# LIMIT THEOREMS FOR SOME ADAPTIVE MCMC ALGORITHMS WITH SUBGEOMETRIC KERNELS: PART II

YVES F. ATCHADÉ AND GERSENDE FORT

Aug. 2009

ABSTRACT. We prove a central limit theorem for a general class of adaptive Markov Chain Monte Carlo algorithms driven by sub-geometrically ergodic Markov kernels. We discuss in detail the special case of stochastic approximation. We use the result to analyze the asymptotic behavior of an adaptive version of the Metropolis Adjusted Langevin algorithm with a heavy tailed target density.

## 1. Introduction

This work is a sequel of Atchade and Fort (2008) and develops central limit theorems for adaptive MCMC (AMCMC) algorithms. Previous works on the subject include Andrieu and Moulines (2006) and Saksman and Vihola (2009) where central limit theorems are proved for certain AMCMC algorithms driven by geometrically ergodic Markov kernels. There is a need to understand the sub-geometric case. Indeed, many Markov kernels routinely used in practice are not geometrically ergodic. For example, if the target distribution of interest has heavy tails, then the Random Walk Metropolis algorithm (RWMA) and the Metropolis Adjusted Langevin algorithm (MALA) result in sub-geometric Markov kernels (Jarner and Roberts (2002a)).

We consider adaptive MCMC algorithms driven by Markov kernels $\{P_\theta,\ \theta \in \Theta\}$ such that each kernel $P_\theta$ enjoys a polynomial rate of convergence towards $\pi$ and satisfies a drift condition of the form $P_\theta V \leq V - cV^{1-\alpha} + b$ for some $\alpha \in (0, 1]$ (uniformly in $\theta$ over compact sets). We obtain a central limit theorem when $\alpha < 1/2$ under some additional stability conditions. This result is very close to what can be proved for Markov chains under similar conditions. Indeed, it is known (Jarner and Roberts (2002b)) that irreducible and

2000 *Mathematics Subject Classification.* 60J10, 65C05.

*Key words and phrases.* Adaptive Markov chain Monte Carlo, Markov chain, Subgeometric ergodicity, Stochastic approximations, Metropolis adjusted Langevin algorithms.

Y. Atchadé: University of Michigan, 1085 South University, Ann Arbor, 48109, MI, United States. *E-mail address* yvesa@umich.edu.

G. Fort: LTCI, CNRS-TELECOM ParisTech, 46 rue Barrault, 75634 Paris Cedex 13, France. *E-mail address* gfort@tsi.enst.fr.

This work is partly supported by NSF grant DMS 0906631 and by the french National Research Agency (ANR) program ANR-05-BLAN-0299.

aperiodic Markov chains for which the drift condition $PV \leq V - cV^{1-\alpha} + b\mathbb{1}_{\mathcal{C}}$ hold for some small set $\mathcal{C}$ satisfy a central limit theorem when $\alpha \leq 1/2$. The slight loss of efficiency in our case ($\alpha < 1/2$ versus $\alpha \leq 1/2$) is typical of martingale approximation-based proofs. The proof of the central limit theorem is based on a martingale approximation technique initiated by Kipnis and Varadhan (1986) and Maxwell and Woodroofe (2000). The method is a Poisson equation-type method but where the Poisson's kernel is replaced by a more general resolvent kernel. We have used a variant of the same technique in Atchade and Fort (2008) to study the strong law of large numbers for AMCMC.

Adaptive MCMC has been studied in a number of recent papers. Beside the above mentioned papers, results related to the convergence of marginal distributions and the law of large numbers can be found e.g. in (Rosenthal and Roberts (2007); Bai (2008)). For specific examples and a review of the methodological developments, see e.g. Roberts and Rosenthal (2006); Andrieu and Thoms (2008); Atchade et al. (2009).

The rest of the paper is organized as follows. The main CLT result is presented in Section 2.3. Adaptive MCMC driven by stochastic approximation is considered in Section 2.6. To illustrate, we apply our theory to an adaptive version of the Metropolis adjusted Langevin algorithm (MALA) with a heavy tailed target distribution (Section 2.7). Most of the proofs are postponed to Section 3.

## 2. Statement of the results

2.1. **Notations.** We start with some notations that will be used through the paper. For a transition kernel $P$ on a measurable general state space $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$, denote by $P^n$, $n \geq 0$, its $n$-th iterate defined as

$$P^0(x, A) \stackrel{\text{def}}{=} \delta_x(A) \;, \qquad P^{n+1}(x, A) \stackrel{\text{def}}{=} \int P(x, dy) P^n(y, A) \;, \quad n \geq 0 \;;$$

$\delta_x(dt)$ stands for the Dirac mass at $\{x\}$. $P^n$ is a transition kernel on $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ that acts both on bounded measurable functions $f$ on $\mathbb{T}$ and on $\sigma$-finite measures $\mu$ on $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ via $P^n f(\cdot) \stackrel{\text{def}}{=} \int P^n(\cdot, dy) f(y)$ and $\mu P^n(\cdot) \stackrel{\text{def}}{=} \int \mu(dx) P^n(x, \cdot)$.

If $V : \mathbb{T} \to [1, +\infty)$ is a function, the $V$-norm of a function $f : \mathbb{T} \to \mathbb{R}$ is defined as $|f|_V \stackrel{\text{def}}{=} \sup_{\mathbb{T}} |f|/V$. When $V = 1$, this is the supremum norm. The set of functions with finite $V$-norm is denoted by $\mathcal{L}_V$.

If $\mu$ is a signed measure on a measurable space $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$, the total variation norm $\|\mu\|_{\text{TV}}$ is defined as

$$\|\mu\|_{\text{TV}} \stackrel{\text{def}}{=} \sup_{\{f, |f|_1 \leq 1\}} |\mu(f)| = 2 \sup_{A \in \mathcal{B}(\mathbb{T})} |\mu(A)| = \sup_{A \in \mathcal{B}(\mathbb{T})} \mu(A) - \inf_{A \in \mathcal{B}(\mathbb{T})} \mu(A) \;;$$

and the $V$-norm, where $V : \mathbb{T} \to [1, +\infty)$ is a function, is defined as $\|\mu\|_V \stackrel{\text{def}}{=} \sup_{\{g, |g|_V \leq 1\}} |\mu(g)|$. Observe that $\|\cdot\|_{\text{TV}}$ corresponds to $\|\cdot\|_V$ with $V \equiv 1$.

In the Euclidean space $\mathbb{R}^n$, we use $\langle a, b\rangle$ to denote the inner product and $|a| \stackrel{\text{def}}{=} \sqrt{\langle a, a\rangle}$ the Euclidean norm. We denote $\mathbb{R}$ the set of real numbers and $\mathbb{N}$ the set of nonnegative integers.

2.2. **Adaptive MCMC: definition.** Let $\mathsf{X}$ be a general state space resp. endowed with a countably generated $\sigma$-field $\mathcal{X}$. Let $\Theta$ be an open subspace of $\mathbb{R}^q$ the $q$-dimensional Euclidean space and $\mathcal{B}(\Theta)$ is its Borel $\sigma$-algebra. Let $\{P_\theta, \theta \in \Theta\}$ be a family of Markov transition kernels on $(\mathsf{X}, \mathcal{X})$ such that for any $(x, A) \in \mathsf{X} \times \mathcal{X}$, $\theta \mapsto P_\theta(x, A)$ is measurable. We assume that for any $\theta \in \Theta$, the Markov kernel $P_\theta$ admits an invariant distribution $\pi$. Let $\{\mathsf{K}_n, n \geq 0\}$ be a family of nonempty compact subspaces of $\Theta$ such that $\mathsf{K}_n \subseteq \mathsf{K}_{n+1}$. Let $\Pi : \mathsf{X} \times \Theta \to \mathsf{X}_0 \times \Theta_0$ be a measurable function, the so-called re-projection function, where $\mathsf{X}_0 \times \Theta_0$ is some measurable subset of $\mathsf{X} \times \Theta$. We assume that $\Pi(x, \theta) = (x, \theta)$ if $\theta \in \Theta_0$. For an integer $k \geq 0$ we define $\Pi_k(x, \theta) = \Pi(x, \theta)$ if $k = 0$ and $\Pi_k(x, \theta) = (x, \theta)$ if $k \geq 1$. Let $\bar{R}(n; \cdot, \cdot) : (\mathsf{X} \times \Theta) \times (\mathcal{X} \times \mathcal{B}(\Theta)) \to [0, 1]$ a sequence of Markov kernels on $\mathsf{X} \times \Theta$ with the following property. For any $n \geq 0$, $A \in \mathcal{X}$, $(x, \theta) \in \mathsf{X} \times \Theta$

$$\bar{R}\left(n; (x, \theta), A \times \Theta\right) = P_\theta(x, A). \tag{1}$$

In most cases in practice, the adaptation is driven by stochastic approximation. One such example of stochastic approximation is obtained by taking $\bar{R}(n; \cdot, \cdot)$ of the form $\bar{R}\left(n; (x, \theta), (dx', d\theta')\right) = P_\theta(x, dx')\delta_{\theta+\gamma_n \Upsilon_\theta(x')}(d\theta')$. But the main example of stochastic approximation considered in this paper is

$$\bar{R}\left(n; (x, \theta), (dx', d\theta')\right) = \int q_\theta^{(1)}(x, dy) q_\theta^{(2)}\left((x, y), dx'\right) \delta_{\theta+\gamma_n \Phi_\theta(x,y)}(d\theta').$$

where $q_\theta^{(1)}$ and $q_\theta^{(2)}$ are Markov kernels. Obviously, in order for (1) to hold, these kernels ought to satisfy the constraint

$$\int q_\theta^{(1)}(x, dy) q_\theta^{(2)}\left((x, y), dx'\right) = P_\theta(x, dx').$$

Throughout the paper and without further mention, we assume that (1) hold. We are interested in the Markov chain $\{(X_n, \theta_n, \nu_n, \xi_n),\ n \geq 0\}$ define on $\mathsf{X} \times \Theta \times \mathbb{N} \times \mathbb{N}$ with transition kernel $\bar{P}$,

$$\bar{P}\left((x, \theta, \nu, \xi), (dx', d\theta', d\nu', d\xi')\right) \stackrel{\text{def}}{=} \bar{R}\left(\nu + \xi; \Pi_\xi(x, \theta), (dx', d\theta')\right) \\ \times \left(\mathbb{1}_{\{\theta' \in \mathsf{K}_\nu\}} \delta_\nu(d\nu') \delta_{\xi+1}(d\xi') + \mathbb{1}_{\{\theta' \notin \mathsf{K}_\nu\}} \delta_{\nu+1}(d\nu') \delta_0(d\xi')\right). \tag{2}$$

Algorithmically, this Markov chain can be described as follows.

**Algorithm 2.1.** *Given* $(X_n, \theta_n, \nu_n, \xi_n)$*:*

- **a:** *generate* $(X_{n+1}, \theta_{n+1}) \sim \bar{R}\left(\nu_n + \xi_n; \Pi_{\xi_n}(X_n, \theta_n), \cdot\right)$*;*
- **b:** *if* $\theta_{n+1} \in \mathsf{K}_{\nu_n}$ *then set* $\nu_{n+1} = \nu_n$*,* $\xi_{n+1} = \xi_n + 1$*,*
- **c:** *if* $\theta_{n+1} \notin \mathsf{K}_{\nu_n}$ *then set* $\nu_{n+1} = \nu_n + 1$ *and* $\xi_{n+1} = 0$*.*

We denote by $\check{\mathbb{P}}_{x,\theta,\nu,\xi}$ and $\check{\mathbb{E}}_{x,\theta,\nu,\xi}$ the probability and expectation operator when the initial distribution of the Markov chain is $\delta_{(x,\theta,\nu,\xi)}$. Throughout the paper, we will assume that the initial state of the process is fixed to $(x_0, \theta_0, 0, 0)$ for some arbitrary element $(x_0, \theta_0) \in \mathsf{X}_0 \times \Theta_0$ and we will systematically write $\check{\mathbb{P}}$ and $\check{\mathbb{E}}$ instead of $\check{\mathbb{P}}_{x_0,\theta_0,0,0}$ and $\check{\mathbb{E}}_{x_0,\theta_0,0,0}$ respectively.

*Remark* 1. Algorithm 2.1 is fairly general and encompasses the two main strategies used in practice to control the adaptation parameter.

(1) For example, one obtains the framework of re-projections on randomly varying compact sets developed in (Andrieu et al. (2005); Andrieu and Moulines (2006)) by taking $\{\mathsf{K}_n, n \geq 0\}$ such that $\Theta = \bigcup_n \mathsf{K}_n$, $\Theta_0 \subseteq \mathsf{K}_0$ and $\mathsf{K}_n \subset \mathrm{int}(\mathsf{K}_{n+1})$, where $\mathrm{int}(A)$ is the interior of $A$.

(2) But we can also set $\Theta_0 = \mathsf{K}_k = \mathsf{K}$ for all $k \geq 0$ for some compact subset $\mathsf{K}$ of $\Theta$. And we then obtain another commonly used approach where the re-projection is done on a fixed compact set $\mathsf{K}$. See e.g. Atchade and Rosenthal (2005).

Let $\{\check{\mathcal{F}}_n,\ n \geq 0\}$ denote the natural filtration of the Markov chain $\{(X_n, \theta_n, \nu_n, \xi_n),\ n \geq 0\}$. It is easy to compute using (1) that for any bounded measurable function $f : \mathsf{X} \to \mathbb{R}$,

$$\check{\mathbb{E}}\left(f(X_{n+1})|\check{\mathcal{F}}_n\right) \mathbb{1}_{\{\xi_n > 0\}} = P_{\theta_n} f(X_n),\ \ \check{\mathbb{P}} - \text{a.s.} \tag{3}$$

Equation (3) together with the strong Markov property are the two main properties of the process $\{(X_n, \theta_n, \nu_n, \xi_n),\ n \geq 0\}$ that will used in the sequel.

We now introduce another stochastic process closely related to the adaptive chain defined above. For $l \geq 0$ an integer, we consider the nonhomogeneous Markov chain $\{(\tilde{X}_n, \tilde{\theta}_n),\ n \geq 0\}$ with initial distribution $\delta_{x,\theta}$ and sequence of transition Markov kernels

$$P_l\left(n; (x_1, \theta_1), (dx', d\theta')\right) = \bar{R}\left(l + n; (x_1, \theta_1), (dx', d\theta')\right).$$

Its distribution and expectation operator are denoted respectively by $\mathbb{P}^{(l)}_{x,\theta}$ and $\mathbb{E}^{(l)}_{x,\theta}$. We will denote $\{\mathcal{F}_n,\ n \geq 0\}$ its natural filtration (for convenience in the notations, we omit its dependence on $(x, \theta, l)$). Again it follows from (1) that for any bounded measurable function $f : \mathsf{X} \to \mathbb{R}$,

$$\mathbb{E}^{(l)}_{x,\theta}\left(f(\tilde{X}_{n+1})|\mathcal{F}_n\right) = P_{\tilde{\theta}_n} f(\tilde{X}_n),\ \ \mathbb{P}^{(l)}_{x,\theta} - \text{a.s.} \tag{4}$$

For $\mathsf{K}$ a compact subset of $\Theta$, we define the stopping time $\overleftarrow{\tau}_{\mathsf{K}}$ (wrt the nonhomogeneous Markov chain $\{(\tilde{X}_n, \tilde{\theta}_n),\ n \geq 0\}$) as

$$\overleftarrow{\tau}_{\mathsf{K}} = \inf\{k \geq 1 :\ \tilde{\theta}_k \notin \mathsf{K}\},$$

with the usual convention that $\inf \emptyset = \infty$. Clearly the two processes defined above are closely related. We will refer to $\{(\tilde{X}_n, \tilde{\theta}_n),\ n \geq 0\}$ as the *re-projection free process*. The general strategy that we adopt to study the Markov chain $\{(X_n, \theta_n, \nu_n, \xi_n),\ n \geq 0\}$ (a strategy borrowed from Andrieu et al. (2005)) consists in first studying the re-projection free process $\{(\tilde{X}_n, \tilde{\theta}_n),\ n \geq 0\}$ and showing that the former process inherits the limit behavior of the latter.

2.3. **General results.** The main assumption of the paper is the following.

**A1** There exist $\alpha \in (0, 1]$, and a measurable function $V :\ \mathsf{X} \to [1, \infty)$, $\sup_{x \in \mathsf{X}_0} V(x) < \infty$ with the following properties. For any compact subset $\mathsf{K}$ of $\Theta$, there exists $b, c \in (0, \infty)$ (that depend on $\mathsf{K}$) such that for any $(x, \theta) \in \mathsf{X} \times \mathsf{K}$,

$$P_\theta V(x) \leq V(x) - cV^{1-\alpha}(x) + b \tag{5}$$

and for any $\beta \in [0, 1-\alpha]$, $\kappa \in [0, \alpha^{-1}(1-\beta) - 1]$, there exists $C = C(V, \kappa, \beta, \mathsf{K})$ such that

$$(n+1)^\kappa \left\| P_\theta^n(x, \cdot) - \pi(\cdot) \right\|_{V^\beta} \leq C\ V^{\beta + \alpha\kappa}(x), \quad n \geq 0. \tag{6}$$

Notice that (5) implies that $\pi(V^{1-\alpha}) < \infty$. We will also assume that the number of re-projection is finite.

**A2**

$$\check{\mathbb{P}}\left(\sup_{n \geq 0} \nu_n < \infty\right) = 1. \tag{7}$$

We introduce a new pseudo-metric on $\Theta$. For $\beta \in [0, 1]$, $\theta, \theta' \in \Theta$, set

$$D_\beta(\theta, \theta') \stackrel{\text{def}}{=} \sup_{|f|_{V^\beta} \leq 1} \sup_{x \in \mathsf{X}} \frac{|P_\theta f(x) - P_{\theta'} f(x)|}{V^\beta(x)}\ .$$

Under A1 and A2 a weak law of large numbers hold.

