Bernoulli **16**(1), 2010, 116–154
DOI: 10.3150/09-BEJ199

Limit theorems for some adaptive MCMC algorithms with subgeometric kernels

YVES 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*

This paper deals with the ergodicity (convergence of the marginals) and the law of large numbers for adaptive MCMC algorithms built from transition kernels that are not necessarily geometrically ergodic. We develop a number of results that significantly broaden the class of adaptive MCMC algorithms for which rigorous analysis is now possible. As an example, we give a detailed analysis of the adaptive Metropolis algorithm of Haario *et al.* [*Bernoulli* **7** (2001) 223–242] when the target distribution is subexponential in the tails.

Keywords: adaptive Markov chain Monte Carlo; Markov chain; subgeometric ergodicity

1. Introduction

This paper deals with the convergence of adaptive Markov chain Monte Carlo (AMCMC) algorithms. Markov chain Monte Carlo (MCMC) is a well-known, widely used method to sample from arbitrary probability distributions. One of the major limitations of the method is the difficulty in finding sensible values for the parameters of the Markov kernels. Adaptive MCMC provides a general framework to tackle this problem where the parameters are adaptively tuned, often using previously generated samples. This approach generates a class of stochastic processes that is the object of this paper.

Denote by π the probability measure of interest on some measure space $(\mathsf{X}, \mathcal{X})$. Let $\{P_\theta, \theta \in \Theta\}$ be a family of ϕ-irreducible and aperiodic Markov kernels, each with invariant distribution π. We are interested in the class of stochastic processes based on non-homogeneous Markov chains $\{(X_n, \theta_n), n \geq 0\}$ with transition kernels $\{\bar{P}(n; (x, \theta); (\mathrm{d}x', \mathrm{d}\theta')), n \geq 0\}$ satisfying $\int_\Theta \bar{P}(n; (x, \theta); (\cdot, \mathrm{d}\theta')) = P_\theta(x, \cdot)$. Often, these transition kernels are of the form $\{P_\theta(x, \mathrm{d}y)\delta_{H_n(\theta,y)}(\mathrm{d}\theta'), n \geq 0\}$, where $\{H_l, l \geq 0\}$ is a family of measurable functions, $H_l : \Theta \times \mathsf{X} \to \Theta$. The stochastic approximation dynamic corresponds to the case $H_l(\theta, x) = \theta + \gamma_l H(\theta, x)$. In this latter case, it is assumed that the best values for θ are the solutions of the equation $\int H(\theta, x)\pi(\mathrm{d}x) = 0$. Since the pioneering work of Gilks *et al.* (1998); Holden (1998); Haario *et al.* (2001); Andrieu and Robert (2001), the number of AMCMC algorithms in the literature has significantly increased. However, despite much recent work on the topic, the asymptotic behavior of these algorithms is still not completely understood. Almost all previous work on the convergence of AMCMC is limited to the case where each kernel P_θ is geometrically er-

1350-7265 © 2010 ISI/BS

godic (see, e.g., Roberts and Rosenthal (2007); Andrieu and Moulines (2006)). In this paper, we weaken this condition and consider the case where each transition kernel is subgeometrically ergodic.

More specifically, we study the ergodicity of the marginal $\{X_n, n \geq 0\}$, that is, the convergence to π of the distribution of X_n irrespective of the initial distribution, and the existence of a strong law of large numbers for AMCMC.

We first show that a diminishing adaptation assumption of the form $|\theta_n - \theta_{n-1}| \to 0$, in a sense to be made precise (assumption B1), together with a uniform-in-θ positive recurrence toward a small set C (assumptions A1(i) and A1(iii)) and a uniform-in-θ ergodicity condition of the kernels $\{P_\theta, \theta \in \Theta\}$ (assumption A1(ii)) are enough to imply the ergodicity of AMCMC.

We believe that this result is close to being optimal. Indeed, it is well documented in the literature that AMCMC can fail to be ergodic if the diminishing assumption does not hold (see, e.g., Roberts and Rosenthal (2007) for examples). Furthermore, the additional assumptions are also fairly weak since in the case where Θ is reduced to the single point $\{\theta_\star\}$ so that $\{X_n, n \geq 0\}$ is a Markov chain with transition kernel $P_{\theta_\star}$, these conditions hold if $P_{\theta_\star}$ is an aperiodic positive kernel that is polynomially ergodic.

We then prove a strong law of large numbers for AMCMC. We show that the diminishing adaptation assumption and a uniform-in-θ polynomial drift condition toward a small set $\mathcal{C}$ of the form $P_\theta V \leq V - cV^{1-\alpha} + b\mathbb{1}_{\mathcal{C}}(x)$, $\alpha \in (0, 1)$, implies a strong law of large numbers for all real-valued measurable functions f for which $\sup_{\mathsf{X}}(|f|/V^\beta) < \infty$, $\beta \in [0, 1-\alpha)$. This result is close to what can be achieved with Markov chains (with fixed transition kernel) under similar conditions; see Meyn and Tweedie (1993).

On a more technical note, this paper makes two key contributions to the analysis of AMCMC. First, to study the ergodicity, we use a more careful coupling technique which extends the coupling approach of Roberts and Rosenthal (2007). Second, we tackle the law of large numbers using a resolvent kernel approach, together with martingale theory. This approach has a decisive advantage over the more classical Poisson equation approach (see Andrieu and Moulines (2006)), in that no continuity property of the resolvent kernels is required. It is also worth noting that the results developed in this paper can be applied to adaptive Markov chains beyond Markov chain Monte Carlo simulation, provided that all of the transition kernels have the same invariant distribution.

The remainder of the paper is organized as follows. In Section 2, we state our assumptions, followed by a statement of our main results. Detailed discussion of the assumptions and some comparisons with the literature are provided in Section 2.4. We apply our results to the analysis of the adaptive random walk Metropolis algorithm of Haario *et al.* (2001) when the target distribution is subexponential in the tails. This is covered in Section 3, together with a toy example taken from Atchade and Rosenthal (2005). All proofs are postponed to Section 4.

2. Statement of the results and discussion

2.1. Notation

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

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

$\delta_x(\mathrm{d}t)$ 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, then 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 μ is a signed measure on a measurable space $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$, then 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, for some function $V : \mathbb{T} \to [1, +\infty)$, is defined as $\|\mu\|_V \stackrel{\text{def}}{=} \sup_{\{g, |g|_V \leq 1\}} |\mu(g)|$.

Let X, Θ be two general state spaces, respectively endowed with countably generated σ-fields $\mathcal{X}$ and $\mathcal{B}(\Theta)$. 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. Let $\{\bar{P}(n; \cdot, \cdot), n \geq 0\}$ be a family of transition kernels on $(\mathsf{X} \times \Theta, \mathcal{X} \otimes \mathcal{B}(\Theta))$, satisfying, for any $A \in \mathcal{X}$,

$$\int_{A \times \Theta} \bar{P}(n; (x, \theta); (\mathrm{d}x', \mathrm{d}\theta')) = P_\theta(x, A). \tag{1}$$

An adaptive Markov chain is a non-homogeneous Markov chain $\{Z_n = (X_n, \theta_n), n \geq 0\}$ on $\mathsf{X} \times \Theta$ with transition kernels $\{\bar{P}(n; \cdot; \cdot), n \geq 0\}$.

