Bernoulli **18**(3), 2012, 975–1001
DOI: 10.3150/11-BEJ360

Limit theorems for some adaptive MCMC algorithms with subgeometric kernels: Part II

YVES F. ATCHADÉ[1] and GERSENDE FORT[2]

[1]*University of Michigan, 1085 South University, Ann Arbor, MI 48109, USA. E-mail: yvesa@umich.edu*
[2]*LTCI, CNRS-TELECOM ParisTech, 46 rue Barrault, 75634 Paris Cedex 13, France.*
E-mail: gfort@tsi.enst.fr

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.

Keywords: adaptive Markov chain Monte Carlo; Markov chain; Metropolis adjusted Langevin algorithms; stochastic approximations; subgeometric ergodicity

1. Introduction

This work is a sequel of Atchadé and Fort [4] and develops central limit theorems for adaptive MCMC (AMCMC) algorithms. As a tool for quantifying the fluctuations of Monte Carlo estimates, central limit theorems play an important role in the practice of Monte Carlo simulation. Previous work on the subject include Andrieu and Moulines [1] and Saksman and Vihola [20] where central limit theorems are obtained 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 [12].

We consider adaptive MCMC algorithms driven by Markov kernels $\{P_\theta, \theta \in \Theta\}$ such that each kernel P_θ enjoys a polynomial rate of convergence towards π and satisfies a drift condition of the form $P_\theta V \le V - cV^{1-\alpha} + b$ for some $\alpha \in (0, 1]$ (uniformly in θ 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 is known for Markov chains under similar conditions. Indeed, Jarner and Roberts [11] that irreducible and aperiodic Markov chains for which the drift condition $PV \le V - cV^{1-\alpha} + b\mathbb{1}_{\mathcal{C}}$ hold for some small set $\mathcal{C}$ satisfy a central limit theorem when $\alpha \le 1/2$. The proof of our results is based on a martingale approximation technique developed by Kipnis and Varadhan [14] and Maxwell and Woodroofe [16] in the Markovian setting. 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 Atchadé and Fort [4] to study the strong law of large numbers for AMCMC.

1350-7265

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 for example, in [8,18]. For specific examples and a review of the methodological developments see, for example, Andrieu and Thoms [3], Atchade *et al.* [6].

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). The proofs are postponed to Section 3. We omit some of the technical details which can be found in a longer version of the paper [5] available from the Arxiv.

2. Statement of the results

2.1. Notations

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

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

where $\delta_x(\mathrm{d}t)$ 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 σ-finite measures μ on $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ via $P^n f(\cdot) \stackrel{\text{def}}{=} \int P^n(\cdot, \mathrm{d}y) f(y)$ and $\mu P^n(\cdot) \stackrel{\text{def}}{=} \int \mu(\mathrm{d}x) P^n(x, \cdot)$.

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 measurable functions $f : \mathsf{X} \to \mathbb{R}$ with finite V-norm is denoted by $\mathcal{L}_V$.

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

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

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 $\|\cdot\|_{\mathrm{TV}}$ 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 X be a general state space endowed with a countably generated σ-field $\mathcal{X}$. Let Θ be an open subspace of $\mathbb{R}^q$ the q-dimensional Euclidean space and $\mathcal{B}(\Theta)$ its Borel σ-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_θ admits an invariant distribution π. Let $\{\mathsf{K}_n, n \geq 0\}$ be a family of nonempty compact subspaces of Θ 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 reprojection 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 $\bar{R}(n; \cdot, \cdot) : (\mathsf{X} \times \Theta) \times (\mathcal{X} \times \mathcal{B}(\Theta)) \to [0, 1]$ be a sequence of Markov kernels on $\mathsf{X} \times \Theta$ with the following property. For any $n \geq 0$, $A \in \mathcal{X}$, $(x, \theta) \in \mathsf{X} \times \Theta$

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

In practice, the kernel $\bar{R}(n; \cdot)$ is commonly designed using stochastic approximation. We give detailed examples in Section 2.6. 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 $\bar{P}$,

$$\begin{aligned}
&\bar{P}((x, \theta, \nu, \xi), (\mathrm{d}x', \mathrm{d}\theta', \mathrm{d}\nu', \mathrm{d}\xi')) \\
&\quad \stackrel{\text{def}}{=} \bar{R}\big(\nu + \xi; \Pi_\xi(x, \theta), (\mathrm{d}x', \mathrm{d}\theta')\big) \\
&\qquad \times \big(\mathbb{1}_{\{\theta' \in \mathsf{K}_\nu\}} \delta_\nu(\mathrm{d}\nu') \delta_{\xi+1}(\mathrm{d}\xi') + \mathbb{1}_{\{\theta' \notin \mathsf{K}_\nu\}} \delta_{\nu+1}(\mathrm{d}\nu') \delta_0(\mathrm{d}\xi')\big).
\end{aligned} \tag{2}$$

Algorithmically, this Markov chain can be described as follows.

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

(a) *generate* $(X_{n+1}, \theta_{n+1}) \sim \bar{R}(\nu_n + \xi_n; \Pi_{\xi_n}(X_n, \theta_n), \cdot)$;
(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* $\xi_{n+1} = 0$.

The dynamics of the algorithm is simple to describe. Assume $\nu_0 = 0, \xi_0 = 0$. As long as $\theta_k \in \mathsf{K}_0$, (X_k, θ_k) is updated using $\bar{R}(k-1; (X_{k-1}, \theta_{k-1}), \cdot)$, $k \geq 1$. If $\theta_{n_0} \notin \Theta_0$, we restart the algorithm: we set K_1 as the new reference compact set and as long as $\theta_k \in \mathsf{K}_1$ $(k \geq n_0 + 1)$, (X_k, θ_k) is updated using $\bar{R}(1 + (k - n_0); (X_{k-1}, \theta_{k-1}), \cdot)$, $k \geq n_0 + 1$; etc...

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 simply 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. The adaptive MCMC algorithm described above is more involved than the algorithm analyzed in Atchadé and Fort [4]. This is due to the fact that in studying the central limit theorem, unavoidably, we need to study the limiting behavior of the adaptation parameter θ_n. When the kernel $\bar{R}$, is based on stochastic approximation, as in most examples, proving the convergence of θ_n under simple assumptions is a rather difficult problem. At issue is the stability of the stochastic approximation algorithm. In order to develop general and easily verifiable results, one has to recourse, as above, to various reprojection tricks. This is a well known issue in the stochastic approximation literature (see, e.g., [2,15]). Algorithm 2.1 encompasses the two main reprojection strategies used in practice to control stochastic approximation algorithms. These reprojection tricks also serve the purpose of building algorithms that are less sensitive to the initial conditions and step-size selected.

(1) For example, Algorithm 2.1 reduces to the framework of re-projections on randomly varying compact sets developed in [2] if we take $\{\mathsf{K}_n, n \geq 0\}$ 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 K of Θ. And we then obtain another commonly used approach where the reprojection is done on a fixed compact set K. See, for example, Kushner and Yin [15] and in the context of AMCMC, see Atchade and Rosenthal [7].