**Theorem 2.1.** *Assume A1-A2. Let $\beta \in [0, 1-\alpha)$ and $f_\theta : \mathsf{X} \to \mathbb{R}$ a family of measurable functions of $\mathcal{L}_{V^\beta}$ such that $\pi(f_\theta) = 0$, $\theta \to f_\theta(x)$ is measurable and $\sup_{\theta \in \mathsf{K}} |f_\theta|_{V^\beta} < \infty$ for any compact subset $\mathsf{K}$ of $\Theta$. Suppose also that there exist $\epsilon > 0$, $\kappa > 0$, $\beta + \alpha\kappa < 1 - \alpha$ such that for any $(x, \theta, l) \in \mathsf{X}_0 \times \Theta_0 \times \mathbb{N}$*

$$\mathbb{E}_{x,\theta}^{(l)}\left[\sum_{k \geq 1} k^{-1+\epsilon}\left(D_\beta(\tilde{\theta}_k, \tilde{\theta}_{k-1}) + |f_{\tilde{\theta}_k} - f_{\tilde{\theta}_{k-1}}|_{V^\beta}\right) \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > k\}} V^{\beta + \alpha\kappa}(\tilde{X}_k)\right] < \infty. \tag{8}$$

*Then $n^{-1} \sum_{k=1}^n f_{\theta_{k-1}}(X_k)$ converges in $\check{\mathbb{P}}$-probability to zero.*

*Proof.* The proof is given in Section 3.5. □

*Remark* 2. A strong law of large numbers also hold under similar assumptions (Atchade and Fort (2008)). It is an open problem whether A1, A2 and (8) imply a weak law of large numbers hold for measurable functions $f$ for which $\pi(|f|) < \infty$ without the additional assumption that $f \in \mathcal{L}_{V^\beta}$, $0 \le \beta < 1-\alpha$.

For the Central limit theorem, we introduce few additional notations. For $f \in \mathcal{L}_{V^\beta}$ with $\pi(f) = 0$, and $a \in [0, 1/2]$ we introduce the resolvent functions

$$g_a(x,\theta) = \sum_{j\ge 0} (1-a)^{j+1} P_\theta^j f(x).$$

Whenever $g_a$ is well defined it satisfies the *approximate Poisson equation*

$$f(x) = (1-a)^{-1} g_a(x,\theta) - P_\theta g_a(x,\theta). \tag{9}$$

When $a = 0$, we write $g(x,\theta)$ which is the usual solution to the Poisson equation $f(x) = g(x,\theta) - P_\theta g(x,\theta)$. Define also

$$H_a(x,y) = g_a(y,\theta) - P_\theta g_a(x,\theta), \tag{10}$$

where $P_\theta g_a(x,\theta) \stackrel{\text{def}}{=} \int P_\theta(x,dz) g_a(z,\theta)$. We start by showing that under A1-A2, the partial sum $\sum_{k=1}^n f(X_k)$ admits a martingale approximation.

**Theorem 2.2.** *Assume A1-A2 with $\alpha < 1/2$. Let $\beta \in [0, \frac{1}{2} - \alpha)$ and $f \in \mathcal{L}_{V^\beta}$ such that $\pi(f) = 0$. Let $\kappa > 1$, $\delta \in (0,1)$ be such that $2\beta + \alpha(\kappa+\delta) < 1-\alpha$. Take $\rho \in (\frac{1}{2}, \frac{1}{2-\delta}]$ and let $\{a_n,\ n \ge 0\}$ be any sequence of positive numbers such that $a_n \in (0, 1/2]$, $a_n \propto n^{-\rho}$. Suppose that for any $(x,\theta,b,l) \in \mathsf{X}_0 \times \Theta_0 \times [0, 1-\alpha] \times \mathbb{N}$*

$$\mathbb{E}_{x,\theta}^{(l)} \left[ \sum_{k\ge 1} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}} > k\}} k^{-1+\rho(2-\delta)} D_b(\tilde\theta_k, \tilde\theta_{k-1}) V^{2\beta+\alpha(\kappa+\delta)}(\tilde X_k) \right] < \infty. \tag{11}$$

*Then*

$$\lim_{n\to\infty} n^{-1/2} \sum_{k=1}^n \left( f(X_k) - H_{a_n,\theta_{k-1}}(X_{k-1}, X_k) \right) = 0, \quad \textit{in } \check{\mathbb{P}}\textit{-probability.}$$

*Proof.* We show in Lemma 3.8 that the same martingale approximation hold for the re-projection free process $\{(\tilde X_n, \tilde\theta_n),\ n \ge 0\}$ and this property transfers to the adaptive chain $\{(X_n, \theta_n, \nu_n, \xi_n),\ n \ge 0\}$ as a consequence of Lemma 3.12. □

The process $\sum_{j=1}^k n^{-1/2} H_{a_n,\theta_{j-1}}(X_{j-1}, X_j) \mathbb{1}_{\{\xi_{j-1}>0\}}$, $1 \le k \le n$ is a martingale array but do not satisfy a CLT in general. To derive a CLT we strengthen A2.

**A3** There exists a $\Theta$-valued random variable $\theta_\star$ such that with $\check{\mathbb{P}}$-probability one, $\{\theta_n,\ n \ge 0\}$ remains in a compact set and $\lim_{n\to\infty} D_\beta(\theta_n, \theta_\star) = 0$ for any $\beta \in [0, 1-\alpha]$.

Notice that the compact set referred to in A3 is sample path dependent.

**Theorem 2.3.** *Assume A1 and A3 with $\alpha < 1/2$. Let $\beta \in [0, \frac{1}{2} - \alpha)$, $f \in \mathcal{L}_{V^\beta}$, $\kappa, \delta, \rho$ and $\{a_n,\ n \geq 0\}$ as in Theorem 2.2. Suppose that the diminishing adaptation condition (11) hold and*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} g_{a_n}^2(X_k, \theta_{k-1}) - P_{\theta_{k-1}} g_{a_n}^2(X_k, \theta_{k-1}) = 0, \quad \text{in} \quad \check{\mathbb{P}}\text{-probability.} \tag{12}$$

*Then there exists a nonnegative random variable $\sigma_\star^2(f)$ such that $n^{-1/2}\sum_{k=1}^n f(X_k)$ converges weakly to a random variable $Z$ with characteristic function $\phi(t) = \check{\mathbb{E}}\left[\exp\left(-\frac{\sigma_\star^2(f)}{2}t^2\right)\right]$. Moreover*

$$\sigma_\star^2(f) = \int \pi(dx)\left\{2f(x)g(x,\theta_\star) - f^2(x)\right\}, \quad \check{\mathbb{P}} - a.s.$$

*Proof.* See Section 3.6. □

2.4. **On assumption (12).** Assumption (12) is needed to establish the weak law of large numbers in the CLT. When $\{X_n,\ n \geq 0\}$ is a stationary Markov chain (12) automatically hold. The proof is based on a result due to Maxwell and Woodroofe (2000). The stationarity assumption is not restrictive in the case of Harris recurrent Markov chain.

**Proposition 2.4.** *Suppose that $\{X_n,\ n \geq 0\}$ is a stationary and ergodic Markov chain with invariant distribution $\pi$ and transition kernel $P$ that satisfies (5) and (6) with $\alpha < 1/2$. Let $f \in \mathcal{L}_{V^\beta}$ with $\beta \in [0, 1/2 - \alpha)$. Then (12) hold.*

*Proof.* See Section 3.7. □

In the general adaptive case, the simplest approach to checking (12) is through appropriate moments condition.

**Proposition 2.5.** *Assume A1 and A3 with $\alpha < 1/2$. Let $\beta \in [0, \frac{1}{2} - \alpha)$, $f \in \mathcal{L}_{V^\beta}$, $\kappa, \delta, \rho$ and $\{a_n,\ n \geq 0\}$ as in Theorem 2.2. Suppose that there exists $\epsilon > 0$ such that for any $(x, \theta, l) \in \mathsf{X}_0 \times \Theta_0 \times \mathbb{N}$*

$$\sup_{n\geq 1}\ n^{-1}\mathbb{E}_{x,\theta}^{(l)}\left[\sum_{k=1}^{n} V^{2(\beta+\alpha)+\epsilon}(\tilde{X}_k)\right] < \infty. \tag{13}$$

*Then (12) hold.*

*Proof.* See Section 3.8. □

One can always check (13) if $\alpha < 1/3$ and $\beta \in [0, 1 - 3\alpha)$.

**Corollary 2.6.** *Assume A1 and A3 with $\alpha < 1/3$. Let $\beta \in [0, 1 - 3\alpha)$, $f \in \mathcal{L}_{V^\beta}$, $\kappa, \delta, \rho$ and $\{a_n,\ n \geq 0\}$ as in Theorem 2.2. Suppose that (11). Then the conclusion of Theorem 2.3 hold.*

*Proof.* If $\alpha < 1/3$ and we take $\beta \in [0, 1-3\alpha)$ then we can find $\epsilon > 0$ such that $2(\beta+\alpha)+\epsilon < 1-\alpha$ and by Proposition 3.4 (ii), Eq. (13) hold. The stated result thus follows from Proposition 2.5 and Theorem 2.3. □

2.5. **Some additional remarks on the assumptions.**

2.5.1. *On Assumption A1.* In many cases, A1 can be checked by establishing a drift and a minorization conditions. For example if uniformly over compact subsets $\mathsf{K}$ of $\Theta$, $P_\theta$ satisfies a polynomial drift condition of the form $P_\theta V \leq V - cV^{1-\alpha} + b\mathbb{1}_{\mathcal{C}}$ for some small set $\mathcal{C}$, $\alpha \in (0, 1]$ and such that the level sets of $V$ are 1-small then (5) and (6) hold. This point is thoroughly discussed in Atchade and Fort (2008) (Section 2.4 and Appendix A) and the references therein.

Assumption A1 also hold for geometrically ergodic Markov kernels and in this case we recover the CLT result of Andrieu and Moulines (2006). Indeed, suppose that uniformly over compact subsets $\mathsf{K}$ of $\Theta$, there exist $\mathcal{C} \in \mathcal{X}$, $\nu$ a probability measure on $(\mathsf{X}, \mathcal{X})$, $b, \epsilon > 0$ and $\lambda \in (0, 1)$ such that $\nu(\mathcal{C}) > 0$, $P_\theta(x, \cdot) \geq \epsilon\nu(\cdot)\mathbb{1}_{\mathcal{C}}(x)$ and $P_\theta V \leq \lambda V + b\mathbb{1}_{\mathcal{C}}$. Then for any $\alpha \in (0, 1]$, $P_\theta V \leq V - (1-\lambda)V^{1-\alpha} + b$, thus (5) hold. Moreover by explicit convergence bounds for geometrically ergodic Markov chains (see e.g. Baxendale (2005)), for any $\beta \in (0, 1]$

$$\sup_{\theta \in \mathsf{K}} \|P_\theta^n(x, \cdot) - \pi(\cdot)\|_{V^\beta} \leq C_\beta(\mathsf{K})\rho_\beta^n V^\beta(x).$$

A fortiori (6) hold. Also under the geometric drift condition, if $\beta \in [0, 1/2)$ then we can find $0 < \alpha < 1/2$ and $\epsilon > 0$ such that $2(\beta + \alpha) + \epsilon < 1$, and since $V^\delta$-moment of geometrically ergodic adaptive MCMC are bounded in $n$ for any $\delta \in [0, 1)$, we get (13). In this case and assuming (11), Theorem 2.3 yields a CLT for all functions $f \in \mathcal{L}_{V^\beta}$ with $\beta \in [0, 1/2)$ which is the same CLT obtained in Andrieu and Moulines (2006) (Theorem 8). Roughly speaking, assuming (11) at no extra cost is similar to setting $\beta = 0$ in their theorem).

2.5.2. *On assumption A2-A3.* Assumption A3 is a natural assumption to make when a CLT is sought. Whether A2 or A3 hold depends on the adaptation strategies. We show below how to check A3 when the adaptation is driven by stochastic approximation.

2.5.3. *On the diminishing adaptation conditions (8) and (11).* It is well known that adaptive MCMC can fail to converge when to so-called *diminishing adaptation* condition (which embodies the idea that one should adapt less and less with the iterations) does not hold. Here, the diminishing adaptation takes the form of conditions (8) and (11). Indeed, (8) and (11) cannot hold unless $D_\beta(\theta_n, \theta_{n-1})$ converges to zero in some sense. These conditions are not difficult to check. Typically $D_b(\theta_k, \theta_{k-1}) \leq C\gamma_k V^\eta(X_k)$ for some positive numbers $\gamma_k$ and $\eta \geq 0$. then we can check (8) or (11) using Proposition 3.5.

2.6. **Checking A3 for AMCMC driven by stochastic approximation.** Adaptive MCMC is often driven by stochastic approximation. We consider an example of stochastic approximation dynamics and show how to check A3. Let $\{\gamma_n\}$ be a sequence of positive numbers. Let $q_\theta^{(1)} : \mathsf{X} \times \mathcal{X} \to [0,1]$ and $q_\theta^{(2)} : \mathsf{X} \times \mathsf{X} \times \mathcal{X} \to [0,1]$ be two Markov kernels such that

$$\int q_\theta^{(1)}(x,dy) q_\theta^{(2)}\left((x,y),dx'\right) = P_\theta(x,dx').$$

Let $\Phi : \ \Theta \times \mathsf{X} \times \mathsf{X} \to \Theta$ be a measurable function. For convenience we write $\Phi_\theta(x,y)$ instead of $\Phi(\theta,x,y)$. We consider the adaptive MCMC algorithm with the kernels $\bar{R}$ are given as

$$\bar{R}\left(n;(x,\theta),(dx',d\theta')\right) = \int q_\theta^{(1)}(x,dy) q_\theta^{(2)}\left((x,y),dx'\right)\delta_{\theta+\gamma_n\Phi_\theta(x,y)}(d\theta'). \tag{14}$$

Under (14), the dynamics on $\theta_n$ in algorithm 2.1 can then be written as

$$\theta_{n+1} = \theta_n + \gamma_{\nu_n+\xi_n}\left(h(\theta_n) + \epsilon_{n+1}^{(1)} + \epsilon_{n+1}^{(2)}\right), \quad \text{on } \{\xi_n > 0\}, \quad \check{\mathbb{P}} - a.s.$$

where $\epsilon_{n+1}^{(1)} = \Upsilon_{\theta_n}(X_n) - h(\theta_n)$, $\epsilon_{n+1}^{(2)} = \Phi_{\theta_n}(X_n, Y_{n+1}) - \Upsilon_{\theta_n}(X_n)$, where $Y_{n+1}$ is a random variable with conditional distribution $q_{\theta_n}^{(1)}(X_n,\cdot)$ given $\check{\mathcal{F}}_n$ and where

$$\Upsilon_\theta(x) = \int q_\theta^{(1)}(x,dy)\Phi_\theta(x,y), \quad \text{and} \quad h(\theta) = \int \pi(dx)\Upsilon_\theta(x).$$

Following Andrieu et al. (2005), we assume that

**B1** (1) $\{\mathsf{K}_n, n \geq 0\}$ is such that $\Theta = \bigcup_n \mathsf{K}_n$, $\Theta_0 \subseteq \mathsf{K}_0$ and $\mathsf{K}_n \subset \text{int}(\mathsf{K}_{n+1})$, where $\text{int}(A)$ is the interior of $A$.

(2) The function $h$ is a continuous function and there exists a continuously differentiable function $w : \ \Theta \to [0,\infty)$ such that

(a) for any $\theta \in \Theta$, $\langle \nabla w(\theta), h(\theta)\rangle \leq 0$, the set $\mathcal{L} \stackrel{\text{def}}{=} \{\theta \in \Theta : \ \langle \nabla w(\theta), h(\theta)\rangle = 0\}$ is non-empty and the closure of $w(\mathcal{L})$ has an empty interior.

(b) there exists $M_0 > 0$ such that $\mathcal{L} \cup \Theta_0 \subset \{\theta : \ w(\theta) < M_0\}$ and for any $M \geq M_0$, $\mathcal{W}_M \stackrel{\text{def}}{=} \{\theta : \ w(\theta) \leq M\}$ is a compact set.

For integers $p \geq 0$, $n \geq 1$ and a compact subset $\mathsf{K}$ of $\Theta$, we define the random variable

$$C_{n,p}(\mathsf{K}) \stackrel{\text{def}}{=} \sup_{l \geq n} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}} > l\}} \left| \sum_{j=n}^{l} \gamma_{p+j-1}\left(\tilde{\epsilon}_j^{(1)} + \tilde{\epsilon}_j^{(2)}\right)\right|,$$

where $\tilde{\epsilon}_{n+1}^{(1)} = \Upsilon_{\tilde{\theta}_n}(\tilde{X}_n) - h(\tilde{\theta}_n)$ and $\tilde{\epsilon}_{n+1}^{(2)} = \Phi_{\tilde{\theta}_n}\left(\tilde{X}_n, \tilde{Y}_{n+1}\right) - \int q_{\tilde{\theta}_n}^{(1)}(\tilde{X}_n, dy)\Phi_{\tilde{\theta}_n}\left(\tilde{X}_n, y\right)$ and where the conditional distribution of $\tilde{Y}_{n+1}$ given $\mathcal{F}_n$ is $q_{\tilde{\theta}_n}^{(1)}(\tilde{X}_n,\cdot)$.

$C_{n,p}(\mathsf{K})$ is the magnitude of the errors in the stochastic approximation. Notice that $C_{n,p}(\mathsf{K})$ is defined from the re-projection free process. A key result shown by Andrieu et al. (2005) is that when B1 hold, the convergence of a SA algorithm depends mainly on $C_{n,p}(\mathsf{K})$. The framework considered here is slightly different from Andrieu et al. (2005)

but the result still hold. The proof follows the same lines as in Andrieu et al. (2005) and we omit the details.

**Proposition 2.7.** *Assume (14), B1,* $\lim_n \gamma_n = 0$ *and* $\sum_n \gamma_n = \infty$. *Suppose that for any* $M > 0$ *large enough and for any* $\delta > 0$

$$\lim_{p\to\infty} \sup_{(x,\theta)\in \mathsf{X}_0\times\Theta_0} \mathbb{P}^{(p)}_{x,\theta}\left(C_{1,p}(\mathcal{W}_M) > \delta\right) = 0, \tag{15}$$

*and for any* $p \geq 0$,

$$\lim_{n\to\infty} \sup_{(x,\theta)\in \mathsf{X}_0\times\Theta_0} \mathbb{P}^{(p)}_{x,\theta}\left(C_{n,p}(\mathsf{K}_p) > \delta\right) = 0. \tag{16}$$

*Then A3 hold.*

We now show that (15)-(16) hold true under A1.

Assume that the function $\Upsilon$ satisfies

**B2** There exists $\eta \geq 0$, $2(\eta + \alpha) < 1$ such that for any compact subset $\mathsf{K}$ of $\Theta$, $b \in [0, 1-\alpha]$, $\theta, \theta' \in \mathsf{K}$,

$$\sup_{\theta\in\mathsf{K}} \sup_{x\in\mathsf{X}} V^{-2\eta}(x) \int q^{(1)}_\theta(x,y) \left|\Phi_\theta(x,y)\right|^2 < \infty, \quad \text{and} \quad D_b(\theta,\theta') + |\Upsilon_\theta - \Upsilon_{\theta'}|_{V^\eta} \leq C|\theta-\theta'|, \tag{17}$$

for some finite constant $C$ that depends possibly on $\mathsf{K}$.