Among examples of such transition kernels, consider the case where $\{(X_n, \theta_n), n \geq 0\}$ is obtained through the algorithm: given (X_n, θ_n), sample $X_{n+1} \sim P_{\theta_n}(X_n, \cdot)$ and set $\theta_{n+1} = \theta_n$ with probability $1 - p_{n+1}$ or set $\theta_{n+1} = \tilde{\Xi}_{n+1}(X_n, \theta_n, X_{n+1})$ with probability p_{n+1}. Then

$$\bar{P}(n; (x, \theta); (\mathrm{d}x', \mathrm{d}\theta')) = P_\theta(x, \mathrm{d}x') \big\{ (1 - p_{n+1}) \delta_\theta(\mathrm{d}\theta') + p_{n+1} \delta_{\tilde{\Xi}_{n+1}(x, \theta, x')}(\mathrm{d}\theta') \big\}.$$

Consider the special case where $p_{n+1} = 1$ and $\theta_{n+1} = H_{n+1}(\theta_n, X_{n+1})$, where $\{H_l, l \geq 0\}$ is a family of measurable functions $H_l : \Theta \times \mathsf{X} \to \Theta$. Then

$$\bar{P}(n; (x, \theta); (\mathrm{d}x', \mathrm{d}\theta')) \stackrel{\text{def}}{=} P_\theta(x, \mathrm{d}x') \delta_{H_{n+1}(\theta, x')}(\mathrm{d}\theta').$$

Such a situation occurs if, for example, θ_{n+1} is updated following a stochastic approximation dynamic: $\theta_{n+1} = \theta_n + \gamma_{n+1} H(\theta_n, X_{n+1})$.

From $\{\bar{P}(n;\cdot,\cdot), n \geq 0\}$ and for any integer $l \geq 0$, we introduce a family – indexed by l – of sequence of transition kernels $\{\bar{P}_l(n;\cdot,\cdot), n \geq 0\}$, where $\bar{P}_l(n;\cdot,\cdot) \stackrel{\text{def}}{=} \bar{P}(l+n;\cdot,\cdot)$, and we denote by $\mathbb{P}^{(l)}_{x,\theta}$ and $\mathbb{E}^{(l)}_{x,\theta}$ the probability and expectation, respectively, on the canonical space $(\Omega, \mathcal{F})$ of the canonical non-homogeneous Markov chain $\{Z_n = (X_n, \theta_n), n \geq 0\}$ with transition kernels $\{\bar{P}_l(n;\cdot;\cdot), n \geq 0\}$ and initial distribution $\delta_{(x,\theta)}$. We denote by $\underline{\theta}$ the shift operator on Ω and by $\{\mathcal{F}_k, k \geq 0\}$ the natural filtration of the process $\{Z_k, k \geq 0\}$. We use the notation $\mathbb{P}_{x,\theta}$ and $\mathbb{E}_{x,\theta}$ as shorthand for $\mathbb{P}^{(0)}_{x,\theta}$ and $\mathbb{E}^{(0)}_{x,\theta}$, respectively.

Set

$$D(\theta, \theta') \stackrel{\text{def}}{=} \sup_{x \in \mathsf{X}} \| P_\theta(x,\cdot) - P_{\theta'}(x,\cdot) \|_{\mathrm{TV}}.$$

2.2. Convergence of the marginals

We assume that minorization, drift conditions and ergodicity are available for P_θ uniformly in θ. For a set $\mathcal{C}$, denote by $\tau_{\mathcal{C}}$ the return time to $\mathcal{C} \times \Theta : \tau_{\mathcal{C}} \stackrel{\text{def}}{=} \inf\{n \geq 1, X_n \in \mathcal{C}\}$.

A1 There exist a measurable function $V : \mathsf{X} \to [1, +\infty)$ and a measurable set $\mathcal{C}$ such that:
  (i) $\sup_l \sup_{\mathcal{C}\times\Theta} \mathbb{E}^{(l)}_{x,\theta}[\mathbf{r}(\tau_{\mathcal{C}})] < +\infty$ for some non-decreasing function $\mathbf{r} : \mathbb{N} \to (0, +\infty)$ such that $\sum_n 1/\mathbf{r}(n) < +\infty$;
  (ii) there exists a probability measure π such that

$$\lim_{n\to+\infty} \sup_{x\in\mathsf{X}} V^{-1}(x) \sup_{\theta\in\Theta} \| P^n_\theta(x,\cdot) - \pi \|_{\mathrm{TV}} = 0;$$

  (iii) $\sup_\theta P_\theta V \leq V$ on $\mathcal{C}^c$ and $\sup_{\mathcal{C}\times\Theta}\{P_\theta V(x) + V(x)\} < +\infty$.

B1 There exist probability distributions ξ_1, ξ_2 on X, Θ, respectively, such that for any $\epsilon > 0$, $\lim_n \mathbb{P}_{\xi_1,\xi_2}(D(\theta_n, \theta_{n-1}) \geq \epsilon) = 0$.

Theorem 2.1. *Assume* A1 *and* B1. *Then*

$$\lim_{n\to+\infty} \sup_{\{f, |f|_1 \leq 1\}} |\mathbb{E}_{\xi_1,\xi_2}[f(X_n) - \pi(f)]| = 0.$$

Sufficient conditions for A1 to hold are the following uniform-in-θ conditions:

A2 (i) the transition kernels P_θ are ϕ-irreducible and aperiodic;
  (ii) there exist a function $V : \mathsf{X} \to [1, +\infty)$, $\alpha \in (0, 1)$ and constants b, c such that for any $\theta \in \Theta$,

$$P_\theta V(x) \leq V(x) - cV^{1-\alpha}(x) + b\mathbb{1}_{\mathcal{C}}(x);$$

  (iii) for any level set $\mathcal{D}$ of V, there exist $\epsilon_{\mathcal{D}} > 0$ and a probability $\nu_{\mathcal{D}}$ such that for any θ, $P_\theta(x,\cdot) \geq \epsilon_{\mathcal{D}} \mathbb{1}_{\mathcal{D}}(x) \nu_{\mathcal{D}}(\cdot)$.

We thus have the following corollary.

Corollary 2.2 (of Theorem 2.1). *Assume A2 and B1. Then*

$$\lim_{n\to+\infty} \sup_{\{f,|f|_1\le 1\}} |\mathbb{E}_{\xi_1,\xi_2}[f(X_n)-\pi(f)]| = 0.$$

Assumptions A1(i) and A1(iii) are designed to control the behavior of the chain "far from the center". When the state space X is "bounded" so that, for example, $V=1$ in A1(ii), then we have the following result.

Lemma 2.3. *If there exists a probability measure π such that $\lim_{n\to+\infty}\sup_{\mathsf{X}\times\Theta}\|P_\theta^n(x,\cdot)-\pi(\cdot)\|_{\mathrm{TV}}=0$, then A1(i) and A1(iii) hold with a bounded function V and $\mathcal{C}=\mathsf{X}$.*

Combining the assumptions of Lemma 2.3 and B1, we deduce from Theorem 2.1 the convergence of the marginals. This result coincides with (Roberts and Rosenthal (2007), Theorem 5). As observed in Bai (2008) (personal communication), assumption A2 also implies the "containment condition" as defined in Roberts and Rosenthal (2007). Consequently, Corollary 2.2 could also be established by applying (Roberts and Rosenthal (2007), Theorem 13): this would yield the following statement, which is adapted from Bai (2008). Define $M_\epsilon(x,\theta)\stackrel{\mathrm{def}}{=}\inf\{n\ge 1, \|P_\theta^n(x,\cdot)-\pi(\cdot)\|_{\mathrm{TV}}\le\epsilon\}$.