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 $f : \mathsf{X} \to \mathbb{R}$,

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

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

$$P_l(n; (x_1, \theta_1), (\mathrm{d}x', \mathrm{d}\theta')) = \bar{R}\big(l + n; (x_1, \theta_1), (\mathrm{d}x', \mathrm{d}\theta')\big).$$

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, θ, l)). Again it follows from (1) that for any bounded measurable function $f : \mathsf{X} \to \mathbb{R}$,

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

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

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

with the usual convention that $\inf \varnothing = \infty$. Clearly, the two processes defined above are closely related. We will refer to $\{(\tilde{X}_n, \tilde{\theta}_n), n \geq 0\}$ as the *reprojection free process*. The general strategy that we adopt to study the Markov chain $\{(X_n, \theta_n, \nu_n, \xi_n), n \geq 0\}$ (adapted from [2]) consists in first studying the reprojection 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 K of Θ, there exists $b, c \in (0,\infty)$ (that depend on K) such that for any $(x,\theta) \in \mathsf{X} \times \mathsf{K}$,

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

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

$$(n+1)^\kappa \| P_\theta^n(x,\cdot) - \pi(\cdot)\|_{V^\beta} \le C\ V^{\beta+\alpha\kappa}(x), \qquad n \ge 0. \tag{6}$$

Notice that (5)–(6) imply that $\pi(V^{1-\alpha}) < \infty$. We will also assume that the number of reprojection is finite.

A2.

$$\check{\mathbb{P}}\Bigl(\sup_{n\ge 0} \nu_n < \infty\Bigr) = 1. \tag{7}$$

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

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

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 K of Θ. Suppose also that there exist $\varepsilon > 0$, $\kappa > 0$, $\beta + \alpha\kappa < 1-\alpha$ such that for any $(x,\theta,l) \in \mathsf{X}_0 \times \Theta_0 \times \mathbb{N}$*

$$\mathbb{E}_{x,\theta}^{(l)}\Biggl[\sum_{k\ge 1} k^{-1+\varepsilon}\bigl(D_\beta(\tilde\theta_k,\tilde\theta_{k-1}) + |f_{\tilde\theta_k} - f_{\tilde\theta_{k-1}}|_{V^\beta}\bigr)\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > k\}} V^{\beta+\alpha\kappa}(\tilde X_k)\Biggr] < \infty. \tag{8}$$

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 3.4. □

Remark 2. A strong law of large numbers also hold under similar assumptions [4], Theorem 2.3. The main difference with Atchadé and Fort [4] is that in the result above, the function of interest is allowed to depend on θ. The summability condition (8) seems complicated but is not hard to check in practice. See, for example, Section 2.5.3. We also point out the fact that (8) is expressed in terms of the non-homogeneous Markov chain $\{(\tilde X_n, \tilde\theta_n), n \ge 0\}$ which is much easier to handle than the actual adaptive MCMC process.

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

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

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 $P_\theta g_a(x,\theta) \stackrel{\text{def}}{=} \int P_\theta(x, \mathrm{d}z) g_a(z,\theta)$.

For the CLT, we need the adaptation parameter θ_n to converge to a limit.

A3. There exists a Θ-valued random variable $\theta_\star$ such that with $\check{\mathbb{P}}$-probability one, $\{\theta_n, n \geq 0\}$ 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. We introduce the nonnegative random variable

$$\sigma_\star^2(f) \stackrel{\text{def}}{=} \int \pi(\mathrm{d}x)\{2f(x)g(x,\theta_\star) - f^2(x)\}.$$

Theorem 2.2. *Assume A1–A3 with $\alpha < 1/2$. Let $\beta \in [0, \frac{1}{2} - \alpha)$ and $f \in \mathcal{L}_{V^\beta}$ such that $\pi(f) = 0$. Suppose that there exist $\kappa > 1$, $\varepsilon > 0$, $\rho \in (\frac{1}{2}, \frac{1}{2(1-\varepsilon)})$, such that $2\beta + \alpha(\kappa+\varepsilon) < 1 - \alpha$, and for any $b \in [0, 1-\alpha]$, any $(x,\theta,l) \in \mathsf{X}_0 \times \Theta_0 \times \mathbb{N}$*

$$\mathbb{E}_{x,\theta}^{(l)}\left[\sum_{k\geq 1} k^{-1/2+\rho(1-\varepsilon)} D_b(\tilde{\theta}_k, \tilde{\theta}_{k-1}) \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > k\}} V^{2\beta+\alpha(\kappa+\varepsilon)}(\tilde{X}_k)\right] < \infty. \tag{11}$$

Suppose also that for a sequence $a_n \propto n^{-\rho}$, $a_n \in (0, \frac{1}{2}]$,

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

Then $n^{-1/2} \sum_{k=1}^n f(X_k)$ converges weakly to $\sqrt{\sigma_\star^2(f)} Z$ where $Z \sim \mathcal{N}(0,1)$ is a standard normal random variable that is independent of $\sigma_\star^2(f)$.

Proof. See Section 3.5. □

2.4. On assumption (12)

Assumption (12) is the most difficult assumption to check in Theorem 2.2. This assumption is needed to establish the weak law of large numbers in the CLT. For stationary Markov chains (12) automatically holds. In the general adaptive case, the simplest approach to checking (12) is through appropriate moments condition.

Proposition 2.3. *Assume* A1 *and* A3 *with* $\alpha < 1/2$ *and let* $\beta \in [0, \frac{1}{2} - \alpha)$. *Suppose that there exists* $\varepsilon > 0$ *such that for any* $(x, \theta, l) \in \mathsf{X}_0 \times \Theta_0 \times \mathbb{N}$

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

Then (12) *hold.*

By Proposition 3.4(ii) below, one can always check (13) if $\alpha < 1/3$ and $\beta \in [0, 1 - 3\alpha)$.

Corollary 2.4. *Assume* A1–A3 *with* $\alpha < 1/3$. *Let* $\beta \in [0, 1 - 3\alpha)$ *and* $f \in \mathcal{L}_{V^\beta}$. *Suppose that* (11) *holds. Then* $n^{-1/2} \sum_{k=1}^{n} f(X_k)$ *converges weakly to* $\sqrt{\sigma_\star^2(f)} Z$ *where* $Z \sim \mathcal{N}(0, 1)$ *is a standard normal random variable that is independent of* $\sigma_\star^2(f)$.

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 K of Θ, P_θ 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 Atchadé and Fort [4] (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 of Andrieu and Moulines [1]. Indeed, suppose that uniformly over compact subsets K of Θ, there exist $\mathcal{C} \in \mathcal{X}$, ν a probability measure on $(\mathsf{X}, \mathcal{X})$, $b, \varepsilon > 0$ and $\lambda \in (0, 1)$ such that $\nu(\mathcal{C}) > 0$, $P_\theta(x, \cdot) \geq \varepsilon\nu(\cdot)\mathbb{1}_{\mathcal{C}}(x)$ 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., [9]), 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 $\varepsilon > 0$ such that $2(\beta + \alpha) + \varepsilon < 1$, and since V^δ-moment of geometrically ergodic adaptive MCMC are bounded in n for any $\delta \in [0, 1)$, we get (13).