**Proposition 2.8.** *Assume A1 with* $\alpha < 1/2$ *and (14). Suppose that B1 and B2 hold. Suppose also that* $\lim_n \gamma_n = 0$ *and* $\sum_n \gamma_n = \infty$ *and for any* $p \geq 0$,

$$\lim_{n\to\infty} (\gamma_{p+n-1} - \gamma_{p+n}) n^{1-\alpha} = 0 \quad \text{and} \quad \sum_{n\geq 1} \left(\gamma_k^2 k^\rho + \gamma_k k^{-\rho} + \gamma_k^{1+\rho}\right) < \infty, \tag{18}$$

*for some* $\rho \in \left(0, (1-\alpha)(\eta+\alpha)^{-1} - 1\right)$. *Then A3 hold.*

*Proof.* See Section 3.9. □

2.7. **Example: Adaptive Langevin algorithms.** We illustrate the theory above with an application to the Metropolis-adjusted Langevin algorithm (MALA). In this section, $\mathsf{X}$ is the $d$-dimensional Euclidean space $\mathbb{R}^d$ and $\pi$ is a positive density on $\mathsf{X}$ with respect to the Lebesgue (denoted $\mu_{Leb}$ or $dx$). The MALA algorithm is an effective Metropolis-Hastings algorithm whose proposal kernel is obtained by discretization of the Langevin diffusion

$$dX_t = \frac{1}{2} e^\theta \nabla \log \pi(X_t) dt + e^\theta dB_t, \quad X_0 = x,$$

where $\theta \in \mathbb{R}$ is a scale parameter and $\{B_t,\ t \geq 0\}$ a $d$-dimensional standard Brownian motion. Denote $q_\theta(x,y)$ the density of the $d$-dimensional Gaussian distribution with mean $b_\theta(x)$ and covariance matrix $e^\theta I_d$ where

$$b_\theta(x) = x + \frac{1}{2} e^\theta \nabla \log \pi(x).$$

The MALA works as follows. Given $X_n = x$, we propose a new value $Y \sim q_\theta(x, \cdot)$. Then with probability $\alpha_\theta(X_n, Y)$, we 'accept $Y$' and set $X_{n+1} = Y$ and with probability $1 - \alpha_\theta(X_n, Y)$, we 'reject $Y$' and set $X_{n+1} = X_n$. The acceptance probability is given by

$$\alpha_\theta(x, y) = 1 \wedge \frac{\pi(y) q_\theta(y, x)}{\pi(x) q_\theta(x, y)}.$$

The convergence and optimal scaling of MALA is studied in detail in Roberts and Tweedie (1996); Roberts and Rosenthal (2001). In practice the performance of this algorithm depends on the choice of the scale parameter $\theta$. In high-dimensional spaces (and under some regularity conditions) it is optimal to set $\theta = \theta_\star$ such that the average acceptance probability of the algorithm in stationarity is 0.574. In general, $\theta_\star$ is not available and its computation would require a tedious fine-tuning of the sampler. Adaptive MCMC provides a straightforward approach to properly scale the algorithm.

The parameter space is $\Theta = \mathbb{R}$. For $\theta \in \Theta$, denote $P_\theta$ the transition kernel of the MALA algorithm with proposal $q_\theta$. We also introduce the functions

$$A_\theta(x) \stackrel{\text{def}}{=} \int_{\mathsf{X}} \alpha_\theta(x, y) q_\theta(x, y) \mu_{Leb}(dy), \quad a(\theta) \stackrel{\text{def}}{=} \int_{\mathsf{X}} A_\theta(x) \pi(x) \mu_{Leb}(dx).$$

Let $\{\mathsf{K}_n,\ n \geq 0\}$ be a family of nonempty compact intervals of $\Theta$ such that $\cup \mathsf{K}_n = \mathbb{R}$, $\mathsf{K}_n \subset \mathrm{int}(\mathsf{K}_{n+1})$. Therefore by construction B1-(1) hold. Let $\Theta_0 = \{\theta_0\}$ and $\mathsf{X}_0 = \{x_0\}$ for some arbitrary point $(x_0, \theta_0) \in \mathsf{X} \times \mathsf{K}_0$. The re-projection function is $\Pi(x, \theta) = (x_0, \theta_0)$ for any $(x, \theta) \in \mathsf{X} \times \Theta$. We also have $\Pi_k(x, \theta) = (x, \theta)$ if $k > 0$ and $\Pi_k(x, \theta) = \Pi(x, \theta)$ if $k = 0$. Obviously many other choices are possible. The adaptive MALA we consider is the following.

**Algorithm 2.2.** **Initialization:** *Let $\bar{\alpha}$ be the target acceptance probability (taken as 0.574). Choose $(X_0, \theta_0) \in \mathsf{X}_0 \times \Theta_0$, $\nu_0 = 0$ and $\xi_0 = 0$.*

**Iteration:** *Given $(X_n, \theta_n, \nu_n, \xi_n)$: set $(\bar{X}, \bar{\theta}) = \Pi_{\xi_n}(X_n, \theta_n)$.*

- **a:** *generate $Y_{n+1} \sim q_{\bar{\theta}}\left(\bar{X}, \cdot\right)$. With probability $\alpha_{\bar{\theta}}(\bar{X}, Y_{n+1})$, set $X_{n+1} = Y_{n+1}$ and with probability $1 - \alpha_{\bar{\theta}}(\bar{X}, Y_{n+1})$, set $X_{n+1} = \bar{X}$.*
- **b:** *Compute*

$$\theta_{n+1} = \bar{\theta} + \frac{1}{1 + \nu_n + \xi_n} \left(\alpha_{\bar{\theta}}(\bar{X}, Y_{n+1}) - \bar{\alpha}\right). \tag{19}$$

- **c:** *If $\theta_{n+1} \in \mathsf{K}_{\nu_n}$ then set $\nu_{n+1} = \nu_n$ and $\xi_{n+1} = \xi_n + 1$. Otherwise if $\theta_{n+1} \notin \mathsf{K}_{\nu_n}$ then set $\nu_{n+1} = \nu_n + 1$ and $\xi_{n+1} = 0$.*

In this algorithm, the kernel $\bar{R}(n; \cdot, \cdot)$ takes the form

$$\bar{R}\left(n; (x, \theta), (dx', d\theta')\right) = \int q_\theta(x, dy) \left(\alpha_\theta(x, y) \delta_y(dx') \right. \\ \left. + (1 - \alpha_\theta(x, y)) \delta_x(dx')\right) \delta_{\Phi_n(\theta, x, y)}(d\theta'),$$

where $\Phi_n(\theta, x, y) = \theta + (n+1)^{-1}(\alpha_\theta(x,y) - \bar{\alpha})$. Thus (14) hold. We make the following assumption.

**C1** $\bar{\alpha} \in (0,1)$, $\lim_{\theta\to+\infty} a(\theta) = 0$, $\lim_{\theta\to-\infty} a(\theta) = 1$.

**Proposition 2.9.** *Under C1, the function $h(\theta) = a(\theta) - \bar{\alpha}$ satisfies B1-(2) with $\mathcal{L} = \{\theta \in \mathbb{R} :\ a(\theta) = \bar{\alpha}\}$ and $w(\theta) = \int_0^\theta \cosh(u)(\bar{\alpha} - a(u))du + K$ for some finite constant $K$ where $\cosh(u) = (e^u + e^{-u})/2$ is the hyperbolic cosine.*

*Proof.* See Section 3.10.1. □

We assume that the target density $\pi$ is heavy tailed as in Kamatani (To appear).

**C2** We assume that $\pi : \mathbb{R}^d \to (0, \infty)$ is of class $\mathcal{C}^2$ and there exists $\eta > d$ such that

$$\limsup_{|x|\to\infty} \langle x, \nabla \log \pi(x) \rangle \le -\eta, \quad \lim_{|x|\to\infty} |\nabla \log \pi(x)| = 0, \quad \lim_{|x|\to\infty} \|\nabla^2 \log \pi(x)\| = 0, \tag{20}$$

where for a matrix $A$, $\|A\|$ denotes its Frobenius norm.

The next proposition is a paraphrase of Theorem 5 of Kamatani (To appear).

**Proposition 2.10.** *Assume C2. For $s \in (2, 2+\eta-d)$, define $V_s(x) = \left(1+|x|^2\right)^{s/2}$ and $\alpha = 2/s$. Let $\mathcal{C}$ be a compact subset of $\mathbb{R}^d$ with $\mu_{Leb}(\mathcal{C}) > 0$. For any compact subset $\mathsf{K}$ of $\Theta$, there exists $\epsilon, c, b \in (0,\infty)$, such that*

$$\inf_{\theta\in\mathsf{K}} P_\theta(x, dy) \ge \epsilon \left[\frac{\mu_{Leb}(dy)\mathbb{1}_{\mathcal{C}}(y)}{\mu_{Leb}(\mathcal{C})}\right] \mathbb{1}_{\mathcal{C}}(x),$$

$$\sup_{\theta\in\mathsf{K}} P_\theta V_s(x) \le V_s(x) - cV^{1-\alpha}(x) + b\mathbb{1}_{\mathcal{C}}(x).$$

For the smoothness we have

**Proposition 2.11.** *Assume that $|\nabla \log \pi(x)|$ is a bounded function. Let $\mathsf{K}$ be a compact convex subset of $\Theta$. There exists a finite constant $C(\mathsf{K})$ such that for any $f \in \mathcal{L}_{V_s^\beta}$, $\beta \in [0,1]$, any $\theta, \theta' \in \mathsf{K}$,*

$$\left|\int \alpha_\theta(x,y)q_\theta(x,y)f(y)dy - \int \alpha_{\theta'}(x,y)q_{\theta'}(x,y)f(y)dy\right| \le C(\mathsf{K})|f|_{V_s^\beta}|\theta-\theta'|V_s^\beta(x). \tag{21}$$

*Proof.* See Section 3.10.2. □

We now apply Theorem 2.3 to get a CLT for the adaptive MALA.

**Theorem 2.12.** *Assume C1 and C2 with $\eta > d+4$. Let $s \in (6, 2+\eta-d)$ and let $f :\ \mathsf{X} \to \mathbb{R}$ be a measurable function such that $\pi(f) = 0$ and $|f(x)| \le C(1+|x|^2)^b$ for some $b \in [0, \frac{s}{2}-3)$ and some finite constant $C$. Then there exists a nonnegative random variable $\sigma^2_\star(f)$ such that $n^{-1/2}\sum_{k=1}^n f(X_k)$ converges weakly to a random variable $Z$ with characteristic function $\phi(t) = \check{\mathbb{E}}\left[\exp\left(-\frac{\sigma^2_\star(f)}{2}t^2\right)\right]$.*

*Remark* 3. If $\pi$ is positive and of class $\mathcal{C}^2$ and $\pi(x) \approx (1+|x|^2)^{-(d+\nu)/2}$ in the tails, then $C2$ hold with $\eta = \nu + d$ and Theorem 2.12 guarantees a CLT for $\nu > 4$. Compare with $\nu > 2$ for Harris recurrent Markov chains satisfying A1.

*Proof.* A1 hold as a consequence of Proposition 2.10 (see e.g. Atchade and Fort (2008) Section 2.4 and Appendix A). Proposition 2.9 shows that B1-(2) hold and Proposition 2.11 implies that B2 hold. Therefore A3 hold as a consequence of Proposition 2.8. (11) is an easy consequence of Proposition 2.11 and Proposition 3.5. We thus conclude with Corollary 2.4. □

In the above theorem the asymptotic variance $\sigma^2_\star(f)$ takes values in the set $\{\sigma^2_\theta(f),\ \theta \in \mathcal{L}\}$, where $\mathcal{L} = \{\theta \in \mathbb{R} :\ a(\theta) = \bar{\alpha}\}$ and

$$\sigma^2_\theta(f) \stackrel{\text{def}}{=} \int \pi(dx) \left\{ f^2(x) + 2\sum_{k\geq 0} f(x) P^k_\theta f(x) \right\}.$$

In particular, if $\mathcal{L} = \{\theta_\star\}$ and $\sigma^2_{\theta_\star}(f) > 0$, then $n^{-1/2}\sum_{k=1}^n f(X_k)$ converges weakly to $\mathcal{N}\left(0, \sigma^2_{\theta_\star}(f)\right)$.

## 3. Proofs

The proofs are organized as follows. The weak law of large numbers (Theorem 2.1) is proved in Section 3.5, the CLT (Theorem 2.3) is proved in Section 3.6. In Section 3.1 we develop some preliminary results on the resolvent functions $g_a$ and we establish some basic results on the asymptotic behavior of the nonhomogeneous process $\{(\tilde{X}_n, \tilde{\theta}_n),\ n \geq 0\}$ in Section 3.2-3.3. The results in Section 3.4 (in particular Lemma 3.12) serve as a link and allow us to reduce the limiting behavior of the adaptive algorithm $\{(X_n, \theta_n, \nu_n, \xi_n),\ n \geq 0\}$ to that of the nonhomogeneous Markov chain $\{(\tilde{X}_n, \tilde{\theta}_n),\ n \geq 0\}$.

Throughout the proof, $C(\mathsf{K})$ denotes a finite constant that depends on the compact set $\mathsf{K}$ and on the constants in the above assumptions. But to simplify the notations, we will not keep track of these constants so the actual value of $C(\mathsf{K})$ might be different from one appearance to the next.

3.1. **Resolvent kernels and approximate Poisson's equations.** In this section, $\mathsf{K}$ is a given compact subset of $\Theta$ and $\beta \in [0, 1-\alpha]$. We consider a family of functions $f_\theta \in \mathcal{L}_{V^\beta}$, $\theta \in \Theta$ such that $\pi(f_\theta) = 0$. For $a \in (0,1)$ we define the resolvent function associated with $f_\theta$ as

$$\tilde{g}_a(x,\theta) = \sum_{j=0}^{\infty} (1-a)^{j+1} P^j_\theta f_\theta(x) = \sum_{j=0}^{\infty} (1-a)^{j+1} \bar{P}^j_\theta f_\theta(x),$$

where $\bar{P}_\theta = P_\theta - \pi$. Similarly we define

$$\tilde{g}(x,\theta) = \sum_{j=0}^{\infty} P_\theta^j f_\theta(x) = \sum_{j=0}^{\infty} \bar{P}_\theta^j f_\theta(x),$$

When $f_\theta \equiv f$ does not depend on $\theta \in \Theta$, and to help keep the notation clear, we write $g_a(x,\theta)$ (resp. $g(x,\theta)$) instead of $\tilde{g}_a(x,\theta)$ (resp. $\tilde{g}$). It is easy to see that when $\tilde{g}_a$ is well defined, it satisfies the following approximate Poisson equation

$$f_\theta(x) = (1-a)^{-1}\tilde{g}_a(x,\theta) - P_\theta \tilde{g}_a(x,\theta). \tag{22}$$

Similarly $\tilde{g}$, when well-defined, satisfies the Poisson equation

$$f_\theta(x) = \tilde{g}(x,\theta) - P_\theta \tilde{g}(x,\theta). \tag{23}$$

We introduce the function

$$\zeta_\kappa(a) \stackrel{\text{def}}{=} \sum_{j\geq 0} (1-a)^{j+1}(1+j)^{-\kappa}.$$

We will need the following lemma.

**Lemma 3.1.** *For any $a \in (0,1/2]$ and $\kappa \geq 0$,*

$$\zeta_\kappa(a) \leq \begin{cases} \sum_{k\geq 1} k^{-\kappa} & \text{if } \kappa > 1 \\ -\log(2a) + 1 & \text{if } \kappa = 1 \\ 2^{-1+\kappa}\Gamma(1-\kappa)a^{-1+\kappa} & \text{if } 0 \leq \kappa < 1 \end{cases},$$

*where $\Gamma(x) := \int_0^\infty u^{x-1}e^{-u}du$ is the Gamma function.*

*Proof.* $(1-a)^j \leq 1$ for all $j \geq 1$. Therefore, for $\kappa > 1$, $\sum_{j\geq 0}(1-a)^{j+1}(1+j)^{-\kappa} \leq \sum_{j\geq 1} j^{-\kappa}$. For $\kappa = 1$, we note that $\frac{d}{d\,a}\left\{\sum_{j\geq 0}(1-a)^{j+1}(1+j)^{-\kappa}\right\} = -a^{-1}$. Therefore for $a \in (0,1/2]$, $\sum_{j\geq 0}(1-a)^{j+1}(1+j)^{-\kappa} = \sum_{j\geq 1}(j2^j)^{-1} - \log(2a) \leq -\log(2a) + 1$. Finally, if $0 \leq \kappa < 1$, by monotonicity, $\sum_{j\geq 0}(1-a)^{j+1}(1+j)^{-\kappa} \leq \int_0^\infty (1-a)^x x^{-\kappa} dx = \int_0^\infty x^{1-\kappa-1}e^{-\beta x}dx = \Gamma(1-\kappa)\beta^{-1+\kappa}$, where $\beta = -\log(1-a)$. For $a \in (0,1/2]$, $-\log(1-a) \leq 2a$ and we conclude that $\sum_{j\geq 0}(1-a)^{j+1}(1+j)^{-\kappa} \leq 2^{-1+\kappa}\Gamma(1-\kappa)a^{-1+\kappa}$. □

**Proposition 3.2.** *Assume A1.*

**(i):** *Let $\kappa \in [0, \alpha^{-1}(1-\beta)-1]$. There exists a finite constant $C(\mathsf{K})$ such that for any $(x,\theta) \in \mathsf{X}\times\mathsf{K}$ and any $a \in (0,1/2]$*

$$|\tilde{g}_a(x,\theta)| \leq C(\mathsf{K})|f_\theta|_{V^\beta}\zeta_\kappa(a)V^{\beta+\alpha\kappa}(x). \tag{24}$$

**(ii):** *Suppose that $\alpha < 1/2$. Let $\kappa \in (1, \alpha^{-1}(1-\beta)-1]$. There exists a finite constant $C(\mathsf{K})$ such that for any $(x,\theta) \in \mathsf{X}\times\mathsf{K}$ and any $a \in (0,1/2]$*

$$|\tilde{g}_a(x,\theta) - \tilde{g}(x,\theta)| \leq C(\mathsf{K})|f_\theta|_{V^\beta}\left(2^{1-\kappa}\int_0^a \zeta_{\kappa-1}(u)du\right)V^{\beta+\alpha\kappa}(x). \tag{25}$$

*Proof.* (i) is a direct consequence of (6).