Proposition 2.4. *Assume A2 and B1. Then, for any $\epsilon>0$, the sequence $\{M_\epsilon(X_n,\theta_n), n\ge 0\}$ is bounded in probability for the probability $\mathbb{P}_{\xi_1,\xi_2}$ and*

$$\lim_{n\to+\infty} \sup_{\{f,|f|_1\le 1\}} |\mathbb{E}_{\xi_1,\xi_2}[f(X_n)-\pi(f)]| = 0.$$

2.3. Strong law of large numbers

Assumptions A1 and B1 are strengthened as follows:

A3 there exist a probability measure ν on X, a positive constant ε and a set $\mathcal{C}\in\mathcal{X}$ such that for any $\theta\in\Theta$, $P_\theta(x,\cdot)\ge \mathbb{1}_{\mathcal{C}}(x)\varepsilon\nu(\cdot)$;

A4 there exist a measurable function $V:\mathsf{X}\to[1,+\infty)$, $0<\alpha<1$ and positive constants b, c such that for any $\theta\in\Theta$, $P_\theta V\le V-cV^{1-\alpha}+b\mathbb{1}_{\mathcal{C}}$;

A5 there exist a probability measure π and some $0\le\beta<1-\alpha$ such that for any level set $\mathcal{D}\stackrel{\mathrm{def}}{=}\{x\in\mathsf{X}, V(x)\le d\}$ of V,

$$\lim_{n\to+\infty}\sup_{\mathcal{D}\times\Theta}\|P_\theta^n(x,\cdot)-\pi\|_{V^\beta}=0;$$

B2 for any level set $\mathcal{D}$ of V and any $\epsilon>0$,

$$\lim_n \sup_{l\ge 0}\sup_{\mathcal{D}\times\Theta}\mathbb{P}_{x,\theta}^{(l)}\big(D(\theta_n,\theta_{n-1})\ge\epsilon\big)=0.$$

Theorem 2.5. *Assume A3–A5 and B2. Then, for any measurable function $f : \mathsf{X} \to \mathbb{R}$ in $\mathcal{L}_{V^\beta}$ and any initial distribution ξ_1 (resp. ξ_2) on X (resp. Θ) such that $\xi_1(V) < +\infty$,*

$$\lim_{n\to+\infty} n^{-1}\sum_{k=1}^{n} f(X_k) = \pi(f), \qquad \mathbb{P}_{\xi_1,\xi_2}\text{-}a.s.$$

As in the case of the convergence of the marginals, when A5 and B2 hold with $\mathcal{D} = \mathsf{X}$ and $\beta = 0$, A3 and A4 can be omitted. We thus have the following.

Proposition 2.6. *Assume that A5 and B2 hold with $\mathcal{D} = \mathsf{X}$ and $\beta = 0$. Then, for any measurable bounded function $f : \mathsf{X} \to \mathbb{R}$ and any initial distribution ξ_1 (resp. ξ_2) on X (resp. Θ),*

$$\lim_{n\to+\infty} n^{-1}\sum_{k=1}^{n} f(X_k) = \pi(f), \qquad \mathbb{P}_{\xi_1,\xi_2}\text{-}a.s.$$

2.4. Discussion

2.4.1. *Non-adaptive case*

We start by comparing our assumptions to assumptions in Markov chain theory under which the law of large numbers hold. In the setup above, taking $\Theta = \{\theta_\star\}$ and $H(\theta_\star, x) = \theta_\star$ reduces $\{X_n, n \geq 0\}$ to a Markov chain with transition kernel $P_{\theta_\star}$. Assume that $P_{\theta_\star}$ is Harris recurrent.

In that case, a condition which is known to be minimal and to imply ergodicity in total variation norm is that $P_{\theta_\star}$ is an aperiodic positive Harris recurrent transition kernel (Meyn and Tweedie (1993), Theorems 11.0.1 and 13.0.1). Condition A1(i) is stronger than positive Harris recurrence since it requires $\sup_{\mathcal{C}} \mathbb{E}_x[\mathbf{r}(\tau_{\mathcal{C}})] < +\infty$ for some rate $\mathbf{r}$, $\mathbf{r}(n) \gg n$. Nevertheless, as discussed in the proof (see Remark 2, Section 4), the condition $\sum_n \{1/\mathbf{r}(n)\} < +\infty$ is really designed for the adaptive case. A1(ii) is stronger than what we want to prove (since A1(ii) implies the conclusion of Theorem 2.1 in the non-adaptive case); indeed, this is due to our proof technique, which is based on the comparison of the adaptive process to a process – namely, a Markov chain with transition kernel P_θ – whose stationary distribution is π. Our proof is thus designed to address the adaptive case. Finally, B1 is trivially true.

For the strong law of large numbers (Theorem 2.5), B2 is still trivially true in the Markovian case and A5 is implied by A3 and A4, combined with the assumption that $P_{\theta_\star}$ is ϕ-irreducible and aperiodic (see Appendix A and references therein). In the Markovian case, whenever $P_{\theta_\star}$ is ϕ-irreducible and aperiodic, A3 and A4 are known sufficient conditions for a strong law of large numbers for $f \in \mathcal{L}_{V^{1-\alpha}}$, which is somewhat stronger than the conclusions of Theorem 2.5. This slight loss of efficiency is due to the proof technique based on martingale theory (see comments in Section 2.4.5). Observe that in the geometric case, there is the same loss of generality as in Theorem 8 of Andrieu and Moulines (2006). More generally, any proof of the law of large numbers based on martingale theory (through, e.g., the use of Poisson's equation or the resolvent kernel) will incur the same loss of efficiency since limit theorems exist only for L^p-martingales with $p > 1$.

2.4.2. *Checking assumptions* A1(ii) *and* A5

A1(ii) and A5 are the most technical of our assumptions. Contrary to the case of a single kernel, the relations between A1(ii) (resp. A5) and A1(i)–A3 (resp. A3, A4) are not completely well understood. Nevertheless, these assumptions can be checked under conditions which are essentially of the form A3, A4 along with the assumptions that each transition kernel P_θ is ϕ-irreducible and aperiodic, as discussed in Appendix A.

2.4.3. *On the uniformity in θ in assumptions* A1(i), A1(ii), A3 *and* A4

We have formulated A1(i), A1(ii), A3 and A4 such that all of the constants involved are independent of θ, for $\theta \in \Theta$. Intuitively, this corresponds to AMCMC algorithms based on kernels with similar overall ergodicity properties. This uniformity assumption might seem unrealistically strong at first. However, the next example shows that when these conditions do not hold uniformly in θ for $\theta \in \Theta$, pathologies can occur if the adaptation parameter can wander to the boundary of Θ.