2.5.2. *On Assumptions 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 $D_b(\theta_k, \theta_{k-1}) \le C\gamma_k V^\eta(X_k)$ for some positive numbers γ_k and $\eta \ge 0$. then we can check (8) or (11) using Proposition 3.5.

2.6. Checking A3 for AMCMC driven by stochastic approximation

One specific form of the kernel $\bar{R}(n; \cdot)$ used in practice can be described as follows. 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

$$\int q_\theta^{(1)}(x, \mathrm{d}y) q_\theta^{(2)}((x, y), \mathrm{d}x') = P_\theta(x, \mathrm{d}x').$$

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 $\Phi(\theta, x, y)$. Set

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

where $\{\gamma_n\}$ is a sequence of positive numbers. Thus, (1) holds. Under (14), Step (a) of Algorithm 2.1 can be described as follows. Given $(\bar{X}_n, \bar{\theta}_n) = \Pi_{\xi_n}(X_n, \theta_n)$, generate $Y_{n+1} \sim q_{\bar{\theta}_n}^{(1)}(\bar{X}_n, \cdot)$ and $X_{n+1} \sim q_{\bar{\theta}_n}^{(2)}(\bar{X}_n, Y_{n+1}, \cdot)$. Then set

$$\theta_{n+1} = \theta_n + \gamma_{\nu_n + \xi_n} \bigl(h(\theta_n) + \varepsilon_{n+1}^{(1)} + \varepsilon_{n+1}^{(2)} \bigr),$$

where $\varepsilon_{n+1}^{(1)} = \Upsilon_{\theta_n}(\bar{X}_n) - h(\bar{\theta}_n)$, $\varepsilon_{n+1}^{(2)} = \Phi_{\bar{\theta}_n}(\bar{X}_n, Y_{n+1}) - \Upsilon_{\bar{\theta}_n}(\bar{X}_n)$ and

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

Most adaptive MCMC algorithms used in practice are special cases of (14). For example, in Metropolis–Hastings type algorithms, $q_\theta^{(1)}$ corresponds to the proposal kernel and $q_\theta^{(2)}$ corresponds to the acceptance probability (see Section 2.7 for a specific example). Also, with the choices $q_\theta^{(1)}(x, \mathrm{d}y) = P_\theta(x, \mathrm{d}y)$, $\Phi_\theta(x, y) = \Phi_\theta(y)$, and $q_\theta^{(2)}((x, y), \mathrm{d}x') = \delta_y(\mathrm{d}x')$, one obtains the stochastic approximation dynamics analyzed in Andrieu and Moulines [1].

We assume the following.

B1. (1) $\{\mathsf{K}_n, n \geq 0\}$ is such that $\Theta = \bigcup_n \mathsf{K}_n$, $\Theta_0 \subseteq \mathsf{K}_0$ and $\mathsf{K}_n \subset \mathrm{int}(\mathsf{K}_{n+1})$, where $\mathrm{int}(A)$ 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 $\mathcal{L} \stackrel{\mathrm{def}}{=} \{\theta \in \Theta\colon \langle \nabla w(\theta), h(\theta)\rangle = 0\}$ is nonempty and the closure of $w(\mathcal{L})$ has an empty interior. In the above, ∇w denote the gradient of w.

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

Assume that the function Υ satisfies

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

$$\sup_{\theta \in \mathsf{K}} \sup_{x \in \mathsf{X}} V^{-2\eta}(x) \int q_\theta^{(1)}(x, \mathrm{d}y) |\Phi_\theta(x, y)|^2 < \infty, \quad \text{and} \\ D_b(\theta, \theta') + |\Upsilon_\theta - \Upsilon_{\theta'}|_{V^\eta} \leq C|\theta - \theta'| \tag{15}$$

for some finite constant C that depends possibly on K.

Proposition 2.5. *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* $p \geq 0$,

$$\lim_{n\to\infty} (\gamma_{p+n-1} - \gamma_{p+n}) n^{1-\alpha} = 0 \quad \text{and} \quad \sum_{n \geq 1} (\gamma_k^2 k^\rho + \gamma_k k^{-\rho} + \gamma_k^{1+\rho}) < \infty \tag{16}$$

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

Proof. See Section 3.6. □

2.7. Example: Adaptive Langevin algorithms

We illustrate the theory above with an application to the Metropolis-adjusted Langevin algorithm (MALA). In this section, X is the d-dimensional Euclidean space $\mathbb{R}^d$ and π is a positive density on X with respect to the Lebesgue (denoted μ_{Leb} or $\mathrm{d}x$). The MALA algorithm is an effective Metropolis–Hastings algorithm whose proposal kernel is obtained by discretization of the Langevin diffusion

$$\mathrm{d}X_t = \tfrac{1}{2} \mathrm{e}^\theta \nabla \log \pi(X_t)\, \mathrm{d}t + \mathrm{e}^\theta\, \mathrm{d}B_t, \qquad X_0 = x,$$

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 $\mathrm{e}^\theta I_d$ where

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

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 [19], Roberts and Rosenthal [17]. In practice, the performance of this algorithm depends on the choice of the scale parameter θ. 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_θ the transition kernel of the MALA algorithm with proposal q_θ. We also introduce the functions

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

Let $\{\mathsf{K}_n, n \geq 0\}$ be a family of nonempty compact intervals of Θ such that $\bigcup \mathsf{K}_n = \mathbb{R}$, $\mathsf{K}_n \subset \text{int}(\mathsf{K}_{n+1})$. 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 reprojection 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* $Y_{n+1} \sim q_{\bar{\theta}}(\bar{X}, \cdot)$. *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*

$$\theta_{n+1} = \bar{\theta} + \frac{1}{1 + \nu_n + \xi_n} \big(\alpha_{\bar{\theta}}(\bar{X}, Y_{n+1}) - \bar{\alpha}\big). \tag{17}$$

(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

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

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

Assumption C1 is slightly redundant. It is easy to check that $\lim_{\theta\to-\infty} a(\theta) = 1$. But the second part of the assumption, although realistic, depends in general on $\nabla \log \pi$ and $\nabla^2 \log \pi$.

Proposition 2.6. *Under C1, the function $h(\theta) = a(\theta) - \bar{\alpha}$ satisfies B1(2) with $\mathcal{L} = \{\theta \in \mathbb{R}\colon a(\theta) = \bar{\alpha}\}$ and $w(\theta) = \int_0^\theta \cosh(u)(\bar{\alpha} - a(u))\,\mathrm{d}u + K$ for some finite constant K where $\cosh(u) = (\mathrm{e}^u + \mathrm{e}^{-u})/2$ is the hyperbolic cosine.*

Proof. See Section 3.7.1. □

We assume that the target density π is heavy tailed as in Kamatani [13].