To prove (ii), we note the identity $1-(1-a)^{j+1} = (j+1)\int_0^a (1-u)^j du$ and then write

$$\begin{aligned}
|\tilde{g}(x,\theta) - \tilde{g}_a(x,\theta)| &\leq \sum_{j\geq 1} \left(1-(1-a)^{j+1}\right) \left|P_\theta^j f_\theta(x)\right| \\
&\leq C(\mathsf{K})|f_\theta|_{V^\beta} V^{\beta+\alpha\kappa}(x) \sum_{j\geq 1} \int_0^a (1-u)^j du (1+j)^{-\kappa+1} \\
&= C(\mathsf{K})|f_\theta|_{V^\beta} V^{\beta+\alpha\kappa}(x) \int_0^a \left\{ \sum_{j\geq 1} (1-u)^j (1+j)^{-\kappa+1} \right\} du \\
&\leq C(\mathsf{K})|f_\theta|_{V^\beta} V^{\beta+\alpha\kappa}(x) 2^{1-\kappa} \int_0^a \zeta_{\kappa-1}(u) du.
\end{aligned}$$

Since $\kappa > 1$ and $a > 0$, the interchange of the summation and integral signs is permitted.

□

*Remark* 4. One can check using Lemma 3.1 that for $\kappa > 1$, $\int_0^a \zeta_{\kappa-1}(u)du \to 0$ as $a \to 0$. Hence a direct consequence of Proposition 3.2 is that for any $\beta \in [0, 1-2\alpha)$ ($\alpha < 1/2$), any $\kappa \in (1, \alpha^{-1}(1-\beta)-1]$, there exists a finite constant $C(\mathsf{K})$ such that for any $(x,\theta) \in \mathsf{X}\times\mathsf{K}$,

$$|\tilde{g}(x,\theta)| \leq C(\mathsf{K})|f_\theta|_{V^\beta} V^{\beta+\alpha\kappa}(x). \tag{26}$$

**Proposition 3.3.** *Assume A1.*

**(i):** *For any $\kappa, \delta \geq 0$ with $\kappa+\delta \leq \alpha^{-1}(1-\beta)-1$, there exists a finite constant $C(\mathsf{K})$ such that for any $\theta, \theta' \in \mathsf{K}$, $x \in \mathsf{X}$ and $a \in (0, 1/2]$*

$$\left|\tilde{g}_a(x,\theta) - \tilde{g}_a(x,\theta')\right| \leq C(\mathsf{K}) \sup_{\theta\in\mathsf{K}} |f_\theta|_{V^\beta} \zeta_\kappa(a) \left(\zeta_\delta(a) D_{\beta+\alpha\delta}(\theta,\theta') + |f_\theta - f_{\theta'}|_{V^\beta}\right) V^{\beta+\alpha(\kappa+\delta)}(x).$$

**(ii):** *Assume $\alpha < 1/2$. For any $\beta \in [0, 1-2\alpha)$, any $\kappa \geq 0$, $\delta > 1$ with $\kappa+\delta \leq \alpha^{-1}(1-\beta)-1$, There exist a finite constant $C(\mathsf{K})$ such that for any $x \in \mathsf{X}$, $\theta, \theta' \in \mathsf{K}$ and any $a \in (0, 1/2]$*

$$\begin{aligned}
\left|\tilde{g}(x,\theta) - \tilde{g}(x,\theta')\right| \leq C(\mathsf{K}) \sup_{\theta\in\mathsf{K}} |f_\theta|_{V^\beta} \Big( \int_0^a \zeta_{\delta-1}(u)du + \zeta_\kappa(a)|f_\theta - f_{\theta'}|_{V^\beta} \\
+ \zeta_\kappa(a) D_{\beta+\alpha\delta}(\theta,\theta')\Big) V^{\beta+\alpha(\kappa+\delta)}(x).
\end{aligned}$$

*Proof.* We have

$$\tilde{g}_a(x,\theta) - \tilde{g}_a(x,\theta') = \sum_{j\geq 0} (1-a)^{j+1} \left(\bar{P}_\theta^j f_\theta(x) - \bar{P}_{\theta'}^j f_\theta(x)\right) + \sum_{j\geq 0} (1-a)^{j+1} \bar{P}_{\theta'}^j \left(f_\theta(x) - f_{\theta'}(x)\right).$$

By Proposition 3.2 (i) we bound the second term in the rhs as follows.

$$\left|\sum_{j\geq 0} (1-a)^{j+1} \bar{P}_{\theta'}^j \left(f_\theta(x) - f_{\theta'}(x)\right)\right| \leq C(\mathsf{K})|f_\theta - f_{\theta'}|_{V^\beta} \zeta_\kappa(a) V^{\beta+\alpha\kappa}(x). \tag{27}$$

The first term in the rhs can be rewritten as

$$\sum_{j\geq 0}(1-a)^{j+1}\left(\bar{P}^j_\theta f_\theta(x)-\bar{P}^j_{\theta'}f_\theta(x)\right) = \sum_{j\geq 1}(1-a)^{j+1}\sum_{l=0}^{j-1}\bar{P}^l_\theta(\bar{P}_\theta-\bar{P}_{\theta'})\bar{P}^{j-l-1}_{\theta'}f_\theta(x).$$

From (6) of A2 with $\kappa=\delta$, we have $|\bar{P}^l_{\theta'}f_\theta(x)|\leq C(\mathsf{K})\sup_{\theta\in\mathsf{K}}|f_\theta|_{V^\beta}(1+l)^{-\delta}V^{\beta+\alpha\delta}(x)$ for all $l\geq 0$. Combined with the definition of $D_{\beta+\alpha\delta}$, we get

$$\left|(\bar{P}_\theta-\bar{P}_{\theta'})\bar{P}^{j-l-1}_{\theta'}f_\theta(x)\right|\leq C(\mathsf{K})\sup_{\theta\in\mathsf{K}}|f_\theta|_{V^\beta}(j-l)^{-\delta}D_{\beta+\alpha\delta}(\theta,\theta')V^{\beta+\alpha\delta}(x).$$

Another application of A2-(6) then yields for any $\kappa\in[0,\alpha^{-1}(1-\beta)-1-\delta]$

$$\left|\bar{P}^j_\theta f_\theta(x)-\bar{P}^j_{\theta'}f_\theta(x)\right|\leq C(\mathsf{K})\sup_{\theta\in\mathsf{K}}|f_\theta|_{V^\beta}D_{\beta+\alpha\delta}(\theta,\theta')V^{\beta+\alpha(\kappa+\delta)}(x)\sum_{l=0}^{j-1}(1+l)^{-\kappa}(j-l)^{-\delta}.$$

It follows that

$$\begin{aligned}&\left|\sum_{j\geq 0}(1-a)^{j+1}\left(\bar{P}^j_\theta f_\theta(x)-\bar{P}^j_{\theta'}f_\theta(x)\right)\right|\\ &\quad\leq C(\mathsf{K})\sup_{\theta\in\mathsf{K}}|f_\theta|_{V^\beta}D_{\beta+\alpha\delta}(\theta,\theta')V^{\beta+\alpha(\kappa+\delta)}(x)\sum_{j\geq 1}(1-a)^{j+1}\sum_{l=0}^{j-1}(1+l)^{-\kappa}(j-l)^{-\delta}\\ &\qquad\leq C(\mathsf{K})\sup_{\theta\in\mathsf{K}}|f_\theta|_{V^\beta}\zeta_\kappa(a)\zeta_\delta(a)D_{\beta+\alpha\delta}(\theta,\theta')V^{\beta+\alpha(\kappa+\delta)}(x).\end{aligned}$$

Combining this with (27) gives part (i).

To prove (ii), we write $|\tilde{g}(x,\theta)-\tilde{g}(x,\theta')|\leq|\tilde{g}_a(x,\theta)-\tilde{g}(x,\theta)|+|\tilde{g}_a(x,\theta)-\tilde{g}_a(x,\theta')|+|\tilde{g}_a(x,\theta')-\tilde{g}(x,\theta')|$. Part (i) gives

$$|\tilde{g}_a(x,\theta)-\tilde{g}_a(x,\theta')|\leq C(\mathsf{K})\sup_{\theta\in\mathsf{K}}|f_\theta|_{V^\beta}\zeta_\kappa(a)\left(D_{\beta+\alpha\delta}(\theta,\theta')+|f_\theta-f_{\theta'}|_{V^\beta}\right)V^{\beta+\alpha(\delta+\kappa)}(x).$$

Then we use $\delta>1$ and Part (ii) of Proposition 24, to get

$$|\tilde{g}_a(x,\theta)-\tilde{g}(x,\theta)|+|\tilde{g}_a(x,\theta')-\tilde{g}(x,\theta')|\leq C(\mathsf{K})\sup_{\theta\in\mathsf{K}}|f_\theta|_{V^\beta}\int_0^a\zeta_{\delta-1}(u)du V^{\beta+\alpha\delta}(x).$$

The conclusion follows. □

3.2. **Modulated moments.** In this section, $\mathsf{K}$ is an arbitrary compact subset of $\Theta$, $(x,\theta)\in\mathsf{X}\times\mathsf{K}$ and $l\geq 0$ an integer. We consider the nonhomogeneous Markov chain $\{(\tilde{X}_n,\tilde{\theta}_n),\ n\geq 0\}$ with initial distribution $\delta_{x,\theta}$ and transition kernels $P_l\left(n;(x_1,\theta_1),(dx',d\theta')\right)=\bar{R}\left(l+n;(x_1,\theta_1),(dx',d\theta')\right)$. Its distribution and expectation operator are denoted respectively by $\mathbb{P}^{(l)}_{x,\theta}$ and $\mathbb{E}^{(l)}_{x,\theta}$. The key property that we will use here is (4) which, as we have seen, is a consequence of (1). The first two propositions below are easy modifications of similar results proved in Atchade and Fort (2008).

**Proposition 3.4.** *Assume A1. There exists a finite constant $C(\mathsf{K})$ such that for any $(x,\theta)\in\mathsf{X}\times\mathsf{K}$, $l,n\geq 1$,* *(i) for any* $0 \leq \beta \leq 1$,

$$\mathbb{E}^{(l)}_{x,\theta}\left[V^{\beta}(\tilde{X}_n)\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>n-1\}}\right] \leq C(\mathsf{K})n^{\beta}\, V^{\beta}(x)\;.$$

*(ii) for any* $0 \leq \beta \leq 1-\alpha$

$$\mathbb{E}^{(l)}_{x,\theta}\left[\sum_{k=1}^{n} V^{\beta}(\tilde{X}_k)\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>k-1\}}\right] \leq C(\mathsf{K})n\, V^{\beta+\alpha}(x)\;.$$

**Proposition 3.5.** *Assume A1. Let* $\{r_n,\ n \geq 0\}$ *be a non-increasing sequence of positive numbers. For* $\beta \in [0, 1-\alpha]$*, there exists a finite constant* $C(\mathsf{K})$ *such that for any* $(x,\theta) \in \mathsf{X}\times\mathsf{K}$, $1 \leq n < N$

$$\mathbb{E}^{(l)}_{x,\theta}\left[\sum_{k=n}^{N-1} r_{k+1}V^{\beta}(\tilde{X}_k)\mathbb{1}_{\{\overleftarrow{\tau}_{K}>k-1\}}\right] \leq C(\mathsf{K})\left(r_n\mathbb{E}^{(l)}_{x,\theta}\left(V^{\beta+\alpha}(\tilde{X}_n)\mathbb{1}_{\{\overleftarrow{\tau}_{K}>n-1\}}\right)+\sum_{k=n}^{N} r_{k+1}\right).$$

The next proposition gives a general standard bound on moments of martingales as a consequence of the Burkholder's inequality.

**Proposition 3.6.** *Let* $M_n = \sum_{k=1}^{n} D_k$, $n \geq 1$ *be a martingale such that* $\mathbb{E}\left(|D_k|^p\right) < \infty$ *for some* $p > 1$*. Then*

$$\mathbb{E}\left[|M_n|^p\right] \leq Cn^{\max(1,p/2)-1}\sum_{k=1}^{n}\mathbb{E}\left(|D_k|^p\right),$$

*for* $C = (18pq^{1/2})^p$, $p^{-1}+q^{-1} = 1$.

3.3. **A Weak law of large numbers.** We fix $l \geq 0$ integer, $\mathsf{K}$ a compact subset of $\Theta$ and $(x,\theta) \in \mathsf{X}\times\mathsf{K}$. This section deals with the weak law of large numbers for the non-homogeneous Markov chain $\{(\tilde{X}_n,\tilde{\theta}_n),\ n\geq 0\}$ with initial distribution $\delta_{x,\theta}$ and transition kernels $P_l\left(n;(x_1,\theta_1),(dx',d\theta')\right) = \bar{R}\left(l+n;(x_1,\theta_1),(dx',d\theta')\right)$.

**Proposition 3.7.** *Assume A1. Let* $\beta \in [0, 1-\alpha)$ *and* $f_\theta \in \mathcal{L}_{V^\beta}$ *a class of functions such that* $\theta \to f_\theta(x)$ *is a measurable map,* $\pi(f_\theta) = 0$ *and* $\sup_{\theta\in\mathsf{K}}|f_\theta|_{V^\beta} < \infty$*. Suppose also that there exist* $\epsilon > 0$, $\kappa > 0$ *such that* $\beta + \alpha\kappa < 1-\alpha$ *and*

$$\mathbb{E}^{(l)}_{x,\theta}\left[\sum_{k\geq 1}\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>k\}}k^{-1+\epsilon}\left(D_\beta(\tilde{\theta}_k,\tilde{\theta}_{k-1})+|f_{\tilde{\theta}_k}-f_{\tilde{\theta}_{k-1}}|_{V^\beta}\right)V^{\beta+\alpha\kappa}(\tilde{X}_k)\right] < \infty. \tag{28}$$

*Then* $n^{-1}\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>n\}}\sum_{k=1}^{n} f_{\tilde{\theta}_{k-1}}(\tilde{X}_k)$ *converges to zero in* $\mathbb{P}^{(l)}_{x,\theta}$*-probability.*

*Proof.* Define $\tilde{H}_{a,\theta}(x,y) = \tilde{g}_a(y,\theta) - P_\theta \tilde{g}_a(x,\theta)$ and $S_n = \sum_{k=1}^n \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}} > k-1\}} f_{\tilde{\theta}_{k-1}}(\tilde{X}_k)$. Note that $\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}} > n\}} n^{-1} \sum_{k=1}^n f_{\tilde{\theta}_{k-1}}(\tilde{X}_k) = \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}} > n\}} n^{-1} S_n$. Then we use (22) to re-write $S_n$ as:

$$\begin{aligned} S_n &= \sum_{k=1}^n \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}} > k-1\}} \tilde{H}_{a_n,\tilde{\theta}_{k-1}}(\tilde{X}_{k-1}, \tilde{X}_k) \;+\; \left((1-a_n)^{-1} - 1\right) \sum_{k=1}^n \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}} > k-1\}} \tilde{g}_{a_n}(\tilde{X}_k, \tilde{\theta}_{k-1}) \\ &\quad + \left(P_{\theta_0} \tilde{g}_{a_n}(\tilde{X}_0, \theta_0) - \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}} > n\}} P_{\tilde{\theta}_n} \tilde{g}_{a_n}(\tilde{X}_n, \tilde{\theta}_n)\right) \\ &\quad + \sum_{k=1}^n \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}} > k\}} \left(P_{\tilde{\theta}_k} \tilde{g}_{a_n}(\tilde{X}_k, \tilde{\theta}_k) - P_{\tilde{\theta}_{k-1}} \tilde{g}_{a_n}(\tilde{X}_k, \tilde{\theta}_{k-1})\right) + \sum_{k=1}^n \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}} = k\}} P_{\tilde{\theta}_{k-1}} \tilde{g}_{a_n}(\tilde{X}_k, \tilde{\theta}_{k-1}). \end{aligned}$$

We take $a_n \propto n^{-\rho} \in (0, 1/2]$ where $\rho > 0$ is such that $\rho(1-\kappa) < \min\left(0.5, \alpha, 1 - p^{-1}\right)$ where $p = (1-\alpha)(\beta + \alpha\kappa)^{-1} > 1$; and $\rho(2-\kappa) < \epsilon$, where $\kappa$ and $\epsilon$ are as in (28). First, we notice that

$$\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}} > n\}} \sum_{k=1}^n \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}} = k\}} P_{\tilde{\theta}_{k-1}} g_{a_n}(\tilde{X}_k, \tilde{\theta}_{k-1}) = 0.$$

Then we consider the term $M_{n,k} \stackrel{\text{def}}{=} \sum_{j=1}^k \tilde{\mathbb{1}}_{\{\overleftarrow{\tau}_{\mathsf{K}} > k-1\}} H_{a_n, \tilde{\theta}_{j-1}}(\tilde{X}_{j-1}, \tilde{X}_j)$. Clearly, $\{(M_{n,k}, \mathcal{F}_k)\}$ is a martingale array. Applying Proposition 3.2 and Proposition 3.6 (with $p = (1-\alpha)/(\beta + \alpha\kappa) > 1$), we get

$$\begin{aligned} \mathbb{E}_{x,\theta}^{(l)}\left[|M_{n,n}|^p\right] \le C(\mathsf{K}) a_n^{p(\kappa-1)} n^{max(1,p/2)-1} \mathbb{E}_{x,\theta}^{(l)}\left(\sum_{k=1}^n \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}} > k-1\}} V^{1-\alpha}(\tilde{X}_k)\right) \\ = O\left(n^{\rho p(1-\kappa)} n^{\max(1,p/2)}\right). \end{aligned}$$

By the choice of $\rho$, $\rho(1-\kappa) + \max(0.5, p^{-1}) < 1$ and we conclude that $M_{n,n}/n$ converges in $L^p$ to zero.

Define $R_n^{(1)} \stackrel{\text{def}}{=} \left((1-a_n)^{-1} - 1\right) \sum_{k=1}^n \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}} > k-1\}} \tilde{g}_{a_n}(\tilde{X}_k, \tilde{\theta}_{k-1})$. Proposition 3.2 (i) implies that

$$\mathbb{E}_{x,\theta}^{(l)}\left[n^{-1}|R_n^{(1)}|\right] \le C a_n^\kappa n^{-1} \mathbb{E}_{x,\theta}^{(l)}\left(\sum_{k=1}^n \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}} > k-1\}} V^{1-\alpha}(\tilde{X}_k)\right) = O\left(a_n^\kappa\right).$$

The rhs converges to zero since $a_n \to 0$ and $\kappa > 0$.

We turn to $R_n^{(2)} \stackrel{\text{def}}{=} P_{\theta_0} \tilde{g}_{a_n}(\tilde{X}_0, \theta_0) - \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}} > n\}} P_{\tilde{\theta}_n} \tilde{g}_{a_n}(\tilde{X}_n, \tilde{\theta}_n)$. Again, by Proposition 3.2 (i), the drift condition in A2, and Proposition 3.4 (i)

$$\begin{aligned} \mathbb{E}_{x,\theta}^{(l)}\left(n^{-1}|R_n^{(2)}|\right) \le C n^{-1} a_n^{-1+\kappa} \mathbb{E}_{x,\theta}^{(l)}\left(V^{\beta+\alpha\kappa}(\tilde{X}_0) + \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}} > n\}} V^{\beta+\alpha\kappa}(\tilde{X}_n)\right) \\ = O\left(n^{-1+\beta+\alpha\kappa} a_n^{-1+\kappa}\right) = O\left(n^{-\alpha+\rho(1-\kappa)}\right). \end{aligned}$$

Given the assumption $\rho(1-\kappa) < \alpha$, it follows that $n^{-1} R_n^{(2)}$ converges in probability to zero.