Example 1. This example is adapted from Winkler (2003). Let $\mathsf{X} = \{0, 1\}$ and $\{P_\theta, \theta \in (0, 1)\}$ be a family of transition matrices with $P_\theta(0, 0) = P_\theta(1, 1) = 1 - \theta$. Let $\{\theta_n, n \geq 0\}$, $\theta_n \in (0, 1)$, be a deterministic sequence of real numbers decreasing to 0 and let $\{X_n, n \geq 0\}$ be a non-homogeneous Markov chain on $\{0, 1\}$ with transition matrices $\{P_{\theta_n}, n \geq 0\}$. One can check that $D(\theta_n, \theta_{n-1}) \leq \theta_{n-1} - \theta_n$ for all $n \geq 1$ so that B1 and B2 hold.

For any compact subset K of $(0, 1)$, it can be checked that A1(i), A1(ii), A3 and A4 hold uniformly for all $\theta \in \mathsf{K}$. However, these assumptions do not hold uniformly for all $\theta \in (0, 1)$. Therefore, Theorems 2.1 and 2.5 do not apply. Actually, one can easily check that $\mathbb{P}_{x,\theta_0}(X_n \in \cdot) \to \pi(\cdot)$ as $n \to \infty$, but that $\mathbb{E}_{x,\theta_0}[(n^{-1} \sum_{k=1}^{n} f(X_k) - \pi(f))^2]$ does not converge to 0 for bounded functions f. That is, the marginal distribution of X_n converges to π, but a weak law of large numbers fails to hold.

This raises the question of how to construct AMCMC when A1(i), A1(ii), A3 and A4 do not hold uniformly for all $\theta \in \Theta$. When these assumptions hold uniformly on any compact subsets of Θ and the adaptation is based on stochastic approximation, one approach is to stop the adaptation or to reproject θ_n back on $\mathcal{K}$ whenever $\theta_n \notin \mathcal{K}$ for some fixed compact $\mathcal{K}$ of Θ. A more elaborate strategy is Chen's truncation method which – roughly speaking – reinitializes the algorithm with a larger compact whenever $\theta_n \notin \mathcal{K}$ (Chen and Zhu (1986); Chen *et al.* (1988)). A third strategy consists of proving a drift condition on the bivariate process $\{(X_n, \theta_n), n \geq 0\}$ in order to ensure the stability of the process (Andrieu and Tadic (2008), see also Benveniste *et al.* (1987)). This question is, however, beyond the scope of this paper; the use of Chen's truncation method to weaken our assumption is addressed in Atchadé and Fort (2008).

2.4.4. *Comparison with the literature*

The convergence of AMCMC has been considered in a number of early works, most under a geometric ergodicity assumption. Haario *et al.* (2001) proved the convergence of the adaptive random walk Metropolis (ARWM) when the state space is bounded. Their results were generalized to unbounded spaces in Atchade and Rosenthal (2005), assuming the diminishing adaptation assumption and a geometric drift condition of the form

$$P_\theta V(x) \leq \lambda V(x) + b\mathbb{1}_C(x), \tag{2}$$

for $\lambda \in (0, 1)$, $b < \infty$ and $\theta \in \Theta$.

Andrieu and Moulines (2006) undertook a thorough analysis of adaptive chains under the geometric drift condition (2) and proved a strong law of large numbers and a central limit theorem. Andrieu and Atchade (2007) provided a theoretical discussion on the efficiency of AMCMC under (2).

Roberts and Rosenthal (2007) improved on the literature by relaxing the convergence rate assumption on the kernels. The authors proved the convergence of the marginal and a weak law of large numbers for bounded functions. However, their analysis requires a uniform control on certain moments of the drift function, a condition which is easily checked in the geometric case (i.e., when A2 or A4 is replaced by (2)). Until recently, it was an open question in the polynomial case, but this was recently solved by Bai (2008) – contemporaneously with our own work – who proved that such a control holds under conditions which are essentially of the form A2.

Yang (2007) tackled some open questions mentioned in Roberts and Rosenthal (2007) by providing sufficient conditions – close to the conditions we give in Theorems 2.1 and 2.5 – to ensure convergence of the marginals and a weak law of large numbers for bounded functions. The conditions in (Yang (2007), Theorems 3.1 and 3.2) are stronger than our conditions, but we have noted some omissions and mistakes in the proofs of these theorems.

2.4.5. *Comments on the methods of proof*

The proof of Theorem 2.1 is based on an argument extended from Roberts and Rosenthal (2007), which can be sketched heuristically as follows. For N large enough, we can expect $P^N_{\theta_n}(X_n, \cdot)$ to be within ϵ of π (by ergodicity). On the other hand, since the adaptation is diminishing, by waiting long enough, we can find n such that the distribution of X_{n+N} given (X_n, θ_n) is within ϵ of $P^N_{\theta_n}(X_n, \cdot)$. Combining these two arguments, we can then conclude that the distribution of X_{n+N} is within 2ϵ of π. This is essentially the argument of Roberts and Rosenthal (2007). The difficulty with this argument is that the distance between $P^N_{\theta_n}(x, \cdot)$ and π depends, in general, on x and can rarely be bounded uniformly in x. We solve this problem here by introducing some level set $\mathcal{C}$ of V and using two basic facts: (i) under A1(i), the process cannot wait too long before coming back into $\mathcal{C}$; (ii) under A1(ii) and (iii), a bound on the distance between $P^N_{\theta_n}(x, \cdot)$ and π uniformly in x, for $x \in \mathcal{C}$, is possible.

The proof of Theorem 2.5 is based on a resolvent kernel approach that we adapted from Merlevede *et al.* (2006) (see also Maxwell and Woodroofe (2000)), combined with martingale theory. Another possible route to the SLLN is the Poisson equation technique which was used to study adaptive MCMC in Andrieu and Moulines (2006). Under A3 and A4, a solution g_θ to Poisson's equation with transition kernel P_θ exists for any $f \in \mathcal{L}_{V^\beta}$, $0 \leq \beta \leq 1 - \alpha$ and $g_\theta \in \mathcal{L}_{V^{\beta+\alpha}}$. However, in order to use $\{g_\theta, \theta \in \Theta\}$ to obtain an SLLN for f, we typically need to control $|g_\theta - g_{\theta'}|$, which, overall, can be expensive. Here, we avoid these pitfalls by introducing the resolvent $\hat{g}_a(x,\theta)$ of the process $\{X_n\}$, defined by

$$\hat{g}_a^{(l)}(x,\theta) \stackrel{\text{def}}{=} \sum_{j\geq 0}(1-a)^{j+1}\mathbb{E}^{(l)}_{x,\theta}[f(X_j)], \qquad x\in\mathsf{X}, \theta\in\Theta, a\in(0,1), l\geq 0.$$

3. Examples

3.1. A toy example

We first consider an example discussed in Atchade and Rosenthal (2005) (see also Roberts and Rosenthal (2007)). Let π be a target density on the integers $\{1,\dots,K\}$, $K\geq 4$. Let $\{P_\theta, \theta\in\{1,\dots,M\}\}$ be a family of random walk Metropolis algorithms with proposal distribution q_θ, the uniform distribution on $\{x-\theta,\dots,x-1,x+1,\dots,x+\theta\}$.