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

$$\limsup_{|x|\to\infty} \langle x, \nabla \log \pi(x)\rangle \le -\eta, \qquad \lim_{|x|\to\infty} |\nabla \log \pi(x)| = 0, \qquad \lim_{|x|\to\infty} \|\nabla^2 \log \pi(x)\| = 0, \tag{18}$$

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

The next proposition is a paraphrase of Theorem 5 of Kamatani [13].

Proposition 2.7. *Assume C2. For $s \in (2, 2+\eta-d)$, define $V_s(x) = (1+|x|^2)^{s/2}$ and $\alpha = 2/s$. Let $\mathcal{C}$ be a compact subset of $\mathbb{R}^d$ with $\mu_{\mathrm{Leb}}(\mathcal{C}) > 0$. For any compact subset K of Θ, there exists $\varepsilon, c, b \in (0,\infty)$, such that*

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

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

For the smoothness, we have the following proposition.

Proposition 2.8. *Assume that $|\nabla \log \pi(x)|$ is a bounded function. Let K be a compact convex subset of Θ. There exists a finite constant $C(\mathsf{K})$ such that for any $f \in \mathcal{L}_{V_s^\beta}$, $\beta \in [0,1]$, any $\theta, \theta' \in \mathsf{K}$,*

$$\left|\int \alpha_\theta(x,y)q_\theta(x,y)f(y)\,\mathrm{d}y - \int \alpha_{\theta'}(x,y)q_{\theta'}(x,y)f(y)\,\mathrm{d}y\right| \le C(\mathsf{K})|f|_{V_s^\beta}|\theta-\theta'|V_s^\beta(x). \tag{19}$$

Proof. See Section 3.7.2. □

We now apply Theorem 2.2 to get a CLT for the adaptive MALA.

Theorem 2.9. *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* $\sigma_\star^2(f)$ *such that* $n^{-1/2} \sum_{k=1}^n f(X_k)$ *converges weakly to a random variable* $\sqrt{\sigma_\star^2(f)} Z$, *where* $Z \sim N(0, 1)$ *is independent of* $\sigma_\star^2(f)$.

Remark 3. If π 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 2.9 guarantees a CLT for $\nu > 4$. Compare with $\nu > 2$ for Harris recurrent Markov chains satisfying A1.

Proof of Theorem 2.9. A1 holds as a consequence of Proposition 2.7 (see, e.g., [4], Section 2.4 and Appendix A). Proposition 2.6 shows that B1(2) holds and Proposition 2.8 implies that B2 holds. Therefore, A3 holds as a consequence of Proposition 2.5. (11) is an easy consequence of Proposition 2.8 and Proposition 3.5. We thus conclude with Corollary 2.4. □

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

$$\sigma_\theta^2(f) \stackrel{\text{def}}{=} \int \pi(\mathrm{d}x) \left\{ f^2(x) + 2 \sum_{k \ge 0} f(x) P_\theta^k f(x) \right\}.$$

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

3. Proofs

Throughout the proof, $C(\mathsf{K})$ denotes a finite constant that depends on the compact set 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, K is a given compact subset of Θ 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_θ as

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

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

$$f_\theta(x) = (1-a)^{-1}\tilde{g}_a(x,\theta) - P_\theta \tilde{g}_a(x,\theta). \tag{20}$$

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

$$f_\theta(x) = \tilde{g}(x,\theta) - P_\theta \tilde{g}(x,\theta). \tag{21}$$

For $a > 0$, we introduce the function

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

where $\Gamma(x) := \int_0^\infty u^{x-1}\mathrm{e}^{-u}\,\mathrm{d}u$ is the Gamma function. We will need the following technical lemma. For lack of space, we omit the details.

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

$$\sum_{j\geq 0}(1-a)^{j+1}(1+j)^{-\kappa} \leq \zeta_\kappa(a).$$

Proposition 3.2. *Assume* A1.

(i) *The function $\tilde{g}_a$ is well defined and there exists a finite constant C such that for any $\kappa \in [0, \alpha^{-1}(1-\beta)-1]$, $(x,\theta) \in \mathsf{X}\times\Theta$ and any $a \in (0,1/2]$*

$$|\tilde{g}_a(x,\theta)| \leq C|f_\theta|_{V^\beta}\zeta_\kappa(a)V^{\beta+\alpha\kappa}(x). \tag{22}$$

(ii) *Suppose that $\alpha < 1/2$. Then the function $\tilde{g}(x,\theta)$ is well-defined and there exists a finite constant C such that for any $\kappa \in (1, \alpha^{-1}(1-\beta)-1]$, $(x,\theta) \in \mathsf{X}\times\Theta$ and any $a \in (0,1/2]$*

$$|\tilde{g}_a(x,\theta) - \tilde{g}(x,\theta)| \leq C|f_\theta|_{V^\beta}\left(\int_0^a \zeta_{\kappa-1}(u)\,\mathrm{d}u\right)V^{\beta+\alpha\kappa}(x). \tag{23}$$

Proof. Follows from A1 and Lemma 3.1. □

Remark 4. One can check using Lemma 3.1 that for $\kappa > 1$, $\int_0^a \zeta_{\kappa-1}(u)\,\mathrm{d}u \to 0$ as $a \to 0$. Hence, a 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 $(x,\theta) \in \mathsf{X}\times\mathsf{K}$,

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

The following bounds also hold true. Again these are easy consequences of A1, Lemma 3.1 and Proposition 3.2. We omit the details.

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* $a \in (0, 1/2]$

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

(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* $a \in (0, 1/2]$

$$\begin{aligned}
&|\tilde{g}(x,\theta) - \tilde{g}(x,\theta')| \\
&\quad \leq C(\mathsf{K}) \sup_{\theta \in \mathsf{K}} |f_\theta|_{V^\beta} \bigg(\int_0^a \zeta_{\delta-1}(u)\,\mathrm{d}u + \zeta_\kappa(a) |f_\theta - f_{\theta'}|_{V^\beta} \\
&\qquad + \zeta_\kappa(a) D_{\beta+\alpha\delta}(\theta,\theta') \bigg) V^{\beta+\alpha(\kappa+\delta)}(x).
\end{aligned}$$

3.2. Modulated moments and weak law of large numbers

In this section, K is an arbitrary compact subset of Θ, $(x,\theta) \in \mathsf{X} \times \mathsf{K}$ 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 $P_l(n; (x_1,\theta_1), (\mathrm{d}x', \mathrm{d}\theta')) = \bar{R}(l+n; (x_1,\theta_1), (\mathrm{d}x', \mathrm{d}\theta'))$. 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 Atchadé and Fort [4].

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* $0 \leq \beta \leq 1$,

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

(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* $(x, \theta) \in \mathsf{X} \times \mathsf{K}$, $1 \leq n < N$

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

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* $\mathbb{E}(|D_k|^p) < \infty$ *for some* $p > 1$. *Then*

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

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

The nonhomogeneous Markov chain $\{(\tilde{X}_n, \tilde{\theta}_n), n \geq 0\}$ satisfies a weak law of large numbers. The proof is similar to the Atchadé and Fort [4], Theorem 2.3, and we omit the details.