We finally turn to $R_n^{(3)} \stackrel{\text{def}}{=} \sum_{k=1}^n \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>k\}} \left(P_{\tilde{\theta}_k}\tilde{g}_{a_n}(\tilde{X}_k,\tilde{\theta}_k) - P_{\tilde{\theta}_{k-1}}\tilde{g}_{a_n}(\tilde{X}_k,\tilde{\theta}_{k-1})\right)$. By definition, $P_\theta\tilde{g}_a(x,\theta) - P_{\theta'}\tilde{g}_a(x,\theta') = f_{\theta'}(x) - f_\theta(x) + (1-a)^{-1}(\tilde{g}_a(x,\theta) - \tilde{g}_a(x,\theta'))$. By Proposition 3.3 (with $\delta = 0$) we have:

$$\begin{aligned}\left|P_\theta\tilde{g}_{a_n}(x,\theta) - P_{\theta'}\tilde{g}_{a_n}(x,\theta')\right| &\leq C(\mathsf{K})\sup_{\theta\in\mathsf{K}}|f_\theta|_{V^\beta} a_n^{-2+\kappa}\left(D_\beta(\theta,\theta') + |f_\theta - f_{\theta'}|_{V^\beta}\right)V^{\beta+\alpha\kappa}(x) \\ &\leq C(\mathsf{K})\sup_{\theta\in\mathsf{K}}|f_\theta|_{V^\beta} n^\epsilon\left(D_\beta(\theta,\theta') + |f_\theta - f_{\theta'}|_{V^\beta}\right)V^{\beta+\alpha\kappa}(x)\end{aligned}$$

Therefore Kronecker's lemma and (28) implies that $n^{-1}R_n^{(3)}$ converge almost surely to zero. □

The next result will be useful in proving the central limit theorem. We take $f \in \mathcal{L}_{V^\beta}$ and let $g_a$ be the resolvent associated with $f$ and $H_{a,\theta}(x,y) := g_a(y,\theta) - P_\theta g_a(x,\theta)$. We will show in the next lemma that $n^{-1/2}\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>n\}}\sum_{k=1}^n f(\tilde{X}_k)$ behaves like the martingale array $n^{-1/2}\sum_{k=1}^n \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>k-1\}} H_{a_n,\tilde{\theta}_{k-1}}(\tilde{X}_{k-1},\tilde{X}_k)$ as $n\to\infty$ for some well chosen sequence $\{a_n,\ n\geq 0\}$.

**Lemma 3.8.** *Assume A1 with $\alpha < 1/2$ and let $\mathsf{K}$ a compact subset of $\Theta$. Let $\beta \geq 0$ such that $2(\beta+\alpha) < 1$ and $f \in \mathcal{L}_{V^\beta}$ such that $\pi(f) = 0$. Let $\kappa > 1$, $\delta \in (0,1)$ be such that $2\beta + \alpha(\kappa+\delta) < 1-\alpha$. Take $\rho \in (1/2, 1/(2-\delta)]$ and let $\{a_n,\ n\geq 0\}$ be a sequence of positive numbers such that $a_n \in (0,1/2]$, $a_n \propto n^{-\rho}$. Suppose that*

$$\mathbb{E}_{x,\theta}^{(l)}\left[\sum_{k\geq 1}\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>k\}} k^{-1+\rho(2-\delta)} D_{\beta+\alpha\delta}(\tilde{\theta}_k,\tilde{\theta}_{k-1}) V^{\beta+\alpha(\kappa+\delta)}(\tilde{X}_k)\right] < \infty. \tag{29}$$

*For any $s \geq 0$, $n^{-1/2}\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>n\}}\sum_{k=1}^n\left(f(\tilde{X}_k) - H_{a_{n+s},\tilde{\theta}_{k-1}}(\tilde{X}_{k-1},\tilde{X}_k)\right)$ converges to zero in $\mathbb{P}_{x,\theta}^{(l)}$-probability.*

*Proof.* Without any loss of generality, we assume that $\kappa$ also satisfies $\beta + \alpha\kappa < 1/2$. For $s \geq 0$ arbitrary, define $S_{n,s} = \sum_{k=1}^n \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>k-1\}}\left(f(\tilde{X}_k) - H_{a_{n+s},\tilde{\theta}_{k-1}}(\tilde{X}_{k-1},\tilde{X}_k)\right)$. Note that

$$\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>n\}} n^{-1/2}\sum_{k=1}^n\left(f(\tilde{X}_k) - H_{a_{n+s},\tilde{\theta}_{k-1}}(\tilde{X}_{k-1},\tilde{X}_k)\right) = \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>n\}} n^{-1/2} S_n.$$

Then we use the approximate Poisson equation (22) to re-write $S_{n,s}$ as:

$$\begin{aligned} S_{n,s} \quad = \quad & \left((1-a_{n+s})^{-1} - 1\right)\sum_{k=1}^n \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>k-1\}} g_{a_{n+s}}(\tilde{X}_k,\tilde{\theta}_{k-1}) \\ & + \left(P_{\theta_0}g_{a_{n+s}}(\tilde{X}_0,\theta_0) - \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>n\}} P_{\tilde{\theta}_n}g_{a_{n+s}}(\tilde{X}_n,\tilde{\theta}_n)\right) \\ & + \sum_{k=1}^n \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>k\}}\left(P_{\tilde{\theta}_k}g_{a_{n+s}}(\tilde{X}_k,\tilde{\theta}_k) - P_{\tilde{\theta}_{k-1}}g_{a_{n+s}}(\tilde{X}_k,\tilde{\theta}_{k-1})\right) \\ & + \sum_{k=1}^n \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}=k\}} P_{\tilde{\theta}_{k-1}}g_{a_{n+s}}(\tilde{X}_k,\tilde{\theta}_{k-1}). \end{aligned}$$

Notice that $\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>n\}} \sum_{k=1}^n \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}=k\}} P_{\tilde{\theta}_{k-1}} g_{a_{n+s}}(\tilde{X}_k, \tilde{\theta}_{k-1}) = 0$. For the rest, consider $R_n^{(1)} \stackrel{\text{def}}{=} \left( P_{\theta_0} g_{a_{n+s}}(\tilde{X}_0, \theta_0) - \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>n\}} P_{\tilde{\theta}_n} g_{a_{n+s}}(\tilde{X}_n, \tilde{\theta}_n) \right)$. By Proposition 3.2, the choice $\kappa > 1$, and by Proposition 3.4 (i) we have

$$\mathbb{E}_{x,\theta}^{(l)} \left( |R_n^{(1)}| \right) \leq C(\mathsf{K}) \mathbb{E}_{x,\theta}^{(l)} \left( V^{\beta+\alpha\kappa}(x) + V^{\beta+\alpha\kappa}(\tilde{X}_n) \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>n\}} \right) = O\left( n^{\beta+\alpha\kappa} \right).$$

Since $\beta + \alpha\kappa < 1/2$ we deduce that $n^{-1/2} R_n^{(1)} \to 0$ in probability.

Now take $R_n^{(2)} \stackrel{\text{def}}{=} \left( (1 - a_{n+s})^{-1} - 1 \right) \sum_{k=1}^n \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>k-1\}} g_{a_{n+s}}(\tilde{X}_k, \tilde{\theta}_{k-1})$. We can apply Proposition 3.2 to obtain

$$|R_n^{(2)}| \leq C(\mathsf{K}) a_{n+s} \sum_{k=1}^n \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>k-1\}} V^{\beta+\alpha\kappa}(\tilde{X}_k)$$

and by Proposition 3.4 (ii), $\mathbb{E}_{x,\theta}^{(l)} \left( n^{-1/2} |R_n^{(2)}| \right) = O\left( n^{1/2} a_n \right)$. By assumption $a_n \propto n^{-\rho}$ with $\rho > 1/2$, thus $n^{-1/2} R_n^{(2)}$ converges in probability to zero.

Finally, we consider $R_n^{(3)} \stackrel{\text{def}}{=} \sum_{k=1}^n \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>k\}} \left( P_{\tilde{\theta}_k} g_{a_{n+s}}(\tilde{X}_k, \tilde{\theta}_k) - P_{\tilde{\theta}_{k-1}} g_{a_{n+s}}(\tilde{X}_k, \tilde{\theta}_{k-1}) \right)$. By definition,

$$P_\theta g_a(x, \theta) - P_{\theta'} g_a(x, \theta') = (1 - a)^{-1} (g_a(x, \theta) - g_a(x, \theta')),$$

and by Proposition 3.2 applied with $\kappa > 1$ and $\delta > 0$, $|P_\theta g_a(x, \theta) - P_{\theta'} g_a(x, \theta')| \leq C(\mathsf{K}) \zeta_\delta(a) D_{\beta+\alpha\delta}(\theta, \theta') V^{\beta+\alpha(\kappa+\delta)}(x)$ so that

$$|n^{-1/2} R_n^{(3)}| \leq C(\mathsf{K}) n^{-1/2} n^{\rho(1-\delta)} \sum_{k=1}^n \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>k-1\}} D_{\beta+\alpha\delta}(\tilde{\theta}_k, \tilde{\theta}_{k-1}) V^{\beta+\alpha(\kappa+\delta)}(\tilde{X}_k).$$

By assumption $n^{-1/2} n^{\rho(1-\delta)} = o(n^{-1+\rho(2-\delta)})$. Kronecker's lemma and (29) then gives that $n^{-1/2} R_n^{(3)}$ converges to 0 with probability one. □

3.4. **Connection with the adaptive MCMC process.** In this section we give a number of results that connects the non-homogeneous Markov chain $\{(\tilde{X}_n, \tilde{\theta}_n),\ n \geq 0\}$ with the adaptive MCMC process $\{(X_n, \theta_n, \nu_n, \xi_n),\ n \geq 0\}$ defined in Section 2.2. This will allow us to transfer the limit results established above to the adaptive chain.

We introduce the sequence of stopping times associated with the adaptive chain

$$T_0 = 0 \qquad T_{j+1} \stackrel{\text{def}}{=} \inf\{k > T_j,\ \xi_k = 0\}, \quad k \geq 1,$$

with the convention that $\inf \emptyset = \infty$. Also define

$$\nu_\infty \stackrel{\text{def}}{=} \sup_{k \geq 0} \nu_k.$$

**Lemma 3.9.** *If A2 hold then* $\check{\mathbb{P}}\left( T_{\nu_\infty} < \infty \right) = 1$.

*Proof.* A2 states that $\check{\mathbb{P}}\left(\nu_\infty < \infty\right) = 1$. Thus under A2

$$\check{\mathbb{P}}\left(T_{\nu_\infty} = +\infty\right) = \sum_{j\geq 0} \check{\mathbb{P}}_\star\left(T_j = +\infty, \nu_\infty = j\right) = 0,$$

the last equality follows from the fact that on the set $\{T_j = +\infty\}$, $\sup_{k\geq 0} \nu_k \leq j-1$. Hence, $\check{\mathbb{P}}\left(T_{\nu_\infty} < +\infty\right) = 1$. □

The following is Lemma 4.1 of Andrieu et al. (2005).

**Proposition 3.10.** *For any $n \in \mathbb{N}$, any $n$-uplet $(t_1, \cdots, t_n)$, any bounded measurable functions $\{f_k, k \leq n\}$ and for any $(x, \theta, j) \in \mathsf{X} \times \mathsf{K}_j \times \mathbb{N}$,*

$$\check{\mathbb{E}}_{x,\theta,j,0}\left[\prod_{k=1}^{n} f_k(X_{t_k}, \theta_{t_k}) \mathbb{1}_{\{T_1 > t_n\}}\right] = \mathbb{E}^{(j)}_{x,\theta}\left[\prod_{k=1}^{n} f_k(\tilde{X}_{t_k}, \tilde{\theta}_{t_k}) \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_j} > t_n\}}\right]$$

One can obtain the finiteness of moments of the adaptive chain as in the following lemma.

**Lemma 3.11.** *Let $\tilde{W}_n = W(\tilde{X}_n, \tilde{\theta}_n, \tilde{X}_{n+1})$ be a sequence of random variables such that for all $l, k \leq n$,*

$$c^{(l)}_k := \sup_{(x,\theta)\in \mathsf{X}_0\times\Theta_0} \mathbb{E}^{(l)}_{x,\theta}\left[\tilde{W}_k \boldsymbol{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > k\}}\right] < \infty.$$

*Then $\check{\mathbb{E}}\left(W(X_n, \theta_n, X_{n+1})\right)$ is finite.*

*Proof.* Denote $W_n = W(X_n, \theta_n, X_{n+1})$. We have

$$\begin{aligned}
\check{\mathbb{E}}\left[W_n\right] &= \sum_{j=0}^{n}\sum_{s=j}^{n} \check{\mathbb{E}}\left[W_n \mathbf{1}_{\{\nu_n = j\}} \mathbf{1}_{\{T_j = s\}}\right] = \sum_{j=0}^{n}\sum_{s=j}^{n} \check{\mathbb{E}}\left[W_n \mathbf{1}_{\{T_j = s\}} \mathbf{1}_{\{T_{j+1} > s + (n-s)\}}\right] \\
&= \sum_{j=0}^{n}\sum_{s=j}^{n} \check{\mathbb{E}}\left[\mathbf{1}_{\{T_j = s\}} \check{\mathbb{E}}_{X_s,\theta_s,j,0}\left(W(X_{n-s}, \theta_{n-s}, X_{n+1-s}) \mathbf{1}_{\{T_1 > n-s\}}\right)\right] \\
&= \sum_{j=0}^{n}\sum_{s=j}^{n} \check{\mathbb{E}}\left[\mathbf{1}_{\{T_j = s\}} \mathbb{E}^{(j)}_{X_s,\theta_s}\left(\tilde{W}_{n-s} \mathbf{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_j} > n-s\}}\right)\right] \leq \sum_{j=0}^{n}\sum_{s=j}^{n} c^{(j)}_{n-s} \check{\mathbb{P}}\left(T_j = s\right) < \infty.
\end{aligned}$$

The last equality uses Proposition 3.10. □

In very general terms, the next result shows that a weak law of large numbers for the re-projection free process $\{(X_n, \theta_n),\ n \geq 0\}$ implies a similar result from the adaptive chain.

**Lemma 3.12.** *Assume A2. Let $\{\tilde{W}_{n,k},\ 1 \leq k \leq n\}$ be a triangular array of random variables of the form $\tilde{W}_{n,k} = W_n(\tilde{\theta}_{k-1}, \tilde{X}_{k-1}, \tilde{\theta}_k, \tilde{X}_k)$ for some measurable functions $W_n : \Theta \times \mathsf{X} \times \Theta \times \mathsf{X} \to \mathbb{R}$. Let $\{b_n,\ n \geq 1\}$ a non-increasing sequence of positive number with*

$\lim_{n\to\infty} b_n = 0$. *Suppose that for any* $k \geq 1$, $\sup_{n\geq 1} |W_n(\theta_{k-1}, X_{k-1}, \theta_k, X_k)| < \infty$ $\check{\mathbb{P}}$*-a.s. and for all* $l \geq 0$, $s \geq 0$, $(x,\theta) \in \mathsf{X}_0 \times \mathsf{K}_l$ *and* $\delta > 0$

$$\lim_{n\to\infty} \mathbb{P}^{(l)}_{x,\theta}\left[b_n \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > n\}} \left|\sum_{k=1}^{n} \tilde{W}_{n+s,k}\right| > \delta\right] = 0,$$

*then* $b_n \sum_{k=1}^n W_n(\theta_{k-1}, X_{k-1}, \theta_k, X_k)$ *converges to zero in* $\check{\mathbb{P}}$*-probability as* $n \to \infty$.

*Proof.* The idea of the proof is similar to the proof of Proposition 6 of Andrieu and Moulines (2006). Write $W_{n,k} = W_n(\theta_{k-1}, X_{k-1}, \theta_k, X_k)$. As shown above A2 implies that $T_{\nu_\infty}$ is finite $\check{\mathbb{P}}$-a.s. With the convention that $\sum_a^b \cdot = 0$ if $a > b$, we write ;

$$\begin{aligned} b_n \sum_{k=1}^{n} W_{n,k} &= b_n \sum_{k=1}^{n\wedge T_{\nu_\infty}} W_{n,k} + b_n \sum_{k=T_{\nu_\infty}+1}^{n} W_{n,k} \\ &= b_n S_n^{(1)} + b_n S_n^{(2)}. \end{aligned}$$

where $S_n^{(1)} = \sum_{k=1}^{n\wedge T_{\nu_\infty}} W_{n,k}$ and $S_n^{(2)} = \sum_{k=T_{\nu_\infty}+1}^{n} W_{n,k}$. Since $\sup_{n\geq 1}|W_{n,k}|$ and $T_{\nu_\infty}$ are finite $\check{\mathbb{P}}$-a.s., we deduce that $|S_n^{(1)}| \leq \sum_{k=1}^{T_{\nu_\infty}} \sup_{n\geq 1}|W_{n,k}|$ is also finite $\check{\mathbb{P}}$-a.s. Therefore $b_n S_n^{(1)}$ converges to zero $\check{\mathbb{P}}$-a.s..