Consider the sequence $\{(X_n,\theta_n), n\geq 0\}$ defined as follows: given X_n,θ_n:

- the conditional distribution of X_{n+1} is $P_{\theta_n}(X_n,\cdot)$;
- if $X_{n+1}=X_n$, set $\theta_{n+1}=\max(1,\theta_n-1)$ with probability p_{n+1} and $\theta_{n+1}=\theta_n$ otherwise; if $X_{n+1}\neq X_n$, set $\theta_{n+1}=\min(M,\theta_n+1)$ with probability p_{n+1} and $\theta_{n+1}=\theta_n$ otherwise.

This algorithm defines a non-homogeneous Markov chain – still denoted $\{(X_n,\theta_n), n\geq 0\}$ – on a canonical probability space endowed with a probability $\mathbb{P}$. The transitions of this Markov process are given by the family of transition kernels $\{\bar{P}(n;(x,\theta),(\mathrm{d}x',\mathrm{d}\theta'), n\geq 0\}$, where

$$\begin{aligned}\bar{P}(n;(x,\theta),(\mathrm{d}x',\mathrm{d}\theta')) = P_\theta(x,\mathrm{d}x')\big(&\mathbb{1}_{x=x'}\big\{p_{n+1}\delta_{1\vee(\theta-1)}(\mathrm{d}\theta') + (1-p_{n+1})\delta_\theta(\mathrm{d}\theta')\big\}\\ &+ \mathbb{1}_{x\neq x'}\big\{p_{n+1}\delta_{M\wedge(\theta+1)}(\mathrm{d}\theta') + (1-p_{n+1})\delta_\theta(\mathrm{d}\theta')\big\}\big).\end{aligned}$$

In this example, each kernel P_θ is uniformly ergodic: P_θ is ϕ-irreducible, aperiodic, possesses an invariant probability measure π and is such that

$$\lim_n \sup_{x\in\mathsf{X}} \|P_\theta^n(x,\cdot)-\pi(\cdot)\|_{\mathrm{TV}} = 0.$$

Since Θ is finite, this implies that A1(ii) (resp. A5) holds with $V=1$ (resp. $\mathcal{D}=\mathsf{X}$ and $\beta=0$). Furthermore, $\mathbb{E}^{(l)}_{x,\theta}[D(\theta_n,\theta_{n+1})]\leq 2p_{n+1}$ so that B1 (resp. B2) holds with any probability measures ξ_1,ξ_2 (resp. with $\mathcal{D}=\mathsf{X}$) provided $p_n\to 0$. By Lemma 2.3 combined with Theorem 2.1, and by Proposition 2.6, we have the following.

Proposition 3.1. *Assume that* $\lim_n p_n = 0$. *For any probability distributions* ξ_1,ξ_2 *on* X,Θ:

(i) $\sup_{\{f,|f|_1\leq 1\}}|\mathbb{E}_{\xi_1,\xi_2}[f(X_n)]-\pi(f)|\to 0$;
(ii) *for any bounded function* f,

$$n^{-1}\sum_{k=1}^n f(X_k)\to\pi(f), \qquad \mathbb{P}_{\xi_1,\xi_2}\text{-}a.s.$$

3.2. The adaptive random walk Metropolis of Haario *et al.* (2001)

We illustrate our results with the adaptive random walk Metropolis of Haario *et al.* (2001). The random walk Metropolis (RWM) algorithm is a popular MCMC algorithm; see Hastings (1970); Metropolis *et al.* (1953). Let a target density π, absolutely continuous with respect to the Lebesgue measure μ_{Leb}. We still denote by π the density. Choose a proposal distribution with density with respect to μ_{Leb} denoted q, and assume that q is a positive symmetric density on $\mathbb{R}^p$. The algorithm generates a Markov chain $\{X_n, n \geq 0\}$ with invariant distribution π as follows. Given $X_n = x$, a new value $Y = x + Z$ is proposed, where Z is generated from $q(\cdot)$. We then either 'accept' Y and set $X_{n+1} = Y$ with probability $\alpha(x, Y) \stackrel{\text{def}}{=} \min(1, \pi(Y)/\pi(x))$ or we 'reject' Y and set $X_{n+1} = x$.

For definiteness, we will assume that q is a zero-mean multivariate Gaussian distribution (this assumption can be replaced by regularity conditions and moment conditions on the proposal distribution). Given a proposal distribution with finite second moments, the convergence rate of the RWM kernel depends mainly on the tail behavior of the target distribution π. If π is superexponential in the tails with regular contours, then the RWM kernel is typically geometrically ergodic (Jarner and Hansen (2000)). Otherwise, it is typically subgeometric (Fort and Moulines (2000, 2003); Douc *et al.* (2004)).

Define

$$\mu_\star \stackrel{\text{def}}{=} \int_{\mathsf{X}} x\pi(x)\mu_{\mathrm{Leb}}(\mathrm{d}x), \qquad \Sigma_\star \stackrel{\text{def}}{=} \int_{\mathsf{X}} xx^{\mathrm{T}}\pi(x)\mu_{\mathrm{Leb}}(\mathrm{d}x) - \mu_\star\mu_\star^{\mathrm{T}},$$

to be, respectively, the expectation and the covariance matrix of π ($\cdot^{\mathrm{T}}$ denotes the transpose operation). Theoretical results suggest setting the variance-covariance matrix Σ of the proposal distribution to be $\Sigma = c_\star\Sigma_\star$, where $c_\star$ is set so as to reach the optimal acceptance rate $\bar{\alpha}$ in stationarity (typically, $\bar{\alpha}$ is set to values around 0.3–0.4); see, for example, Roberts and Rosenthal (2001) for more details. Haario *et al.* (2001) proposed an adaptive algorithm to find Σ_* adaptively during the simulation. This algorithm has been studied in detail in Andrieu and Moulines (2006), under the assumption that π is superexponential in the tails. An adaptive algorithm to find the optimal value $c_\star$ was proposed in Atchade and Rosenthal (2005) (see also Atchade (2006)) and studied under the assumption that π is superexponential in the tails. We extend these results to cases where π is subexponential in the tails.

Let Θ_+ be a convex compact of the cone of $p \times p$ symmetric positive definite matrices endowed with the Shur norm $|\cdot|_{\mathrm{s}}$, $|A|_{\mathrm{s}} \stackrel{\text{def}}{=} \sqrt{\mathrm{Tr}(A^{\mathrm{T}}A)}$. For example, for a, $M > 0$, $\Theta_+ = \{A + \mathsf{a}\mathrm{Id} :$ A is symmetric positive semidefinite and $|A|_s \leq M\}$. Next, for $-\infty < \kappa_l < \kappa_u < \infty$ and Θ_μ a compact subset of X, we introduce the space $\Theta \stackrel{\text{def}}{=} \Theta_\mu \times \Theta_+ \times [\kappa_l, \kappa_u]$. For $\theta = (\mu, \Sigma, c) \in \Theta$, denote by P_θ the transition kernel of the RWM algorithm with proposal q_θ, where q_θ stands for the multivariate Gaussian distribution with variance-covariance matrix $\mathrm{e}^c\Sigma$.

Consider the adaptive RWM algorithm defined as follows.