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* $\varepsilon > 0$, $\kappa > 0$ *such that* $\beta + \alpha\kappa < 1 - \alpha$ *and*

$$\mathbb{E}^{(l)}_{x,\theta}\left[\sum_{k \geq 1} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}} > k\}} k^{-1+\varepsilon} \bigl(D_\beta(\tilde{\theta}_k, \tilde{\theta}_{k-1}) + |f_{\tilde{\theta}_k} - f_{\tilde{\theta}_{k-1}}|_{V^\beta}\bigr) V^{\beta+\alpha\kappa}(\tilde{X}_k)\right] < \infty. \tag{25}$$

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.*

3.3. 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

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

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

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

Lemma 3.8. *If* A2 *holds then* $\check{\mathbb{P}}(T_{\nu_\infty} < \infty) = 1$.

The following is Lemma 4.1 of Andrieu *et al.* [2]. This result relates the distribution of the nonhomogeneous Markov chain $\{(\tilde{X}_n, \tilde{\theta}_n), n \geq 0\}$ to the distribution of the adaptive MCMC process.

Proposition 3.9. *For any* $n \in \mathbb{N}$, *any* n*-uplet* $(t_1, \ldots, t_n)$, *any bounded measurable functions* $\{f_k, k \leq n\}$ *and for any* $(x, \theta, j) \in \mathsf{X} \times \mathsf{K}_j \times \mathbb{N}$,

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

Proposition 3.9 has a number of interesting consequence. For example, one can easily obtain the finiteness of moments of the adaptive chain.

Lemma 3.10. *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* $l, k \leq n$,

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

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

Another consequence of Proposition 3.9 is that many asymptotic results for the nonhomogeneous Markov chain $\{(\tilde{X}_n, \tilde{\theta}_n), n \geq 0\}$ implies a similar result from the adaptive chain. The idea of the proof is similar to Proposition 6 of Andrieu and Moulines [1] and we omit the details.

Lemma 3.11. *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 nonincreasing 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* $\delta > 0$

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

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* $n \to \infty$.

3.4. Proof of Theorem 2.1

Since A1 and (8) hold, we can apply Proposition 3.7 which implies that

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

for any $\delta > 0$, $l \geq 0$ and $(x, \theta) \in \mathsf{X} \times \mathsf{K}_l$. Theorem 2.1 then follows from Lemma 3.11.

3.5. Proof of Theorem 2.2

As in the statement of the theorem, let $\kappa > 1$, $\delta > 0$, $\rho \in (\frac{1}{2}, \frac{1}{2(1-\delta)})$, be such that $2\beta + \alpha(\kappa + \delta) < 1 - \alpha$; and let $a_n \propto n^{-\rho}$, $a_n \in (0, \frac{1}{2}]$. 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

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

Note that $\xi_k = 0$ indicates a reprojection at time k. By A2, the number of reprojection is finite almost surely (see Lemma 3.8). Thus, $n^{-1/2}\sum_{k=1}^n H_{a_n,\theta_{k-1}}(X_{k-1}, X_k)\mathbb{1}_{\{\xi_{k-1}=0\}}$ converges to zero $\check{\mathbb{P}}$-a.s.

We apply Lemma 3.11 to the term $n^{-1/2}\sum_{k=1}^n (f(X_k) - H_{a_n,\theta_{k-1}}(X_{k-1}, X_k))$. Clearly, $|f(X_k) - H_{a_n,\theta_{k-1}}(X_{k-1}, X_k)| \leq C(\mathsf{K}_{\nu_\infty})|f|_{V^\beta}(V^{\beta+\alpha\kappa}(X_{k-1}) + V^{\beta+\alpha\kappa}(X_{k-1}))$ which is uniformly bounded in n. Thus, by Lemma 3.11, it suffices to show that for any $s \geq 0$, $n^{-1/2}\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > n\}} \times \sum_{k=1}^n (f(\tilde{X}_k) - H_{a_{n+s},\tilde{\theta}_{k-1}}(\tilde{X}_{k-1}, \tilde{X}_k))$ converges to zero in $\mathbb{P}^{(l)}_{x,\theta}$-probability for all $(x, \theta) \in \mathsf{X}_0 \times \mathsf{K}_l$, $l \geq 0$. We do it as follows.

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

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

Then we use the approximate Poisson equation (20) to rewrite $S_{n,s}$ as:

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

Using Propositions 3.2, 3.3 and 3.4, it is easy to show that each of these terms above, multiplied by $\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>n\}}$, converges in probability to zero.

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 $\{(M_{n,k},\mathcal{F}_k), 1\le k\le n\}$ is a martingale array. We will show that

$$\{(M_{n,k},\check{\mathcal{F}}_k), 1\le k\le n\} \text{ is a square-integrable martingale array;} \tag{26}$$

$$\lim_{n\to\infty}\sum_{j=1}^n \check{\mathbb{E}}(D_{n,j}^2|\check{\mathcal{F}}_{j-1})=\sigma_\star^2(f) \qquad (\text{in } \check{\mathbb{P}}\text{-probab.}), \tag{27}$$

where

$$\sigma_\star^2(f)\stackrel{\text{def}}{=}\int \pi(\mathrm{d}x)\bigl(-f^2(x)+2f(x)g(x,\theta_\star)\bigr), \tag{28}$$

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

$$\lim_{n\to\infty}\sum_{k=1}^n \check{\mathbb{E}}\bigl(D_{n,j}^2\mathbb{1}_{\{|D_{n,j}|\ge\varepsilon\}}|\check{\mathcal{F}}_{j-1}\bigr)=0 \qquad (\text{in } \check{\mathbb{P}}\text{-probab.}). \tag{29}$$

The theorem then follows by the central limit theorem for martingales (see, e.g., [10], Corollary 3.1).

Proof of (26). By Proposition 3.2(i) (applied with both $\kappa>1$ and $\delta>0$),

$$\begin{aligned}
&\mathbb{E}_{x,\theta}^{(l)}\bigl(H_{a_n,\tilde\theta_{k-1}}^2(\tilde X_{k-1},\tilde X_k)\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l}>k-1\}}|\mathcal{F}_{k-1}\bigr)\\
&\quad\le \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l}>k-1\}}P_{\tilde\theta_{k-1}}g_{a_n}^2(\tilde X_{k-1},\tilde\theta_{k-1})\\
&\quad\le C(\mathsf{K}_l)\zeta_\delta(a_n)\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l}>k-1\}}V^{1-\alpha}(\tilde X_{k-1}).
\end{aligned}$$

From Proposition 3.4(i), we thus obtain

$$\sup_{(x,\theta)\in\mathsf{X}_0\times\mathsf{K}_l}\mathbb{E}_{x,\theta}^{(l)}\bigl(H_{a_n,\tilde\theta_{k-1}}^2(\tilde X_{k-1},\tilde X_k)\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l}>k-1\}}\bigr)\le C(\mathsf{K}_l)\zeta_\delta(a_n)k^{1-\alpha}\sup_{x\in\mathsf{X}_0}V^{1-\alpha}(x)<\infty.$$

To obtain (26), then apply Lemma 3.10. □

Proof of (27).