Take $\epsilon > 0$. From Lemma 3.9, we can find $L_2 > L_1 > 0$ such that $\check{\mathbb{P}}\left[\nu_\infty \geq L_1\right] + \check{\mathbb{P}}\left[T_{\nu_\infty} \geq L_2\right] \leq \epsilon$. For any $\delta > 0$ and $n > L_2$, we have:

$$\check{\mathbb{P}}\left[b_n |S_n^{(2)}| \geq \delta\right] \leq \sum_{l=0}^{L_1}\sum_{s=0}^{L_2} \check{\mathbb{P}}\left[b_n|S_n^{(2)}| > \delta, T_l = s, \nu_\infty = l\right] + \epsilon.$$

We then observe that the event

$$\left\{b_n|S_n^{(2)}| > \delta, T_l = s, \nu_\infty = l\right\} \subseteq \left\{b_{n-s}|\sum_{k=T_l+1}^{T_l+n-s} W_{n,k}|\mathbb{1}_{\{T_{l+1} > T_l+n-s\}} > \delta\right\}.$$

Therefore by conditioning on $\mathcal{F}_{T_l}$, we get:

$$\begin{aligned} &\check{\mathbb{P}}\left[b_n|S_n^{(2)}| > \delta, T_l = s, \nu_\infty = l\right] \\ &\qquad \leq \check{\mathbb{E}}\left[\check{\mathbb{P}}_{X_{T_l},\theta_{T_l},l,0}\left(b_{n-s}\left|\sum_{k=1}^{n-s} W_{n,k}\right| \mathbb{1}_{\{T_1 > n-s\}} > \delta\right)\right] \\ &\qquad\qquad = \check{\mathbb{E}}\left[\mathbb{P}^{(l)}_{\tilde{X}_{T_l},\tilde{\theta}_{T_l}}\left(b_{n-s}\left|\sum_{k=1}^{n-s} \tilde{W}_{n,k}\right| \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > n-s\}} > \delta\right)\right]. \end{aligned}$$

The last equality follows from Proposition 3.10. By assumption, the inner term in the last expectation above converges almost surely to zero. It follows from Lebesgue's dominated convergence theorem that $\lim_{n\to\infty} \check{\mathbb{P}}\left(b_n|S_n^{(2)}| \geq \delta\right) \leq \epsilon$. Since $\epsilon > 0$ is arbitrary, the results follows. □

3.5. **Proof of Theorem 2.1.** Since A1 and (8) hold, we can apply Proposition 3.7 which implies that

$$\lim_{n\to\infty} \mathbb{P}^{(l)}_{x,\theta}\left[n^{-1}\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l}>n\}}\left|\sum_{k=1}^{n} f_{\theta_{k-1}}(X_k)\right|>\delta\right]=0,$$

for any $\delta > 0$, $l \geq 0$ and $(x,\theta) \in \mathsf{X}\times\mathsf{K}_l$. Theorem 2.1 then follows from Lemma 3.12.

3.6. **Proof of Theorem 2.3.** Throughout the proof, we take $\kappa > 1, \delta \in (0,1)$, $\rho \in (1/2, (2-\delta)^{-1}]$ and $\{a_n,\ n \geq 0\}$ as in the statement of the theorem. Denote $S_n = \sum_{k=1}^n f(X_k)$. Without any loss of generality, we will assume that $|f|_{V^\beta} \leq 1$. We have

$$S_n = \sum_{k=1}^{n} H_{a_n,\theta_{k-1}}(X_{k-1},X_k)\mathbb{1}_{\{\xi_{k-1}\neq 0\}} + \sum_{k=1}^{n} H_{a_n,\theta_{k-1}}(X_{k-1},X_k)\mathbb{1}_{\{\xi_{k-1}=0\}} + \sum_{k=1}^{n}\left(f(X_k) - H_{a_n,\theta_{k-1}}(X_{k-1},X_k)\right).$$

By Theorem 2.2, $n^{-1/2}\sum_{k=1}^n \left(f(X_k) - H_{a_n,\theta_{k-1}}(X_{k-1},X_k)\right)$ converges in $\check{\mathbb{P}}$-probability to zero.

Note that $\xi_k = 0$ signals a re-projection at time $k$. By Proposition 3.2 (i) applied with $\kappa > 1$,

$$\left|\sum_{k=1}^{n} H_{a_n,\theta_{k-1}}(X_{k-1},X_k)\mathbb{1}_{\{\xi_{k-1}=0\}}\right| \leq C(\mathsf{K}_{\nu_\infty})\sum_{k=1}^{T_{\nu_\infty}}\left(V^{1-\alpha}(X_{k-1}) + V^{1-\alpha}(X_k)\right),\ \ \check{\mathbb{P}}-\text{a.s.}$$

and the rhs is finite $\check{\mathbb{P}}$-a.s. We then conclude that $n^{-1/2}\sum_{k=1}^n H_{a_n,\theta_{k-1}}(X_{k-1},X_k)\mathbb{1}_{\{\xi_{k-1}=0\}}$ converges to zero $\check{\mathbb{P}}$-a.s..

Define $M_{n,k} = \sum_{j=1}^k D_{n,j}$, where $D_{n,j} = n^{-1/2}H_{a_n,\theta_{j-1}}(X_{j-1},X_j)\mathbb{1}_{\{\xi_{j-1}\neq 0\}}$. It is straightforward to see that $\{(M_{n,k},\mathcal{F}_k),\ 1 \leq k \leq n\}$ is a martingale array. We will show that

$$\{(M_{n,k},\check{\mathcal{F}}_k),\ 1\leq k\leq n\} \quad \text{is a square-integrable martingale array;} \tag{30}$$

$$\lim_{n\to\infty}\sum_{j=1}^{n}\check{\mathbb{E}}\left(D^2_{n,j}|\check{\mathcal{F}}_{j-1}\right) = \sigma^2_\star(f), \quad \left(\text{in } \check{\mathbb{P}}\text{-probab.}\right) \tag{31}$$

where

$$\sigma^2_\star(f) \stackrel{\text{def}}{=} \int \pi(dx)\left(-f^2(x) + 2f(x)g(x,\theta_\star)\right), \tag{32}$$

is finite $\check{\mathbb{P}}$-almost surely and that for any $\epsilon > 0$,

$$\lim_{n\to\infty}\sum_{k=1}^{n}\check{\mathbb{E}}\left(D^2_{n,j}\mathbb{1}_{\{|D_{n,j}|\geq\epsilon\}}|\check{\mathcal{F}}_{j-1}\right) = 0, \quad \left(\text{in } \check{\mathbb{P}}\text{-probab.}\right) \tag{33}$$

By the central limit theorem for martingales (see e.g. Hall and Heyde (1980), Corollary 3.1), (30)-(33) implies that $M_{n,n}$ converges weakly to $Z$ ($M_{n,n} \stackrel{w}{\to} Z$) where $Z$ is a random variable with characteristic function $\phi(t) = \check{\mathbb{E}}\left(e^{-\frac{1}{2}\sigma^2_\star(f)t^2}\right)$. This will end the proof.

`Proof of (30).` It suffices to show that for all $l \geq 0$, $k, n \geq 1$,

$$\sup_{(x,\theta)\in \mathsf{X}_0\times \mathsf{K}_l} \mathbb{E}^{(l)}_{x,\theta}\left(H^2_{a_n,\tilde{\theta}_{k-1}}(\tilde{X}_{k-1},\tilde{X}_k)\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l}>k-1\}}\right) < \infty$$

and to apply Lemma 3.11. By Proposition 3.2 (i) (applied with both $\kappa > 1$ and $\delta > 0$), $\sup_{\theta\in\mathsf{K}} |g_a(x,\theta)|^2 \leq C(\mathsf{K})\zeta_\delta(a)V^{2\beta+\alpha(\kappa+\delta)}(x) \leq C(\mathsf{K})\zeta_\delta(a)V^{1-\alpha}(x)$ since by assumption $2\beta+\alpha(\kappa+\delta) \leq 1-\alpha$. Thus for any $l \geq 0$, $k, n \geq 1$ and $(x,\theta) \in \mathsf{X}\times\mathsf{K}_l$,

$$\begin{aligned}\mathbb{E}^{(l)}_{x,\theta}\left(H^2_{a_n,\tilde{\theta}_{k-1}}(\tilde{X}_{k-1},\tilde{X}_k)\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l}>k-1\}}|\mathcal{F}_{k-1}\right) &\leq \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l}>k-1\}}P_{\tilde{\theta}_{k-1}}g^2_{a_n}(\tilde{X}_{k-1},\tilde{\theta}_{k-1}) \\ &\leq C(\mathsf{K}_l)\zeta_\delta(a_n)\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l}>k-1\}}V^{1-\alpha}(\tilde{X}_{k-1}).\end{aligned}$$

From Proposition 3.4 (i) we thus obtain

$$\sup_{(x,\theta)\in \mathsf{X}_0\times \mathsf{K}_l} \mathbb{E}^{(l)}_{x,\theta}\left(H^2_{a_n,\tilde{\theta}_{k-1}}(\tilde{X}_{k-1},\tilde{X}_k)\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l}>k-1\}}\right) \leq C(\mathsf{K}_l)\zeta_\delta(a_n)k^{1-\alpha}\sup_{x\in\mathsf{X}_0} V^{1-\alpha}(x) < \infty.$$

`Proof of (31).`

$$\begin{aligned}\check{\mathbb{E}}\left(D^2_{n,j}|\check{\mathcal{F}}_{j-1}\right) &= \mathbb{1}_{\{\xi_{j-1}\neq 0\}}n^{-1}P_{\theta_{j-1}}H^2_{a_n,\theta_{j-1}}(X_{j-1}) \\ &= n^{-1}P_{\theta_{j-1}}H^2_{a_n,\theta_{j-1}}(X_{j-1}) \;\; - \mathbb{1}_{\{\xi_{j-1}=0\}}n^{-1}P_{\theta_{j-1}}H^2_{a_n,\theta_{j-1}}(X_{j-1}).\end{aligned}$$

The same argument as above shows that

$$n^{-1}\sum_{j=1}^n \mathbb{1}_{\{\xi_{j-1}=0\}}P_{\theta_{j-1}}H^2_{a_n,\theta_{j-1}}(X_{j-1}) \leq n^{-1}\zeta_\delta(a_n)C(\mathsf{K}_{\nu_\infty})\sum_{j=1}^{T_{\nu_\infty}} V^{1-\alpha}(X_{j-1}),$$

which converges almost surely to zero since $T_{\nu_\infty}$ is finite $\check{\mathbb{P}}$-almost surely, $\zeta_\delta(a_n) = O(n^{\rho(1-\delta)})$ and $\rho(1-\delta) < 1/2$.

For the first term, we note that $P_\theta H^2_{a,\theta}(x,\theta) = P_\theta g^2_a(x,\theta) - (P_\theta g_a(x,\theta))^2 = P_\theta g^2_a(x,\theta) - \left((1-a)^{-1}g_a(x,\theta) - f(x)\right)^2$. We thus have the decomposition:

$$\frac{1}{n}\sum_{k=1}^n P_{\theta_k-1}H^2_{a_n,\theta_{k-1}}(X_{k-1}) = \frac{1}{n}\sum_{i=1}^6 T_n^{(i)} + \int \pi(dx)\left(-f^2(x) + 2f(x)g(x,\theta_\star)\right),$$

where

$$T_n^{(1)} = \sum_{k=1}^n P_{\theta_{k-1}}g^2_{a_n}(X_{k-1},\theta_{k-1}) - g^2_{a_n}(X_{k-1},\theta_{k-1}),$$

$$T_n^{(2)} = \left(1-(1-a_n)^{-2}\right)\sum_{k=1}^n g^2_{a_n}(X_{k-1},\theta_{k-1}),$$

$$T_n^{(3)} = 2\left((1-a_n)^{-1}-1\right)\sum_{k=1}^n f(X_{k-1})g_{a_n}(X_{k-1},\theta_{k-1}),$$

$$T_n^{(4)} = 2\sum_{k=1}^n f(X_{k-1})\left(g_{a_n}(X_{k-1},\theta_{k-1}) - g(X_{k-1},\theta_{k-1})\right),$$

$$T_n^{(5)} = 2 \sum_{k=1}^{n} \int \pi(dx) f(x) \left( g(x, \theta_{k-1}) - g(x, \theta_\star) \right).$$

$$T_n^{(6)} = \sum_{k=1}^{n} \left[ -f^2(X_{k-1}) + 2f(X_{k-1}) g(X_{k-1}, \theta_{k-1}) - \int \pi(dx) \left( -f^2(x) + 2f(x) g(x, \theta_{k-1}) \right) \right].$$

By assumption $n^{-1}T_n^{(1)}$ converges in $\check{\mathbb{P}}$-probability to zero. We will use the same technique to study the term $T_n^{(2)}$ to $T_n^{(5)}$. For example for $T_n^{(2)}$, the idea is to introduce its counterpart $\tilde{T}_{n,s}^{(1)}$ in the space of the re-projection free process $\{(\tilde{X}_n, \tilde{\theta}_n),\ n \geq 0\}$, to show that $\lim_{n\to\infty} \mathbb{P}_{x,\theta}^{(l)} \left( |\tilde{T}_{n,s}^{(2)}| > \delta \right) = 0$ for any $l \geq 0$, $\delta > 0$ and any $(x, \theta) \in \mathsf{X}_0 \times \Theta_l$ and then to argue that $\lim_{n\to\infty} \check{\mathbb{P}} \left( |T_n^{(1)}| > \delta \right) = 0$ for all $\delta > 0$ using Lemma 3.12.

**Lemma 3.13.** $n^{-1} \left( T_n^{(2)} + T_n^{(3)} \right)$ *converges in probability to zero.*

*Proof.* For $l, s \geq 0$, define

$$\tilde{T}_{n,s} \stackrel{\text{def}}{=} \left( 1 - (1 - a_{n+s})^{-2} \right) \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > n\}} \sum_{k=1}^{n} g_{a_{n+s}}^2 (\tilde{X}_{k-1}, \tilde{\theta}_{k-1})$$
$$+ \left( (1 - a_{n+s})^{-1} - 1 \right) \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > n\}} \sum_{k=1}^{n} f(\tilde{X}_{k-1}) g_{a_{n+s}} (\tilde{X}_{k-1}, \tilde{\theta}_{k-1}).$$

We show that for any $\mu > 0$, and any $(x, \theta) \in \mathsf{X} \times \mathsf{K}_l$, $\lim_{n\to\infty} \mathbb{P}_{x,\theta}^{(l)} \left( n^{-1} |\tilde{T}_{n,s}| > \mu \right) = 0$. Then we can apply Lemma 3.12 to conclude that $n^{-1}T_n^{(1)}$ converges in $\check{\mathbb{P}}$-probability to zero. As above, for any $(x, \theta) \in \mathsf{X} \times \mathsf{K}_l$ and by Proposition 3.2 (i), we get

$$\mathbb{E}_{x,\theta}^{(l)} \left( |\tilde{T}_{n,s}| \right) \leq$$
$$C(\mathsf{K}_l) \left( \zeta_\delta(a_{n+s}) + 1 \right) a_{n+s} \mathbb{E}_{x,\theta}^{(l)} \left( \sum_{k=1}^{n} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > k-1\}} V^{2\beta + \alpha(\kappa + \delta)}(\tilde{X}_k) \right) = O\left( n a_n^\delta \right).$$

The rest of the proof follows from the usual bounds on the $V$-moments. □

**Lemma 3.14.** $n^{-1}T_n^{(4)}$ *converges in probability to zero.*

*Proof.* For $l, s \geq 0$, define

$$\tilde{T}_{n,s}^{(4)} \stackrel{\text{def}}{=} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > n\}} \sum_{k=1}^{n} f(\tilde{X}_{k-1}) \left( g_{a_{n+s}}(\tilde{X}_{k-1}, \tilde{\theta}_{k-1}) - g(\tilde{X}_{k-1}, \tilde{\theta}_{k-1}) \right)$$
$$= \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > n\}} \sum_{k=1}^{n} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > k-1\}} \tilde{f}(X_{k-1}) \left( g_{a_{n+s}}(\tilde{X}_{k-1}, \tilde{\theta}_{k-1}) - g(\tilde{X}_{k-1}, \tilde{\theta}_{k-1}) \right).$$

Again, for any $(x, \theta) \in \mathsf{X} \times \mathsf{K}_l$ and by Proposition 3.2 (ii) we get

$$\mathbb{E}_{x,\theta}^{(l)} \left( n^{-1} |\tilde{T}_{n,s}^{(4)}| \right) \leq C(\mathsf{K}_l) a_{n+s} \zeta_{\kappa-1}(a_{n+1}) n^{-1} \mathbb{E}_{x,\theta}^{(l)} \left( \sum_{k=1}^{n} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > k-1\}} V^{2\beta + \alpha\kappa}(\tilde{X}_k) \right) = O\left( a_n \zeta_{\kappa-1}(a_n) \right).$$

The rest of the proof is similar to the above upon noticing that for $\kappa > 1$, $a\zeta_{\kappa-1}(a) \to 0$ as $a \to 0$. $\square$

**Lemma 3.15.** *$n^{-1}T_n^{(5)}$ converges $\check{\mathbb{P}}$-almost surely to zero.*

*Proof.* By Proposition 3.3 (ii), there exists a finite constant $C(\mathsf{K})$ such that for any $\theta, \theta' \in \mathsf{K}$, $x \in \mathsf{X}$ and any $a \in (0, 1/2]$

$$\left|g(x,\theta) - g(x,\theta')\right| \le C(\mathsf{K}) \left(a\zeta_{\kappa-1}(a) + a^{-1}D_{\beta+\alpha\kappa}(\theta,\theta')\right) V^{\beta+\alpha\kappa}(x).$$

Therefore

$$\left|\int \pi(dx)f(x)\left(g(x,\theta) - g(x,\theta')\right)\right| \le C(\mathsf{K}) \left(a\zeta_{\kappa-1}(a) + a^{-1}D_{\beta+\alpha\kappa}(\theta,\theta')\right) \pi\left(V^{2\beta+\alpha\kappa}\right).$$

Let $\epsilon > 0$. Since $a\zeta_{\kappa-1}(a) \to 0$ as $a \to 0$, we can find $a_0 \in (0, 1/2]$ such that $a_0\zeta_{\kappa-1}(a_0) < \epsilon$. Then for $\check{\mathbb{P}}$-almost every sample path

$$\lim_{n\to\infty} \left|\int \pi(dx)f(x)\left(g(x,\tilde{\theta}_n) - g(x,\theta_\star)\right)\right|$$
$$\le C(\mathsf{K}_{\nu_\infty}) \lim_{n\to\infty} \left(\epsilon + a_0^{-1}D_{\beta+\alpha\kappa}(\tilde{\theta}_n,\theta_\star)\right) \pi\left(V^{2\beta+\alpha\kappa}\right) = \epsilon C(\mathsf{K}_{\nu_\infty})\pi\left(V^{2\beta+\alpha\kappa}\right).$$

Since $\epsilon > 0$ is arbitrary and $\pi\left(V^{2\beta+\alpha\kappa}\right) < \infty$, we are finished. $\square$

**Lemma 3.16.** *$n^{-1}T_n^{(6)}$ converges in probability to zero.*

*Proof.* We would like to apply the law of large number (Theorem 2.2) to show that $n^{-1}T_n^{(6)}$ converges to zero. By Proposition 3.2 (ii), for any compact subset $\mathsf{K}$ of $\Theta$, $\sup_{\theta\in\mathsf{K}} |f^2 + 2fg_\theta|_{V^{2\beta+\alpha\kappa}} < \infty$ and $2\beta + \alpha\kappa < 1 - \alpha$. To check (8), it is enough to find $\epsilon > 0$ such that

$$\mathbb{E}_{x,\theta}^{(l)}\left[\sum_{k\ge1} k^{-1+\epsilon}|fg_{\tilde{\theta}_{k-1}} - fg_{\tilde{\theta}_k}|_{V^{2\beta+\alpha\kappa}}\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l}>k\}}V^{2\beta+\alpha(\kappa+\delta)}(\tilde{X}_k)\right] < \infty. \tag{34}$$