Algorithm 3.1. Initialization: *Let $\bar{\alpha}$ be the target acceptance probability. Choose $X_0 \in \mathsf{X}$, $(\mu_0, \Sigma_0, c_0) \in \Theta$.*

Iteration: *Given $(X_n, \mu_n, \Sigma_n, c_n)$:*

1. *Generate* $Z_{n+1} \sim q_{\theta_n}\, \mathrm{d}\mu_{\mathrm{Leb}}$ *and set* $Y_{n+1} = X_n + Z_{n+1}$. *With probability* $\alpha(X_n, Y_{n+1})$, *set* $X_{n+1} = Y_{n+1}$ *and with probability* $1 - \alpha(X_n, Y_{n+1})$, *set* $X_{n+1} = X_n$.
2. *Set*:

$$\mu = \mu_n + (n+1)^{-1}(X_{n+1} - \mu_n), \tag{3}$$

$$\Sigma = \Sigma_n + (n+1)^{-1}[(X_{n+1} - \mu_n)(X_{n+1} - \mu_n)^{\mathrm{T}} - \Sigma_n], \tag{4}$$

$$c = c_n + \frac{1}{n+1}\big(\alpha(X_n, Y_{n+1}) - \bar{\alpha}\big). \tag{5}$$

3. *If* $(\mu, \Sigma, c) \in \Theta$, *set* $\mu_{n+1} = \mu$, $\Sigma_{n+1} = \Sigma$ *and* $c_{n+1} = c$. *Otherwise, set* $\mu_{n+1} = \mu_n$, $\Sigma_{n+1} = \Sigma_n$ *and* $c_{n+1} = c_n$.

This is an algorithmic description of a random process $\{(X_n, \theta_n), n \ge 0\}$, which is a non-homogeneous Markov chain with successive transitions kernels $\{\bar{P}(n; (x, \theta), (\mathrm{d}x', \mathrm{d}\theta')), n \ge 0\}$ given by

$$\begin{aligned}
&\bar{P}(n; (x, \theta), (\mathrm{d}x', \mathrm{d}\theta')) \\
&\quad = \int q_\theta(z)\big\{\alpha(x, x+z)\delta_{x+z}(\mathrm{d}x') + \big(1 - \alpha(x, x+z)\big)\delta_x(\mathrm{d}x')\big\} \cdots \\
&\qquad \times \big(\mathbb{1}_{\{\phi(\theta, x+z, x') \in \Theta\}}\delta_{\phi(\theta, x+z, x')}(\mathrm{d}\theta') + \mathbb{1}_{\{\phi(\theta, x+z, x') \notin \Theta\}}\delta_\theta(\mathrm{d}\theta')\big)\, \mathrm{d}\mu_{\mathrm{Leb}}(\mathrm{d}z),
\end{aligned}$$

where ϕ is the function defined from the right-hand side expressions of (3)–(5). Integrating over θ', we see that for any $A \in \mathcal{X}$,

$$\int_{A \times \Theta} \bar{P}(n; (x, \theta), (\mathrm{d}x', \mathrm{d}\theta')) = P_\theta(x, A).$$

Lemma 3.2. *Assume that* π *is bounded from below and from above on compact sets. Then any compact subset* $\mathcal{C}$ *of* X *with* $\mu_{\mathrm{Leb}}(\mathcal{C}) > 0$ *satisfies* A3.

Proof. See Theorem 2.2 of Roberts and Tweedie (1996). □

Following (Fort and Moulines (2000)), we assume that π is subexponential in the tails:

D1 π is positive and continuous on $\mathbb{R}^p$, and twice continuously differentiable in the tails;
D2 there exist $m \in (0, 1)$, positive constants $d_i < D_i$, $i = 0, 1, 2$ and $r, R > 0$ such that for $|x| \ge R$:

(i) $\langle \frac{\nabla\pi(x)}{|\nabla\pi(x)|}, \frac{x}{|x|} \rangle \le -r$;
(ii) $d_0|x|^m \le -\log\pi(x) \le D_0|x|^m$;
(iii) $d_1|x|^{m-1} \le |\nabla \log\pi(x)| \le D_1|x|^{m-1}$;
(iv) $d_2|x|^{m-2} \le |\nabla^2 \log\pi(x)| \le D_2|x|^{m-2}$.

Examples of target densities that satisfy D1 and D2 are the Weibull distributions on $\mathbb{R}$ with density $\pi(x) \propto |x|^{m-1}\exp(-\beta|x|^m)$ (for large $|x|$), $\beta > 0$, $m \in (0, 1)$. Multidimensional examples are provided in Fort and Moulines (2000).

3.2.1. *Law of large numbers for exponential functions*

In this subsection, we assume that

D3 there exist $s_\star > 0$, $0 < \upsilon < 1 - m$ and $0 < \eta < 1$ such that as $|x| \to +\infty$,

$$\sup_{\theta\in\Theta} \int_{\{z, |z| \geq \eta |x|^\upsilon\}} \left(1 \vee \frac{\pi(x)}{\pi(x+z)}\right)^{s_\star} q_\theta(z) \mu_{\mathrm{Leb}}(\mathrm{d}z) = \mathrm{o}\big(|x|^{2(m-1)}\big).$$

A sufficient condition for D3 is that $\pi(x+z) \geq \pi(x)\pi(z)$ for any x large enough and $|z| \geq \eta |x|^\upsilon$ (which holds true for Weibull distributions with $0 < m < 1$). Indeed, we then have

$$\begin{aligned} &\int_{\{z, |z| \geq \eta |x|^\upsilon\}} \left(1 \vee \frac{\pi(x)}{\pi(x+z)}\right)^{s_\star} q_\theta(z) \mu_{\mathrm{Leb}}(\mathrm{d}z) \\ &\quad \leq C \exp(-\lambda_\star \eta^2 |x|^{2\upsilon}) \sup_{\theta\in\Theta} \int \exp(s_\star D_0 |z|^m) \exp(\lambda_\star |z|^2) q_\theta(z) \mu_{\mathrm{Leb}}(\mathrm{d}z) \end{aligned}$$

for some constant $C < +\infty$, and $\lambda_\star > 0$ such that the right-hand side is finite.

Lemma 3.3. *Assume* D1–D3. *For* $0 < s \leq s_\star$, *define* $V_s(x) \stackrel{\mathrm{def}}{=} 1 + \pi^{-s}(x)$. *There exists* $0 < s \leq s_\star$ *and, for any* $\alpha \in (0,1)$, *there exist positive constants* b, c *and a compact set* $\mathcal{C}$ *such that*

$$\sup_{\theta\in\Theta} P_\theta V_s(x) \leq V_s(x) - c V_s^{1-\alpha}(x) + b \mathbb{1}_{\mathcal{C}}(x).$$

Hence, A2–A5 *hold*.

Lemma 3.4. *Assume* D1–D3. B2 *holds and* B1 *holds for any probability measures* ξ_1, ξ_2 *such that* $\int |\ln \pi|^{2/m} \, \mathrm{d}\xi_1 < +\infty$.

The proofs of Lemmas 3.3 and 3.4 appear in Appendix C.

Proposition 3.5. *Assume* D1–D3. *Consider the sequence* $\{X_n, n \geq 0\}$ *given by Algorithm* 3.1.