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

The same argument as above shows that

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

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

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

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

where

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

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

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

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

$$T_n^{(5)}=2\sum_{k=1}^{n}\int\pi(\mathrm{d}x)f(x)\bigl(g(x,\theta_{k-1})-g(x,\theta_\star)\bigr),$$

$$T_n^{(6)}=\sum_{k=1}^{n}\Bigl[-f^2(X_{k-1})+2f(X_{k-1})g(X_{k-1},\theta_{k-1})-\int\pi(\mathrm{d}x)\bigl(-f^2(x)+2f(x)g(x,\theta_{k-1})\bigr)\Bigr].$$

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 $\tilde{T}^{(2)}_{n,s}$ in terms of the nonhomogeneous Markov chain $\{(\tilde{X}_n,\tilde{\theta}_n),n\ge 0\}$, to show that $\lim_{n\to\infty}\mathbb{P}^{(l)}_{x,\theta}(|\tilde{T}^{(2)}_{n,s}|>\delta)=0$ for any $l\ge 0$, $\delta>0$ and any $(x,\theta)\in\mathsf{X}_0\times\Theta_l$ and then to argue that $\lim_{n\to\infty}\check{\mathbb{P}}(|T_n^{(2)}|>\delta)=0$ for all $\delta>0$ using Lemma 3.11.

Lemma 3.12. *$n^{-1}(T_n^{(2)}+T_n^{(3)})$ converges in probability to zero.*

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

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

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

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

Then we apply Lemma 3.11 to conclude that $n^{-1}(T_n^{(2)}+T_n^{(3)})$ converges in $\check{\mathbb{P}}$-probability to zero. □

Lemma 3.13. *$n^{-1}T_n^{(4)}$ converges in probability to zero.*

Proof. For $l, s \geq 0$, define $\tilde{T}^{(4)}_{n,s} \stackrel{\text{def}}{=} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l}>n\}}\sum_{k=1}^{n} f(\tilde{X}_{k-1})(g_{a_{n+s}}(\tilde{X}_{k-1},\tilde{\theta}_{k-1}) - g(\tilde{X}_{k-1}, \tilde{\theta}_{k-1}))$. Again, for any $(x,\theta)\in\mathsf{X}\times\mathsf{K}_l$ and by Proposition 3.2(ii) we get

$$\mathbb{E}^{(l)}_{x,\theta}\bigl(n^{-1}\bigl|\tilde{T}^{(4)}_{n,s}\bigr|\bigr)\leq C(\mathsf{K}_l)\left(\int_0^{a_{n+s}}\zeta_{\kappa-1}(u)\,\mathrm{d}u\right)n^{-1}\mathbb{E}^{(l)}_{x,\theta}\left(\sum_{k=1}^{n}\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l}>k-1\}}V^{2\beta+\alpha\kappa}(\tilde{X}_k)\right).$$

The rest of the proof is similar to the above upon noticing that for $\kappa > 1$, $\int_0^a \zeta_{\kappa-1}(u)\,\mathrm{d}u \to 0$ as $a \to 0$. □

Lemma 3.14. *$n^{-1}T_n^{(5)}$ 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 $a\in(0,1/2]$

$$|g(x,\theta)-g(x,\theta')|\leq C(\mathsf{K})\left(\int_0^a\zeta_{\kappa-1}(u)\,\mathrm{d}u+a^{-1}D_{\beta+\alpha\kappa}(\theta,\theta')\right)V^{\beta+\alpha\kappa}(x).$$

Therefore,

$$\left|\int\pi(\mathrm{d}x)f(x)\bigl(g(x,\theta)-g(x,\theta')\bigr)\right|\leq C(\mathsf{K})\left(\int_0^a\zeta_{\kappa-1}(u)\,\mathrm{d}u+a^{-1}D_{\beta+\alpha\kappa}(\theta,\theta')\right)\pi(V^{2\beta+\alpha\kappa}).$$

Let $\varepsilon > 0$. Since $\int_0^a \zeta_{\kappa-1}(u)\,\mathrm{d}u \to 0$ as $a \to 0$, we can find $a_0 \in (0, 1/2]$ such that $\int_0^{a_0} \zeta_{\kappa-1}(u)\,\mathrm{d}u < \varepsilon$. Then for $\check{\mathbb{P}}$-almost every sample path

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

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

Lemma 3.15. *$n^{-1}T_n^{(6)}$ 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 K of Θ, $\sup_{\theta\in\mathsf{K}} |f^2 + 2fg_\theta|_{V^{2\beta+\alpha\kappa}} < \infty$ and $2\beta + \alpha\kappa < 1 - \alpha$. To check (8), it is enough to find $\varepsilon > 0$ such that

$$\mathbb{E}_{x,\theta}^{(l)}\left[\sum_{k\ge 1} k^{-1+\varepsilon} |fg_{\tilde{\theta}_{k-1}} - fg_{\tilde{\theta}_k}|_{V^{2\beta+\alpha\kappa}} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > k\}} V^{2\beta+\alpha(\kappa+\delta)}(\tilde{X}_k)\right] < \infty. \tag{30}$$

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 $a \in (0, 1/2]$, $|f(\cdot)g(\cdot,\theta) - f(\cdot)g(\cdot,\theta')|_{V^{2\beta+\alpha\kappa}} \le C(\mathsf{K}) \int_0^a \zeta_{\kappa-1}(u)\,\mathrm{d}u + a^{-1} D_{\beta+\alpha\kappa}(\theta,\theta')$. We let a depend on k by taking $a = a_k$, therefore

$$\mathbb{E}_{x,\theta}^{(l)}\left[\sum_{k\ge 1} k^{-1+\varepsilon} |fg_{\tilde{\theta}_{k-1}} - fg_{\tilde{\theta}_k}|_{V^{2\beta+\alpha\kappa}} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > k\}} V^{2\beta+\alpha(\kappa+\delta)}(\tilde{X}_k)\right]$$
$$\le \mathbb{E}_{x,\theta}^{(l)}\left[\sum_{k\ge 1} k^{-1+\varepsilon} \int_0^{a_k} \zeta_{\kappa-1}(u)\,\mathrm{d}u \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > k\}} V^{1-\alpha}(\tilde{X}_k)\right]$$
$$+ \mathbb{E}_{x,\theta}^{(l)}\left[\sum_{k\ge 1} k^{-1+\varepsilon} a_k^{-1} D_{\beta+\alpha\kappa}(\tilde{\theta}_{k-1},\tilde{\theta}_k) \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}_l} > k\}} V^{2\beta+\alpha(\kappa+\delta)}(\tilde{X}_k)\right].$$

We can then find $\varepsilon > 0$ such that $\int_0^{a_n} \zeta_{\kappa-1}(u)\,\mathrm{d}u = \mathrm{O}(n^{-\varepsilon})$ and $n^{-1+\varepsilon} a_n^{-1} = \mathrm{O}(n^{-1/2+\rho(1-\delta)})$, and (30) follows. □