But by Proposition 3.3 (ii), there exists a finite constant $C(\mathsf{K})$ such that for any $\theta, \theta' \in \mathsf{K}$, $x \in \mathsf{X}$ and any $a \in (0, 1/2]$

$$\left|f(\cdot)g(\cdot,\theta) - f(\cdot)g(\cdot,\theta')\right|_{V^{2\beta+\alpha\kappa}} \le C(\mathsf{K})a\zeta_{\kappa-1}(a) + a^{-1}D_{\beta+\alpha\kappa}(\theta,\theta').$$

We let $a$ depend on $k$ by taking $a = a_k$, therefore

$$\mathbb{E}_{x,\theta}^{(l)}\left[\sum_{k\ge1} k^{-1+\epsilon}|fg_{\tilde{\theta}_{k-1}} - fg_{\tilde{\theta}_k}|_{V^{2\beta+\alpha\kappa}}\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l}>k\}}V^{2\beta+\alpha(\kappa+\delta)}(\tilde{X}_k)\right]$$
$$\le \mathbb{E}_{x,\theta}^{(l)}\left[\sum_{k\ge1} k^{-1+\epsilon}a_k\zeta_{\kappa-1}(a_k)\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l}>k\}}V^{1-\alpha}(\tilde{X}_k)\right]$$
$$+ \mathbb{E}_{x,\theta}^{(l)}\left[\sum_{k\ge1} k^{-1+\epsilon}a_k^{-1}D_{\beta+\alpha\kappa}(\tilde{\theta}_{k-1},\tilde{\theta}_k)\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l}>k\}}V^{2\beta+\alpha(\kappa+\delta)}(\tilde{X}_k)\right].$$

We can then find $\epsilon > 0$ such that $n^{\epsilon} a_n \zeta_{\kappa-1}(a_n) + n^{-1+\epsilon} a_n^{-1} = O(n^{-\epsilon})$ and (34) follows. □

`Proof of (33).` It is suffices to show that

$$\lim_{n\to\infty} n^{-1} \sum_{k=1}^{n} \int P_{\theta_{k-1}}(X_{k-1}, dy) H^2_{a_n,\theta_{k-1}}(X_{k-1}, y) \mathbb{1}_{\{|Ha_n,\theta_{k-1}(X_{k-1},y)| \geq \epsilon\sqrt{n}\}} = 0,$$

in $\check{\mathbb{P}}$-probability. We will do so by applying Lemma 3.12 again. By a lemma due to Dvoretzky (Lemma 9 of Andrieu and Moulines (2006))

$$\int P_{\theta_{k-1}}(X_{k-1}, dy) H^2_{a_n,\theta_{k-1}}(X_{k-1}, y) \mathbb{1}_{\{|H_{a_n,\theta_{k-1}}(X_{k-1},y)| > \epsilon\sqrt{n}\}} \leq 4W_{n,k},$$

where

$$W_{n,k} = \int P_{\theta_{k-1}}(X_{k-1}, dy) g^2_{a_n}(y, \theta_{k-1}) \mathbb{1}_{\{|g_{a_n}(y,\theta_{k-1})| > \epsilon\sqrt{n}/2\}}.$$

It is thus enough to show that for any $s, l \geq 0$, any $(x, \theta) \in \mathsf{X}_0 \times \mathsf{K}_l$,

$$\lim_{n\to\infty} n^{-1} \sum_{k=1}^{n} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > k-1\}} \tilde{W}_{n+s,k} = 0, \quad (\text{in } \mathbb{P}^{(l)}_{x,\theta}\text{-probability}).$$

Take $p > 2$ such that $p(\beta + \alpha/2) < 1 - \alpha$. Then

$$\begin{aligned}
\mathbb{E}^{(l)}_{x,\theta}\left(\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > k-1\}} \tilde{W}_{n+s,k}\right) &= \mathbb{E}^{(l)}_{x,\theta}\left(\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > k-1\}} \left|g_{a_{n+s}}(\tilde{X}_k, \tilde{\theta}_{k-1})\right|^2 \mathbb{1}_{\{|g_{a_{n+s}}(\tilde{X}_k,\tilde{\theta}_{k-1})| > \epsilon\sqrt{n+s}/2\}}\right) \\
&\leq (2/\epsilon)^{-p}(n+s)^{-p/2}\mathbb{E}^{(l)}_{x,\theta}\left(\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > k-1\}} \left|g_{a_{n+s}}(\tilde{X}_k, \tilde{\theta}_{k-1})\right|^p\right) \\
&\leq (2/\epsilon)^{-p} C(\mathsf{K}_l) n^{-p/2} \left(\zeta_{1/2}(a_n)\right)^p \mathbb{E}^{(l)}_{x,\theta}\left(\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > k-1\}} V^{1-\alpha}(\tilde{X}_k)\right).
\end{aligned}$$

It follows that

$$n^{-1}\mathbb{E}^{(l)}_{x,\theta}\left(\sum_{k=1}^{n} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > k-1\}} \tilde{W}_{n+s,k}\right) = O\left(n^{-p(1-\rho)/2}\right).$$

and since $\rho < 1$, we are done.

### 3.7. **Proof of Proposition 2.4.**

*Proof.* Denote $g_a(x) = \sum_{j\geq 0}(1-a)^{j+1} P^j f(x)$, $H_a(x, y) = g_a(y) - Pg_a(x)$ and write $g$ and $H$ respectively when $a = 0$. Denote $L^2(\pi\times P)$ the $L^2$-space with respect to the joint measure $\pi(dx)P(x, dy)$ on $\mathsf{X} \times \mathsf{X}$. It is shown by Maxwell and Woodroofe (2000) (Proposition 1) that if $f \in L^2(\pi)$ and $\sum_{j\geq 1} j^{-1/2}\|P^j f\|_{L^2(\pi)} < \infty$ then there exists $H_\star \in L^2(\pi \times P)$ such that $\lim_{a\to 0} \|H_a - H_\star\|_{L^2(\pi\times P)} = 0$.

Under (6) and with $f \in \mathcal{L}_{V^\beta}$, $\beta \in [0, 1/2 - \alpha)$, $\sum_{j\geq 1} j^{-1/2}\|P^j f\|_{L^2(\pi)} < \infty$ and thus there exists $H_\star \in L^2(\pi\times P)$ such that $\lim_{a\to 0} \|H_a - H_\star\|_{L^2(\pi\times P)} = 0$. Moreover $\pi\times P(H_\star^2) = \pi(f(2g-f))$ (see e.g. Holzmann (2005) for a derivation of this formula). From Proposition 3.2 (ii), we see that $H_\star = H$ ($\pi\times P$-a.s.). Note that $\pi\times P(H^2) = \pi\left(Pg^2 - g^2 + f(2g-f)\right)$ and $\pi(|f(2g-f)|) < \infty$ by Proposition 3.2 (i) and the fact that $\pi(V^{1-\alpha}) < \infty$. Thus it follows from $\pi\times P(H^2) < \infty$ and $\pi\times P(H^2) = \pi(f(2g-f))$ that $Pg^2 - g^2$ is $\pi$-integrable and $\pi(Pg^2 - g^2) = 0$.

On the other hand we have $PH_a^2(x) = Pg_a^2(x) - (Pg_a(x))^2 = Pg_a^2(x) - g_a^2(x) + g_a^2(x) - (Pg_a(x))^2$. Similarly $PH^2(x) = Pg^2(x) - g^2(x) + g^2(x) - (Pg(x))^2$. After some algebra we get

$$\left(Pg_a^2(x) - g_a^2(x)\right) - \left(Pg^2(x) - g^2(x)\right) = PH_a^2(x) - PH^2(x) + 2\left((1-a)^{-1} - 1\right) f(x) g_a(x) \\ + 2f(x)\left(g_a(x) - g(x)\right) - \left((1-a)^{-1} + 1\right)\left((1-a)^{-1} - 1\right) g_a^2(x).$$

We take $\kappa > 1$ and $\delta > 0$ such that $2\beta + \alpha(\kappa+\delta) < 1-\alpha$ and apply Proposition 3.2 to get

$$\left|\left(Pg_a^2(x) - g_a^2(x)\right) - \left(Pg^2(x) - g^2(x)\right)\right| \le \left|PH_a^2(x) - PH^2(x)\right| + Ca^\delta V^{2\beta+\alpha(\kappa+\delta)}(x),$$

for some finite constant $C$. It follows that

$$\int \pi(dx)\left|\left(Pg_a^2(x) - g_a^2(x)\right) - \left(Pg^2(x) - g^2(x)\right)\right| \le \\ \|H_a - H\|^2_{L^2(\pi\times P)} + 2\|H_a - H\|_{L^2(\pi\times P)}\|H\|_{L^2(\pi\times P)} + Ca^\delta\pi(V^{1-\alpha}). \quad (35)$$

Then we have

$$n^{-1}\sum_{k=1}^n g_{a_n}^2(X_k) - Pg_{a_n}^2(X_k) = n^{-1}\sum_{k=1}^n g^2(X_k) - Pg^2(X_k) \\ + n^{-1}\sum_{k=1}^n g_{a_n}^2(X_k) - Pg_{a_n}^2(X_k) - \left(g^2(X_k) - Pg^2(X_k)\right).$$

Since $\pi(|g^2 - Pg^2|) < \infty$ and $\pi(g^2 - Pg^2) = 0$, the weak law of large numbers for Markov chains implies that $n^{-1}\sum_{k=1}^n g^2(X_k) - Pg^2(X_k)$ converges in probability to zero. And

$$n^{-1}\mathbb{E}\left[\left|\sum_{k=1}^n g_{a_n}^2(X_k) - P_{\theta_\star}g_{a_n}^2(X_k) - g^2(X_k) - Pg^2(X_k)\right|\right] \\ \le \mathbb{E}\left[\left|g_{a_n}^2(X_0) - P_{\theta_\star}g_{a_n}^2(X_0) - g^2(X_0) - Pg^2(X_0)\right|\right]$$

and the rhs converges to zero as a consequence of (35). □

### 3.8. **Proof of Proposition 2.5.**

*Proof.* Write

$$\sum_{k=1}^n P_{\theta_{k-1}}g_{a_n}^2(X_{k-1},\theta_{k-1}) - g_{a_n}^2(X_{k-1},\theta_{k-1}) = \sum_{k=1}^n P_{\theta_{k-1}}g_{a_n}^2(X_{k-1},\theta_{k-1}) - g_{a_n}^2(X_k,\theta_{k-1}) \\ + \sum_{k=1}^n g_{a_n}^2(X_k,\theta_{k-1}) - g_{a_n}^2(X_k,\theta_k) + \left(g_{a_n}^2(X_n,\theta_n) - g_{a_n}^2(X_0,\theta_0)\right). \quad (36)$$

We first deal with the first term. For $l, s \geq 0$, Define

$$\tilde{T}^{(1)}_{n,s} = \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > n\}} \sum_{k=1}^{n} P_{\tilde{\theta}_{k-1}} g^2_{a_{n+s}}(\tilde{X}_{k-1}, \tilde{\theta}_{k-1}) - g^2_{a_{n+s}}(\tilde{X}_k, \tilde{\theta}_{k-1}).$$

We show that for any $\mu > 0$, and any $(x, \theta) \in \mathsf{X} \times \mathsf{K}_l$, $\lim_{n\to\infty} \mathbb{P}^{(l)}_{x,\theta}\left(n^{-1}|\tilde{T}^{(1)}_{n,s}| > \mu\right) = 0$. Then we can apply Lemma 3.12 to conclude that $\sum_{k=1}^{n} P_{\theta_{k-1}} g^2_{a_n}(X_{k-1}, \theta_{k-1}) - g^2_{a_n}(X_k, \theta_{k-1})$ converges in $\check{\mathbb{P}}$-probability to zero. We have

$$\tilde{T}^{(1)}_{n,s} = \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > n\}} \sum_{k=1}^{n} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > k-1\}} \left(P_{\tilde{\theta}_{k-1}} g^2_{a_{n+s}}(\tilde{X}_{k-1}, \tilde{\theta}_{k-1}) - g^2_{a_{n+s}}(\tilde{X}_k, \tilde{\theta}_{k-1})\right)$$

and $\mathbb{E}^{(l)}_{x,\theta}\left[\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > k-1\}} \left(P_{\tilde{\theta}_{k-1}} g^2_{a_{n+s}}(\tilde{X}_{k-1}, \tilde{\theta}_{k-1}) - g^2_{a_{n+s}}(\tilde{X}_k, \tilde{\theta}_{k-1})\right) |\mathcal{F}_{k-1}\right] = 0$. Let $\kappa > 1$ such that $2(\beta + \alpha\kappa) < 2(\beta + \alpha) + \epsilon$. Set $p = (2(\beta + \alpha) + \epsilon)(2\beta + 2\alpha\kappa)^{-1}$. By Proposition 3.6, we get

$$\mathbb{E}^{(l)}_{x,\theta}\left(|\tilde{T}^{(1)}_{n,s}|^p\right) \leq C(\mathsf{K}_l) n^{1\vee 0.5p} n^{-1} \mathbb{E}^{(l)}_{x,\theta}\left(\sum_{j=1}^{n} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > k-1\}} V^{2(\beta+\alpha)+\epsilon}(\tilde{X}_k)\right) = O\left(n^{1\vee 0.5p}\right),$$

by assumption. Since $p > 1$, the result follows.

We use the same strategy to deal with the second term on the rhs of (36). For $l, s \geq 0$, Define

$$\tilde{T}^{(2)}_{n,s} \stackrel{\text{def}}{=} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > n\}} \sum_{k=1}^{n} g^2_{a_{n+s}}(\tilde{X}_k, \tilde{\theta}_{k-1}) - g^2_{a_{n+s}}(\tilde{X}_k, \tilde{\theta}_k)$$

$$= \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > n\}} \sum_{k=1}^{n} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > k-1\}} \left(g_{a_{n+s}}(\tilde{X}_k, \tilde{\theta}_{k-1}) - g_{a_{n+s}}(\tilde{X}_k, \tilde{\theta}_k)\right) \left(g_{a_{n+s}}(\tilde{X}_k, \tilde{\theta}_{k-1}) + g_{a_{n+s}}(\tilde{X}_k, \tilde{\theta}_k)\right).$$

We apply Proposition 3.2 (i) with $\kappa = \delta/2$ to get $\sup_{\theta,\theta' \in \mathsf{K}_l} |g_a(x, \theta) + g_a(x, \theta')| \leq C(\mathsf{K}_l) a^{-1+\delta/2} V^{\beta+\alpha\delta/2}(x)$. This together with Proposition 3.3 (i) (with $\kappa > 1$ and $\delta/2 > 0$) gives:

$$\left|\tilde{T}^{(2)}_{n,s}\right| \leq C(\mathsf{K}_l) \left(\zeta_{\delta/2}(a_{n+s})\right)^2 \sum_{k=1}^{n} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > k-1\}} D_{\beta+\alpha\delta/2}(\theta_{k-1}, \theta_k) V^{2\beta+\alpha(\kappa+\delta)}(X_k)$$

$n^{-1}\left(\zeta_{\delta/2}(a_{n+s})\right)^2 = O\left(n^{-1+\rho(2-\delta)}\right)$ then Kronecker's lemma and (11) implies that $n^{-1}\tilde{T}^{(2)}_{n,s}$ converges in probability to zero.

For the last term on the rhs of (36), define

$$\tilde{T}^{(3)}_{n,s} \stackrel{\text{def}}{=} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > n\}} \left(g^2_{a_{n+s}}(\tilde{X}_n, \tilde{\theta}_n) - g^2_{a_{n+s}}(\tilde{X}_0, \tilde{\theta}_0)\right).$$

Then with $\kappa_0 > 1$ such that $2(\beta + \alpha\kappa_0) < 1$, we get the bound $\mathbb{E}^{(l)}_{x,\theta}\left(n^{-1}|\tilde{T}^{(3)}_{n,s}|\right) \leq C(\mathsf{K}_l) n^{-1} \mathbb{E}^{(l)}_{x,\theta}\left(V^{2(\beta+\alpha\kappa_0)}(\tilde{X}_n) \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > n\}}\right) = O(n^{1-2(\beta+\alpha\kappa_0)})$. The rest of the proof is similar to the above. □

3.9. **Proof of Proposition 2.8.**

*Proof.* We will show that for any $p \geq 0$, $n \geq 1$, any compact subset $\mathsf{K}$ of $\Theta$ and any $\delta > 0$,

$$\sup_{(x,\theta)\in \mathsf{X}_0\times\Theta_0} \mathbb{P}^{(p)}_{x,\theta}\left(C_{n,p}(\mathsf{K}) > \delta\right) \leq \mathcal{B}(n,p), \tag{37}$$

where the bound $\mathcal{B}(n,p)$ satisfies $\lim_{n\to\infty}\mathcal{B}(n,p) = 0$ for any $p \geq 0$ and $\lim_{p\to\infty}\mathcal{B}(n,p) = 0$ for any $n \geq 1$. This clearly implies (15) and (16) and the result will follow from Proposition 2.7. We have

$$C_{n,p}(\mathsf{K}) \leq \sup_{l\geq n} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>l\}}\left|\sum_{j=n}^{l}\gamma_{p+j-1}\tilde{\epsilon}^{(1)}_j\right| + \sup_{l\geq n}\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>l\}}\left|\sum_{j=n}^{l}\gamma_{p+j-1}\tilde{\epsilon}^{(2)}_j\right|. \tag{38}$$

We start with the second term on the rhs of (38). By Doob's inequality and B2, for $N > n$,

$$\begin{aligned}
&\mathbb{P}^{(l)}_{x,\theta}\left(\sup_{n\leq l\leq N}\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>l\}}\left|\sum_{j=n}^{l}\gamma_{p+j-1}\tilde{\epsilon}^{(2)}_j\right| > \delta\right)\\
&\qquad\leq \delta^{-2}\mathbb{E}^{(l)}_{x,\theta}\left(\sum_{j=n}^{N}\gamma^2_{p+j-1}\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>j\}}\int\Phi^2_{\tilde{\theta}_j}(\tilde{X}_j,y)q^{(1)}_{\tilde{\theta}_j}(\tilde{X}_j,dy)\right)\\
&\qquad\leq C(\mathsf{K})\delta^{-2}\mathbb{E}^{(l)}_{x,\theta}\left(\sum_{j=n}^{N}\gamma^2_{p+j-1}\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>j\}}V^{2\eta}(\tilde{X}_j)\right)\\
&\qquad\leq C(\mathsf{K})\delta^{-2}\left(\gamma^2_{p+n}\mathbb{E}^{l}_{x,\theta}\left(\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>n-1\}}V^{2\eta+\alpha}(\tilde{X}_n)\right) + \sum_{j=n}^{N}\gamma^2_{p+j}\right).
\end{aligned}$$

