] ROYDEN, H. L. (1988). *Real Analysis*, 3rd ed. Prentice-Hall, Englewood Cliffs, NJ. MR1013117

DEPARTMENT OF STATISTICS
UNIVERSITY OF MICHIGAN
1085 SOUTH UNIVERSITY
ANN ARBOR, MICHIGAN 48109
USA
E-MAIL: yvesa@umich.edu