□

Proof of (29). It is suffices to show that

$$\lim_{n\to\infty} n^{-1} \sum_{k=1}^{n} \int P_{\theta_{k-1}}(X_{k-1},\mathrm{d}y) H^2_{a_n,\theta_{k-1}}(X_{k-1},y) \mathbb{1}_{\{|H a_n,\theta_{k-1}(X_{k-1},y)| \ge \varepsilon\sqrt{n}\}} = 0,$$

in $\check{\mathbb{P}}$-probability. By Lemma 9 of Andrieu and Moulines [1]

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

where

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

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

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

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

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

It follows that

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

and since $\rho < 1$, we are done. □

3.6. Proof of Proposition 2.5

For integers $p \ge 0$, $n \ge 1$ and a compact subset K of Θ, we define the random variables

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

We will show that for any $p \ge 0$, $n \ge 1$, any compact subset K of Θ and any $\delta > 0$,

$$\sup_{(x,\theta)\in \mathsf{X}_0 \times \Theta_0} \mathbb{P}^{(p)}_{x,\theta}\big(C_{n,p}(\mathsf{K}) > \delta\big) \le \mathcal{B}(n, p), \tag{31}$$

where the upper bound $\mathcal{B}(n, p)$ satisfies $\lim_{n\to\infty} \mathcal{B}(n, p) = 0$ for any $p \ge 0$ and $\lim_{p\to\infty} \mathcal{B}(n, p) = 0$ for any $n \ge 1$. This implies the stated result by the same argument used in Andrieu *et al.*

[2], Proposition 4.2 and Theorem 5.5. We have

$$C_{n,p}(\mathsf{K}) \le \sup_{l\ge n} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>l\}}\left|\sum_{j=n}^{l}\gamma_{p+j-1}\tilde{\varepsilon}_j^{(1)}\right| + \sup_{l\ge n} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>l\}}\left|\sum_{j=n}^{l}\gamma_{p+j-1}\tilde{\varepsilon}_j^{(2)}\right|. \tag{32}$$

We start with the second term on the rhs of (32). By Doob's inequality and B2, for $N > n$,

$$\begin{aligned}
&\mathbb{P}_{x,\theta}^{(l)}\left(\sup_{n\le l\le N} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>l\}}\left|\sum_{j=n}^{l}\gamma_{p+j-1}\tilde{\varepsilon}_j^{(2)}\right| > \delta\right)\\
&\quad\le \delta^{-2}\mathbb{E}_{x,\theta}^{(l)}\left(\sum_{j=n}^{N}\gamma_{p+j-1}^2\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>j\}}\int \Phi_{\tilde{\theta}_j}^2(\tilde{X}_j,y)q_{\tilde{\theta}_j}^{(1)}(\tilde{X}_j,\mathrm{d}y)\right)\\
&\quad\le C(\mathsf{K})\delta^{-2}\left(\gamma_{p+n}^2\mathbb{E}_{x,\theta}^{l}\big(\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>n-1\}}V^{2\eta+\alpha}(\tilde{X}_n)\big) + \sum_{j=n}^{N}\gamma_{p+j}^2\right).
\end{aligned}$$

It follows that

$$\mathbb{P}_{x,\theta}^{(l)}\left(\sup_{l\ge n} \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>l\}}\left|\sum_{j=n}^{l}\gamma_{p+j-1}\tilde{\varepsilon}_j^{(2)}\right| > \delta\right) \le C(\mathsf{K})\delta^{-p}\left(\gamma_{p+n}^2 n^{2\eta+\alpha} + \sum_{j\ge n}\gamma_{p+j}^2\right). \tag{33}$$

To deal with the first term on the right-hand side of (32), we proceed as in the proof of Theorem 2.1. We consider the sequence $\{a_n, n\ge 0\}$ 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 $1 \le n \le l$ and $p \ge 0$, 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, Υ_θ admits an approximate Poisson equation $\tilde{g}_a$ for any $j \ge 1$ 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

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

where

$$\begin{aligned}
T_{n,l}^{(1)} &= \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>l\}}\sum_{j=n}^{l}\mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>j\}}\gamma_{p+j}\big((1-a_j)^{-1}-1\big)\tilde{g}_{a_j}(\tilde{X}_j,\tilde{\theta}_j),\\
T_{n,l}^{(2)} &= \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>n\}}\gamma_{p+n}\tilde{g}_{a_n}(\tilde{X}_n,\tilde{\theta}_n) - \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}}>l\}}\gamma_{p+l}P_{\tilde{\theta}_l}\tilde{g}_{a_l}(\tilde{X}_l,\tilde{\theta}_l),
\end{aligned}$$

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

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

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

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

We deal with each of these terms using similar techniques as in the proofs of Theorem 2.1 and Theorem 2.2. Some of the details are thus omitted. Let $\delta > 0$ arbitrary.

On term $T_{n,l}^{(1)}$. 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 we have

$$\mathbb{P}_{x,\theta}^{(p)}\Big(\sup_{l\ge n}\big|T_{n,l}^{(1)}\big| > \delta\Big) \le \delta^{-1} C(\mathsf{K}) V(x) \left(\gamma_{n+p} n^{1-\alpha-\rho} + \sum_{j\ge n} \gamma_{p+j} j^{-\rho}\right). \tag{34}$$

On term $T_{n,l}^{(2)}$. Let $\varepsilon > 0$, $\kappa > 1$ such that $\varepsilon \in (\rho, (1-\alpha)(\eta+\kappa\alpha)^{-1} - 1)$. That is $(1+\varepsilon)(\eta + \alpha\kappa) < 1 - \alpha$ and $\varepsilon > \rho$. Then

$$\mathbb{P}_{x,\theta}^{(p)}\Big(\sup_{l\ge n}\big|T_{n,l}^{(2)}\big| > \delta\Big) \le (2/\delta)^{1+\varepsilon} C(\mathsf{K}) V(x) \left(\gamma_{p+n}^{1+\varepsilon} n^{1-\alpha} + \sum_{j\ge n-1} \gamma_{p+j}^{1+\varepsilon}\right). \tag{35}$$

On term $T_{n,l}^{(3)}$. 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 $|\tilde{g}_a(x,\theta) - \tilde{g}_a(x,\theta')| \le C(\mathsf{K}) \sup_{\theta\in\mathsf{K}} |\Upsilon_\theta|_{V^\eta} \zeta_\delta(a) |\theta - \theta'| V^{\eta+\alpha(\kappa+\delta)}(x)$. Then

$$\mathbb{P}_{x,\theta}^{(p)}\Big(\sup_{l\ge n}\big|T_{n,l}^{(3)}\big| > \delta\Big) \le (1/\delta) C(\mathsf{K}) \left(\gamma_{p+n-1}^2 n^{1+\rho-\alpha} + \sum_{j\ge n} \gamma_{p+j-1}^2 j^\rho\right) V(x). \tag{36}$$