It follows that

$$\mathbb{P}^{(l)}_{x,\theta}\left(\sup_{l\geq n}\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>l\}}\left|\sum_{j=n}^{l}\gamma_{p+j-1}\tilde{\epsilon}^{(2)}_j\right| > \delta\right) \leq C(\mathsf{K})\delta^{-p}\left(\gamma^2_{p+n}n^{2\eta+\alpha} + \sum_{j\geq n}\gamma^2_{p+j}\right). \tag{39}$$

To deal with the first term on the rhs of (38), we proceed as in the proof of Theorem 2.1. We consider the sequence $\{a_n,\ n \geq 0\}$ such that $a_n \propto n^{-\rho}$, $a_n \in (0, 1/2]$ where $\rho \in (0,1)$ is as in the statement of the Proposition. For $1 \leq n \leq l$ and $p \geq 0$, we introduce the partial sum

$$S_{n,l}(p,\mathsf{K}) \stackrel{\text{def}}{=} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>l\}}\sum_{j=n}^{l}\gamma_{p+j}\bar{\Upsilon}_{\tilde{\theta}_j}(\tilde{X}_j).$$

where $\bar{\Upsilon}_\theta(x) = \Upsilon_\theta(x) - h(\theta)$. Under B2, $\Upsilon_\theta$ admits an approximate Poisson equation $\tilde{g}_a$ for any $j \geq 1$ and we have $\bar{\Upsilon}_{\tilde{\theta}_j}(\tilde{X}_j) = (1-a_j)^{-1}\tilde{g}_{a_j}(\tilde{X}_j,\tilde{\theta}_j) - P_{\tilde{\theta}_j}\tilde{g}_{a_j}(\tilde{X}_j,\tilde{\theta}_j)$. Using this and following the same approach as in the proof of Theorem 2.1, we decompose $S_{n,l}(p,\mathsf{K})$ as

$$S_{n,l}(p,\mathsf{K}) = T^{(1)}_{n,l} + T^{(2)}_{n,l} + T^{(3)}_{n,l} + T^{(4)}_{n,l} + T^{(5)}_{n,l} + T^{(6)}_{n,l}$$

where

$$T^{(1)}_{n,l} = \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>l\}} \sum_{j=n}^{l} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>j\}} \gamma_{p+j} \left((1-a_j)^{-1}-1\right) \tilde{g}_{a_j}(\tilde{X}_j, \tilde{\theta}_j).$$

$$T^{(2)}_{n,l} = \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>n\}} \gamma_{p+n} \tilde{g}_{a_n}(\tilde{X}_n, \tilde{\theta}_n) - \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>l\}} \gamma_{p+l} P_{\tilde{\theta}_l} \tilde{g}_{a_l}(\tilde{X}_l, \tilde{\theta}_l).$$

$$T^{(3)}_{n,l} = \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>l\}} \sum_{j=n}^{l-1} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>j+1\}} \gamma_{p+j+1} \left(\tilde{g}_{a_{j+1}}(\tilde{X}_{j+1}, \tilde{\theta}_{j+1}) - \tilde{g}_{a_{j+1}}(\tilde{X}_{j+1}, \tilde{\theta}_j)\right).$$

$$T^{(4)}_{n,l} = \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>l\}} \sum_{j=n}^{l-1} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>j\}} \left(\gamma_{p+j+1} - \gamma_{p+j}\right) \tilde{g}_{a_{j+1}}(\tilde{X}_{j+1}, \tilde{\theta}_j).$$

$$T^{(5)}_{n,l} = \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>l\}} \sum_{j=n}^{l-1} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>j\}} \gamma_{p+j} \left(\tilde{g}_{a_j+1}(\tilde{X}_{j+1}, \tilde{\theta}_j) - \tilde{g}_{a_j}(\tilde{X}_{j+1}, \tilde{\theta}_j)\right).$$

$$T^{(6)}_{n,l} = \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>l\}} \sum_{j=n}^{l-1} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>j\}} \gamma_{p+j} \left(\tilde{g}_{a_j}(\tilde{X}_{j+1}, \tilde{\theta}_j) - P_{\tilde{\theta}_j} \tilde{g}_{a_j}(\tilde{X}_j, \tilde{\theta}_j)\right).$$

We deal with each of these terms using similar techniques as in the proofs of Theorem 2.1 and Theorem 2.3. Some of the details are thus omitted. Let $\delta > 0$ arbitrary.

`On Term` $T^{(1)}_{n,l}$. Take $\kappa > 1$ such that $\eta + \alpha\kappa < 1 - \alpha$. Then Proposition 3.2 yields $|\tilde{g}_{a_j}(\tilde{X}_j, \tilde{\theta}_j)| \le C(\mathsf{K}) V^{\eta+\alpha\kappa}(\tilde{X}_j)$ on $\{\tilde{\theta}_j \in \mathsf{K}\}$. Then by Markov's inequality, we have

$$\mathbb{P}^{(p)}_{x,\theta}\left(\sup_{l\ge n}\left|T^{(1)}_{n,l}\right| > \delta\right) \le \delta^{-1}\mathbb{E}^{(p)}_{x,\theta}\left(\sum_{j\ge n} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>j-1\}} \gamma_{p+j-1}\left((1-a_j)^{-1}-1\right)\left|\tilde{g}_{a_j}(\tilde{X}_j, \tilde{\theta}_j)\right|\right)$$
$$\le \delta^{-1} C(\mathsf{K}) V(x) \left(\gamma_{n+p} n^{1-\alpha-\rho} + \sum_{j\ge n} \gamma_{p+j} j^{-\rho}\right). \quad (40)$$

The last inequality uses Proposition 3.5 and Proposition 3.4 (i).

`On Term` $T^{(2)}_{n,l}$. Let $\epsilon > 0$, $\kappa > 1$ such that $\epsilon \in (\rho, (1-\alpha)(\eta+\kappa\alpha)^{-1} - 1)$. That is $(1+\epsilon)(\eta + \alpha\kappa) < 1 - \alpha$ and $\epsilon > \rho$. Then

$$\mathbb{P}^{(p)}_{x,\theta}\left(\sup_{l\ge n}\left|T^{(2)}_{n,l}\right| > \delta\right)$$
$$\le (2/\delta)^{1+\epsilon}\mathbb{E}^{(p)}_{x,\theta}\left(\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>n\}} \gamma^{1+\epsilon}_{p+n}\left|\tilde{g}_{a_n}(\tilde{X}_n, \tilde{\theta}_n)\right|^{1+\epsilon} + \sum_{l\ge n} \gamma^{1+\epsilon}_{p+l} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>l\}} \left|P_{\tilde{\theta}_l}\tilde{g}_{a_l}(\tilde{X}_l, \tilde{\theta}_l)\right|^{1+\epsilon}\right)$$
$$\le (2/\delta)^{1+\epsilon} C(\mathsf{K}) V(x) \left(\gamma^{1+\epsilon}_{p+n} n^{1-\alpha} + \sum_{j\ge n-1} \gamma^{1+\epsilon}_{p+j}\right). \quad (41)$$

`On Term` $T^{(3)}_{n,l}$. Take $\kappa > 1$ and $\delta > 0$ such that $2\eta+\alpha(\kappa+\delta) < 1-\alpha$ and $\eta+\alpha(\kappa+\delta) < 1/2$. By Proposition 3.3 and B2 $|\tilde{g}_a(x,\theta) - \tilde{g}_a(x,\theta')| \leq C(\mathsf{K}) \sup_{\theta\in\mathsf{K}} |\Upsilon_\theta|_{V^\eta} \zeta_\delta(a) \, |\theta-\theta'| \, V^{\eta+\alpha(\kappa+\delta)}(x)$. Then by Markov's inequality

$$\mathbb{P}^{(p)}_{x,\theta}\left(\sup_{\geq n}\left|T^{(3)}_{n,l}\right| > \delta\right) \leq (1/\delta)\mathbb{E}^{(p)}_{x,\theta}\left(\sum_{j\geq n} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>j\}}\gamma^2_{p+j+1}\zeta_\delta(a_j)\left|\Phi_{\tilde{\theta}_j}(\tilde{X}_j, Y_{j+1})\right| V^{\eta+\alpha(\kappa+\delta)}(\tilde{X}_{j+1})\right).$$

From B2 and the structure of the algorithm we compute that

$$\mathbb{E}^{(p)}_{x,\theta}\left(\left|\Phi_{\tilde{\theta}_j}(\tilde{X}_j, Y_{j+1})\right| V^{\eta+\alpha(\kappa+\delta)}(\tilde{X}_{j+1})|\mathcal{F}_j\right)\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>j\}} \leq C(\mathsf{K}) V^{2\eta+\alpha(\kappa+\delta)}(\tilde{X}_j).$$

It follows

$$\mathbb{P}^{(p)}_{x,\theta}\left(\sup_{\geq n}\left|T^{(3)}_{n,l}\right| > \delta\right) \leq (1/\delta)C(\mathsf{K})\left(\gamma^2_{p+n-1}n^{1+\rho-\alpha} + \sum_{j\geq n}\gamma^2_{p+j-1}j^\rho\right)V(x). \tag{42}$$

`On Term` $T^{(4)}_{n,l}$. By Markov's inequality,

$$\begin{aligned}\mathbb{P}^{(p)}_{x,\theta}\left(\sup_{l\geq n}\left|T^{(4)}_{n,l}\right| > \delta\right) &\leq (1/\delta)\mathbb{E}^{(p)}_{x,\theta}\left(\sum_{j\geq n}(\gamma_{p+j}-\gamma_{p+j+1})\,\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>j\}}\left|\tilde{g}_{a_{j+1}}(\tilde{X}_{j+1},\tilde{\theta}_j)\right|\right)\\ &\leq (1/\delta)C(\mathsf{K})\mathbb{E}^{(p)}_{x,\theta}\left(\sum_{j\geq n}(\gamma_{p+j}-\gamma_{p+j+1})\,\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>j\}}V^{1-\alpha}(\tilde{X}_{j+1})\right)\\ &\leq (1/\delta)C(\mathsf{K})V(x)\left(n^{1-\alpha}(\gamma_{p+n}-\gamma_{p+n+1})+\gamma_{p+n}\right).\end{aligned} \tag{43}$$

`On Term` $T^{(5)}_{n,l}$. Take $\kappa \in (1,2)$ such that $\eta+\alpha\kappa < 1-\alpha$. One can check as in Proposition 3.3 that for any compact $\mathsf{K}$ $|P_\theta\tilde{g}_a(x,\theta) - P_\theta\tilde{g}_{a'}(x,\theta)| \leq C(\mathsf{K})|a-a'|a^{\kappa-2}V^{\eta+\alpha\kappa}(x)$. And for $a_j \propto j^{-\rho}$, $|a_j - a_{j-1}|a_j^{\kappa-2} \propto j^{-1}a_j^{\kappa-1} = o(j^{-1})$. Hence, by Markov's inequality, we get:

$$\mathbb{P}^{(p)}_{x,\theta}\left(\sup_{l\geq n}\left|T^{(5)}_{n,l}\right| > \delta\right) \leq \delta^{-1}C(\mathsf{K})V(x)\left(n^{-\alpha}\gamma_{p+n} + \sum_{j\geq n}\gamma_{p+j}j^{-1}\right). \tag{44}$$

`On Term` $T^{(6)}_{n,l}$. Let $\kappa > 1$ such that $2(\eta+\alpha\kappa/2) < 1-\alpha$. Consider the term $D_j = \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>j\}}\gamma_{p+j}\left(\tilde{g}_{a_j}(\tilde{X}_{j+1},\tilde{\theta}_j) - P_{\tilde{\theta}_j}\tilde{g}_{a_j}(\tilde{X}_j,\tilde{\theta}_j)\right)$ so that $T^{(2)}_{n,l} = \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>l\}}\sum_{j=n}^{l-1}D_j$. We note that $D_j$ is a martingale difference and by Doob's inequality we get:

$$\begin{aligned}\mathbb{P}^{(p)}_{x,\theta}\left(\sup_{l\geq n}\left|T^{(6)}_{n,l}\right| > \delta\right) &\leq (1/\delta)^2\sum_{j\geq n}\mathbb{E}^{(l)}_{x,\theta}\left(|D_j|^2\right)\\ &\leq (1/\delta)^2C(\mathsf{K})V(x)\left(\gamma^2_{p+n-1}n^{1-\alpha+\rho} + \sum_{j\geq n}\gamma^2_{p+j-1}j^\rho\right).\end{aligned} \tag{45}$$

By combining (39)-(45) and (18), we get (37) as claimed. □

3.10. **Proof of the results of Section 2.7.**

3.10.1. *Proof of Proposition 2.9.* The function $a(\theta)$ is of class $\mathcal{C}^1$. Hence by Assumption C1 and the Mean Value Theorem $\mathcal{L} = \{\theta \in \mathbb{R}: \ a(\theta) = \bar{\alpha}\}$ is not empty. It also follows from C1 that the function $\theta \to \int_0^\theta \cosh(u)(\bar{\alpha} - a(u))du$ is bounded from below; so we can find $K_1$ such that $w(\theta) = \int_0^\theta \cosh(u)(\bar{\alpha} - a(u))du + K_1 \geq 0$. Moreover $(a(u) - \bar{\alpha})w'(\theta) = -\cosh(\theta)(a(\theta) - \bar{\alpha})^2 \leq 0$ with equality iif $\theta \in \mathcal{L}$. By Sard's theorem $w(\mathcal{L})$ has an empty interior. Again from C1, it follows that $\mathcal{L}$ is included in a bounded interval of $\mathbb{R}$ and since $\lim_{\theta\to\pm\infty} w(\theta) = \infty$, we can find $M_0$ such that $\mathcal{L} \subset \{\theta \in \mathbb{R}: \ w(\theta) < M_0\}$ and $\mathcal{W}_M$ is bounded thus compact for any $M > 0$.

3.10.2. *Proof of Proposition 2.11.* A straightforward calculation using the boundedness of $|\nabla \log \pi(x)|$ implies that for any $\theta \in \mathsf{K}$,

$$\left|\frac{\partial}{\partial\theta} \log\left(\alpha_\theta(x,y)q_\theta(x,y)\right)\right| \leq C(\mathsf{K})\left(1 + |y-x|^2\right),$$

for some finite constant $C(\mathsf{K})$. It follows that

$$\int \left|\frac{\partial}{\partial\theta}\left(\alpha_\theta(x,y)q_\theta(x,y)\right)f(y)\right| dy \leq C(\mathsf{K})|f|_{V_s^\beta} \int \left(1 + |y-x|^2\right) V_s^\beta(y) q_\theta(x,y)dy.$$

We do a change of variable $y = b(x) + e^{\theta/2}z$, where $b(x) = x + 0.5e^\theta \nabla \log \pi(x)$ and using the boundedness of $|\nabla \log \pi(x)|$, we get:

$$\sup_{\theta\in\mathsf{K}} \int \left|\frac{\partial}{\partial\theta}\left(\alpha_\theta(x,y)q_\theta(x,y)\right)f(y)\right| dy \leq C(\mathsf{K})|f|_{V_s^\beta} V_s^\beta(x) \int \left(1+|z|^2\right)^{\beta s/2} g(z)dz,$$

where $g$ is the density of the mean zero $d$-dimensional Gaussian distribution with covariance matrix $I_d$. The stated result follows by an application of the Mean Value Theorem.

**Acknowledgment:** We are grateful to Michael Woodroofe for helpful discussions on martingale approximation techniques and to Shukri Osman for helpful conversations.

## References

Andrieu, C. and Moulines, É. (2006). On the ergodicity properties of some adaptive MCMC algorithms. *Ann. Appl. Probab.* **16** 1462–1505.

Andrieu, C., Moulines, É. and Priouret, P. (2005). Stability of stochastic approximation under verifiable conditions. *SIAM J. Control Optim.* **44** 283–312 (electronic).

Andrieu, C. and Thoms, J. (2008). A tutorial on adaptive MCMC. *Statistics and Computing* **18** 343–373.

Atchade, Y. and Fort, G. (2008). Limit theorems for some adaptive mcmc algorithms with sub-geometric kernels. *Bernoulli (To appear)* .

Atchade, Y. F., Fort, G., Moulines, E. and Priouret, P. (2009). Adaptive Markov Chain Monte Carlo: Theory and methods. Tech. rep., Univ. of Michigan.

Atchade, Y. F. and Rosenthal, J. S. (2005). On adaptive markov chain monte carlo algorithm. *Bernoulli* **11** 815–828.

Bai, Y. (2008). The simultaneous drift conditions for Adaptive Markov Chain Monte Carlo algorithms. Tech. rep., Univ. of Toronto. (personal communication).

Baxendale, P. H. (2005). Renewal theory and computable convergence rates for geometrically ergodic markov chains. *Annals of Applied Probability* **15** 700–738.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit theory and its application.* Academic Press, New York.

Holzmann, H. (2005). Martingale approximations for continuous-time and discrete-time stationary markov processes. *Stochastic processes and their applications* **115** 1518–1539.

Jarner, S. F. and Roberts, G. O. (2002a). Convergence of heavy tailed mcmc algorithms. *MCMC Preprints* .

Jarner, S. F. and Roberts, G. O. (2002b). Polynomial convergence rates of Markov chains. *Ann. Appl. Probab.* **12** 224–247.

Kamatani, K. (To appear). Metropolis-Hastings algorithms with acceptance ratios of nearly 1. *Annals of the Institute of Statistical Mathematics* .

Kipnis, C. and Varadhan, S. R. S. (1986). Central limit theorem for additive functionals of reversible markov processes and applications to simple exclusions. *Comm. Math. Phys.* **104** 1–19.

Maxwell, M. and Woodroofe, M. (2000). Central limit theorems for additive functional of markov chains. *Annals of Probability* **28** 713–724.

Roberts, G. and Rosenthal, J. (2006). Examples of adaptive MCMC. Tech. rep., Univ. of Toronto. To appear in J. Computational Graphical Statistics.

Roberts, G. and Tweedie, R. (1996). Exponential convergence of langevin distributions and their discrete approximations. *Bernoulli* **2** 341–363.

Roberts, G. O. and Rosenthal, J. S. (2001). Optimal scaling of various metropolis-hastings algorithms. *Statistical Science* **16**.

Rosenthal, J. S. and Roberts, G. O. (2007). Coupling and ergodicity of adaptive mcmc. *Journal of Applied Probablity* **44** 458–475.

Saksman, E. and Vihola, M. (2009). On the ergodicity of the adaptive metropolis algorithm on unbounded domains. *Technical report arXiv:0806.2933v2* .