(i) *For any probability measures* ξ_1, ξ_2 *such that* $\int |\ln \pi|^{2/m} \, \mathrm{d}\xi_1 < +\infty$,

$$\sup_{\{f, |f|_1 \leq 1\}} |\mathbb{E}_{\xi_1,\xi_2}[f(X_n)] - \pi(f)| \to 0.$$

(ii) *There exists* $0 < s \leq s_\star$ *such that for any probability measures* ξ_1, ξ_2 *such that* $\int |\pi|^{-s} \, \mathrm{d}\xi_1 < +\infty$, *and any function* $f \in \mathcal{L}_{1+\pi^{-r}}$, $0 \leq r < s$,

$$n^{-1} \sum_{k=1}^n f(X_k) \to \pi(f), \qquad \mathbb{P}_{\xi_1,\xi_2}\text{-}a.s.$$

The drift function V_s exhibited in Lemma 3.3. is designed for limit theorems relative to functions f increasing as $\exp(\beta|x|^m)$. This implies a condition on the initial distribution ξ_1 which has to possess subexponential moments (see Proposition 3.5(ii)), something which always holds when $\xi_1 = \delta_x$, $x \in \mathsf{X}$.

3.2.2. *Law of large numbers for polynomially increasing functions*

Proposition 3.5 also addresses the case where f is of the form $1 + |x|^r$, $r > 0$. Nevertheless, the conditions on ξ_1 and the assumptions D3 can be weakened in that case.

We have to find a drift function V such that $V^{1-\alpha}(x) \sim 1 + |x|^{r+\iota}$ for some $\alpha \in (0,1)$, $\iota > 0$. Under D3, this can be obtained from the proof of Lemma 3.3 and this yields $V(x) \sim 1 + |x|^{r+\iota+2-m}$ (apply Jensen's inequality to the drift inequality (24) with the concave function $\phi(t) \sim [\ln t]^{(r+\iota+2)/m-1}$; see Lemma 3.5 in Jarner and Roberts (2002) for similar calculations). Hence, the condition on ξ_1 becomes $\xi_1(|x|^{r+\iota+2-m}) < +\infty$ for some $\iota > 0$.

Drift inequalities with $V \sim (-\ln \pi)^s$ for some $s > 2/m - 1$ can also be derived by direct computation: in that case, D3 can be removed. Details are omitted and left to the interested reader.

To conclude, observe that these discussions relative to polynomially increasing functions can be extended to any function f which is a concave transformation of π^{-s}.

4. Proofs of the results of Section 2

For a set $\mathcal{C} \in \mathcal{X}$, define the hitting time on $\mathcal{C} \times \Theta$ of $\{Z_n, n \geq 0\}$ by $\sigma_{\mathcal{C}} \stackrel{\text{def}}{=} \inf\{n \geq 0, Z_n \in \mathcal{C} \times \Theta\}$. If $\pi(|f|) < +\infty$, we set $\bar{f} \stackrel{\text{def}}{=} f - \pi(f)$.

4.1. Preliminary results

We gather some useful preliminary results in this section. Section 4.1.1 provides an approximation of the marginal distribution of the adaptive chain by the distribution of a related Markov chain. In Section 4.1.2, we develop various bounds for modulated moments of the adaptive chain as consequences of the drift conditions. In Section 4.1.3, we bound the expected return times of the adaptive chain to level sets of the drift function V. The culminating result of this subsection is Theorem 4.10, which gives an explicit bound on the resolvent function $g_a^{(l)}(x,\theta)$.

4.1.1. *Optimal coupling*

Lemma 4.1. *For any integers $l \geq 0$, $N \geq 2$, any measurable bounded function f on X^N and any $(x,\theta) \in \mathsf{X} \times \Theta$,*

$$\Delta \stackrel{\text{def}}{=} \left| \mathbb{E}^{(l)}_{x,\theta}[f(X_1,\ldots,X_N)] - \int_{\mathsf{X}^N} P_\theta(x, \mathrm{d}x_1) \prod_{k=2}^{N} P_\theta(x_{k-1}, \mathrm{d}x_k) f(x_1,\ldots,x_n) \right|$$
$$\leq |f|_1 \sum_{j=1}^{N-1} \sum_{i=1}^{j} \mathbb{E}^{(l)}_{x,\theta}[D(\theta_i, \theta_{i-1})].$$

Proof. We can assume without loss of generality that $|f|_1 \le 1$. Set $z_k = (x_k, t_k)$. With the convention that $\prod_{k=a}^{b} a_k = 1$ for $a > b$ and upon noting that $\int_{\mathsf{X}} P_\theta(x, \mathrm{d}x')h(x') = \int_{\mathsf{X}\times\Theta} \bar{P}_l(0; (x,\theta), (\mathrm{d}x', \mathrm{d}\theta'))h(x')$ for any bounded measurable function $h : \mathsf{X} \to \mathbb{R}$, we have

$$\Delta = \Bigg| \int_{(\mathsf{X}\times\Theta)^N} \sum_{j=1}^{N-1} \bar{P}_l(0; (x,\theta), \mathrm{d}z_1) \prod_{k=2}^{j} \bar{P}_l(k-1; z_{k-1}, \mathrm{d}z_k) \cdots$$

$$\times \{\bar{P}_l(j; z_j, \mathrm{d}z_{j+1}) - \bar{P}_l(0; (x_j,\theta), \mathrm{d}z_{j+1})\} \prod_{k=j+2}^{N} \bar{P}_l(0; (x_{k-1},\theta), \mathrm{d}z_k) f(x_1, \ldots, x_N) \Bigg|$$

$$\le \sum_{j=1}^{N-1} \int_{\mathsf{X}^j} \bar{P}_l(0; (x,\theta), \mathrm{d}z_1) \prod_{k=2}^{j} \bar{P}_l(k-1; z_{k-1}, \mathrm{d}z_k) \sup_{x\in\mathsf{X}} \|P_{t_j}(x,\cdot) - P_\theta(x,\cdot)\|_{\mathrm{TV}},$$

where we have used the fact that

$$\int_{(\mathsf{X}\times\Theta)^{N-j-1}} \prod_{k=j+2}^{N} \bar{P}_l(0; (x_{k-1},\theta), \mathrm{d}z_k) f(x_1, \ldots, x_N)$$

is bounded by a function $\Xi(x_1, \ldots, x_{j+1})$ that does not depend on t_k, $k \le N$, and for any bounded function Ξ on X^{j+1},

$$\int_{\mathsf{X}\times\Theta} \{\bar{P}_l(j; z_j, \mathrm{d}z_{j+1}) - \bar{P}_l(0; (x_j,\theta), \mathrm{d}z_{j+1})\} \Xi(x_1, \ldots, x_{j+1})$$

$$= \int_{\mathsf{X}} \{P_{t_j}(x_j, \mathrm{d}x_{j+1}) - P_\theta(x_j, \mathrm{d}x_{j+1})\} \Xi(x_1, \ldots, x_{j+1}) \le \sup_{x\in\mathsf{X}} \|P_{t_j}(x,\cdot) - P_\theta(x,\cdot)\|_{\mathrm{TV}} |\Xi|_1.$$

Hence,

$$\Delta \le \sum_{j=1}^{N-1} \mathbb{E}_{x,\theta}^{(l)}\Big[\sup_{x\in\mathsf{X}} \|P_{\theta_j}(x,\cdot) - P_{\theta_0}(x,\cdot)\|_{\mathrm{TV}}\Big]$$

$$\le \sum_{j=1}^{N-1} \mathbb{E}_{x,\theta}^{(l)}\Bigg[\sum_{i=1}^{j} \sup_{x\in\mathsf{X}} \|P_{\theta_i}(x,\cdot) - P_{\theta_{i-1}}(x,\cdot)\|_{\mathrm{TV}}\Bigg] = \sum_{j=1}^{N-1} \sum_{i=1}^{j} \mathbb{E}_{x,\theta}^{(l)}[D(\theta_i, \theta_{i-1})]. \qquad \square$$

Lemma 4.2. *Let μ, ν be two probability distributions. There exist a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and random variables X, Y on $(\Omega, \mathcal{F})$ such that $X \sim \mu$, $Y \sim \nu$ and $\mathbb{P}(X = Y) = 1 - \|\mu - \nu\|_{\mathrm{TV}}$.*

The proof can be found in, for example, Proposition 3 of Roberts and Rosenthal (2004). As a consequence of Lemmas 4.1 and 4.2, we have the following.