On term $T_{n,l}^{(4)}$. We have

$$\mathbb{P}_{x,\theta}^{(p)}\Big(\sup_{l\ge n}\big|T_{n,l}^{(4)}\big| > \delta\Big) \le (1/\delta) C(\mathsf{K}) V(x) \big(n^{1-\alpha}(\gamma_{p+n} - \gamma_{p+n+1}) + \gamma_{p+n}\big). \tag{37}$$

On term $T_{n,l}^{(5)}$. Take $\kappa \in (1,2)$ such that $\eta + \alpha\kappa < 1 - \alpha$. One can check as in Proposition 3.3 that for any compact K $|P_\theta \tilde{g}_a(x,\theta) - P_\theta \tilde{g}_{a'}(x,\theta)| \le C(\mathsf{K}) |a - a'| a^{\kappa-2} V^{\eta+\alpha\kappa}(x)$. And for $a_j \propto$

$j^{-\rho}$, $|a_j - a_{j-1}|a_j^{\kappa-2} \propto j^{-1}a_j^{\kappa-1} = \mathrm{o}(j^{-1})$. Hence, by Markov's inequality, we get:

$$\mathbb{P}_{x,\theta}^{(p)}\Bigl(\sup_{l\geq n}\bigl|T_{n,l}^{(5)}\bigr| > \delta\Bigr) \leq \delta^{-1}C(\mathsf{K})V(x)\Bigl(n^{-\alpha}\gamma_{p+n} + \sum_{j\geq n}\gamma_{p+j}j^{-1}\Bigr). \tag{38}$$

On term $T_{n,l}^{(6)}$ Let $\kappa > 1$ such that $2(\eta + \alpha\kappa/2) < 1 - \alpha$. Consider the term $D_j = \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}} > j\}}\gamma_{p+j}(\tilde{g}_{a_j}(\tilde{X}_{j+1}, \tilde{\theta}_j) - P_{\tilde{\theta}_j}\tilde{g}_{a_j}(\tilde{X}_j, \tilde{\theta}_j))$ so that $T_{n,l}^{(2)} = \mathbb{1}_{\{\overleftarrow{\tau}_{\mathsf{K}} > l\}}\sum_{j=n}^{l-1} D_j$. We note that D_j is a martingale difference and by Doob's inequality we get:

$$\mathbb{P}_{x,\theta}^{(p)}\Bigl(\sup_{l\geq n}\bigl|T_{n,l}^{(6)}\bigr| > \delta\Bigr) \leq (1/\delta)^2 C(\mathsf{K})V(x)\Bigl(\gamma_{p+n-1}^2 n^{1-\alpha+\rho} + \sum_{j\geq n}\gamma_{p+j-1}^2 j^{\rho}\Bigr). \tag{39}$$

By combining (33)–(39) and (16), we get (31) as claimed.

3.7. Proof of the results of Section 2.7

3.7.1. *Proof of Proposition* 2.6

The function $a(\theta)$ is of class $\mathcal{C}^1$. Hence, by Assumption C1 and the mean value theorem $\mathcal{L} = \{\theta \in \mathbb{R}\colon a(\theta) = \bar{\alpha}\}$ is not empty. It also follows from C1 that the function $\theta \to \int_0^\theta \cosh(u)(\bar{\alpha} - a(u))\,\mathrm{d}u$ is bounded from below; so we can find K_1 such that $w(\theta) = \int_0^\theta \cosh(u)(\bar{\alpha} - a(u))\,\mathrm{d}u + K_1 \geq 0$. 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 $\mathcal{L} \subset \{\theta \in \mathbb{R}\colon w(\theta) < M_0\}$ and $\mathcal{W}_M$ is bounded thus compact for any $M > 0$.

3.7.2. *Proof of Proposition* 2.8

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

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

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

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

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

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

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.

Acknowledgements

We are grateful to Michael Woodroofe for helpful discussions on martingale approximation techniques and to Shukri Osman for helpful conversations. This work is partly supported by NSF Grant DMS 0906631 and by the French National Research Agency (ANR) program ANR-05-BLAN-0299.

References

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

[2] Andrieu, C., Moulines, É. and Priouret, P. (2005). Stability of stochastic approximation under verifiable conditions. *SIAM J. Control Optim.* **44** 283–312. MR2177157

[3] Andrieu, C. and Thoms, J. (2008). A tutorial on adaptive MCMC. *Stat. Comput.* **18** 343–373. MR2461882

[4] Atchadé, Y. and Fort, G. (2010). Limit theorems for some adaptive MCMC algorithms with subgeometric kernels. *Bernoulli* **16** 116–154. MR2648752

[5] Atchade, Y.F. and Fort, G. (2009). Limit theorems for some adaptive mcmc algorithms with subgeometric kernels: Part ii. Technical report. Univ. Michigan. Available at arXiv:0911.0221.

[6] Atchade, Y.F., Fort, G., Moulines, E. and Priouret, P. (2009). Adaptive Markov chain Monte Carlo: Theory and methods. Technical report, Univ. Michigan.

[7] Atchadé, Y.F. and Rosenthal, J.S. (2005). On adaptive Markov chain Monte Carlo algorithms. *Bernoulli* **11** 815–828. MR2172842

[8] Bai, Y. (2008). The simultaneous drift conditions for Adaptive Markov Chain Monte Carlo algorithms. Technical report, Univ. Toronto. (Personal communication.)

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

[10] Hall, P. and Heyde, C.C. (1980). *Martingale Limit Theory and Its Application: Probability and Mathematical Statistics*. New York: Academic Press [Harcourt Brace Jovanovich Publishers]. MR0624435

[11] Jarner, S.F. and Roberts, G.O. (2002). Polynomial convergence rates of Markov chains. *Ann. Appl. Probab.* **12** 224–247. MR1890063

[12] Jarner, S.F. and Roberts, G.O. (2007). Convergence of heavy-tailed Monte Carlo Markov chain algorithms. *Scand. J. Statist.* **34** 781–815. MR2396939

[13] Kamatani, K. (2009). Metropolis–Hastings algorithms with acceptance ratios of nearly 1. *Ann. Inst. Statist. Math.* **61** 949–967. MR2556773

[14] 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. MR0834478

[15] Kushner, H.J. and Yin, G.G. (2003). *Stochastic Approximation and Recursive Algorithms and Applications*, 2nd ed. *Applications of Mathematics (New York)* **35**. New York: Springer. MR1993642

[16] Maxwell, M. and Woodroofe, M. (2000). Central limit theorems for additive functionals of Markov chains. *Ann. Probab.* **28** 713–724. MR1782272

[17] Roberts, G.O. and Rosenthal, J.S. (2001). Optimal scaling for various Metropolis–Hastings algorithms. *Statist. Sci.* **16** 351–367. MR1888450

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

[19] Roberts, G.O. and Tweedie, R.L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2** 341–363. MR1440273

[20] Saksman, E. and Vihola, M. (2010). On the ergodicity of the adaptive Metropolis algorithm on unbounded domains. *Ann*. *Appl*. *Probab.* **20** 2178–2203. MR2759732

Received August 2009 and revised January 2011