Proposition 4.3. *Let $l \geq 0, N \geq 2$ and set $z = (x, \theta)$. There exists a process $\{(X_k, \tilde{X}_k), 0 \leq k \leq N\}$ defined on a probability space endowed with the probability $\overline{\mathbb{P}}^{(l)}_{z,z}$ such that*

$$\overline{\mathbb{P}}^{(l)}_{z,z}(X_k = \tilde{X}_k, 0 \leq k \leq N) \geq 1 - \sum_{j=1}^{N-1} \sum_{i=1}^{j} \mathbb{E}^{(l)}_z[D(\theta_i, \theta_{i-1})],$$

$(X_0, \ldots, X_N)$ has the X-marginal distribution of $\mathbb{P}^{(l)}_z$ restricted to the time interval $\{0, \ldots, N\}$, and $(\tilde{X}_0, \ldots, \tilde{X}_N)$ has the same distribution as a homogeneous Markov chain with transition kernel P_θ and initial distribution δ_x.

4.1.2. *Modulated moments for the adaptive chain*

Let $V : \mathsf{X} \to [1, +\infty)$ be a measurable function and assume that there exist $\mathcal{C} \in \mathcal{X}$, positive constants b, c and $0 < \alpha \leq 1$ such that for any $\theta \in \Theta$,

$$P_\theta V \leq V - cV^{1-\alpha} + b\mathbb{1}_{\mathcal{C}}. \tag{6}$$

Lemma 4.4. *Assume (6). There exists $\bar{b}$ such that for any $0 \leq \beta \leq 1$, $\theta \in \Theta$: $P_\theta V^\beta \leq V^\beta - \beta c V^{\beta-\alpha} + \bar{b}\mathbb{1}_{\mathcal{C}}$.*

Proof. See Lemma 3.5 in Jarner and Roberts (2002). □

Proposition 4.5. *Assume (6). For any $l \geq 0$, $(x, \theta) \in \mathsf{X} \times \Theta$ and any stopping time τ,*

$$c\mathbb{E}^{(l)}_{x,\theta}\left[\sum_{k=0}^{\tau-1} (k\alpha c + 1)^{\alpha^{-1}-1}\right] \leq V(x) + b\mathbb{E}^{(l)}_{x,\theta}\left[\sum_{k=0}^{\tau-1} \big((k+1)\alpha c + 1\big)^{\alpha^{-1}-1} \mathbb{1}_{\mathcal{C}}(X_k)\right].$$

Proof. The proof can be adapted from Proposition 2.1 in Douc *et al.* (2004) and Proposition 11.3.2 in Meyn and Tweedie (1993), and is thus omitted. □

Proposition 4.6. *Assume (6).*

(i) *There exists $\bar{b}$ such that for any $j \geq 0$, $0 \leq \beta \leq 1$, $l \geq 0$ and $(x, \theta) \in \mathsf{X} \times \Theta$,*

$$\mathbb{E}^{(l)}_{x,\theta}[V^\beta(X_j)] \leq V^\beta(x) + \bar{b} j^\beta.$$

(ii) *Let $0 \leq \beta \leq 1$ and $0 \leq a \leq 1$. For any stopping time τ,*

$$\begin{aligned}
&\mathbb{E}^{(l)}_{x,\theta}[(1-a)^\tau V^\beta(X_\tau)\mathbb{1}_{\tau<+\infty}] + \mathbb{E}^{(l)}_{x,\theta}\left[\sum_{j=0}^{\tau-1} (1-a)^j \{aV^\beta(X_j) + \beta c(1-a)V^{\beta-\alpha}(X_j)\}\right] \\
&\quad \leq V^\beta(x) + \bar{b}(1-a)\mathbb{E}^{(l)}_{x,\theta}\left[\sum_{j=0}^{\tau-1} (1-a)^j \mathbb{1}_{\mathcal{C}}(X_j)\right].
\end{aligned}$$

(iii) *Let* $0 \leq \beta \leq 1 - \alpha$ *and* $0 < a < 1$. *For any stopping time* τ *and any* $q \in [1, +\infty]$,

$$\mathbb{E}^{(l)}_{x,\theta}\left[\sum_{j=0}^{\tau-1}(1-a)^j V^\beta(X_j)\right]$$
$$\leq a^{1/q-1}(1-a)^{-1/q} V^{\beta+\alpha/q}(x)\left(1 + \bar{b}\mathbb{E}^{(l)}_{x,\theta}\left[\sum_{j=0}^{\tau-1}(1-a)^j \mathbb{1}_{\mathcal{C}}(X_j)\right]\right)(\alpha c)^{-1/q}$$

(*with the convention that* $1/q = 0$ *when* $q = +\infty$).

Proof. The proof is already done in the case $l = 0$. The general case is similar and omitted. (i) is a trivial consequence of Lemma 4.4. (ii) Let $\beta \leq 1$. Set $\tau_N = \tau \wedge N$ and $Y_n = (1-a)^n V^\beta(X_n)$. Then

$$Y_{\tau_N} = Y_0 + \sum_{j=1}^{\tau_N}(Y_j - Y_{j-1}) = Y_0 + \sum_{j=1}^{\tau_N}(1-a)^{j-1}\big((1-a)V^\beta(X_j) - V^\beta(X_{j-1})\big)$$
$$= Y_0 + \sum_{j=1}^{\tau_N}(1-a)^j\big(V^\beta(X_j) - V^\beta(X_{j-1})\big) - a\sum_{j=1}^{\tau_N}(1-a)^{j-1}V^\beta(X_{j-1}).$$

Hence,

$$\mathbb{E}_{x,\theta}[Y_{\tau_N}] + a\mathbb{E}_{x,\theta}\left[\sum_{j=0}^{\tau_N-1}(1-a)^j V^\beta(X_j)\right]$$
$$= V^\beta(x) + \sum_{j\geq 1}(1-a)^j \mathbb{E}_{x,\theta}\big[\big(V^\beta(X_j) - V^\beta(X_{j-1})\big)\mathbb{1}_{j\leq\tau_N}\big]$$
$$\leq V^\beta(x) + \sum_{j\geq 1}(1-a)^j \mathbb{E}_{x,\theta}\big[\big(-\beta c V^{\beta-\alpha}(X_{j-1}) + \bar{b}\mathbb{1}_{\mathcal{C}}(X_{j-1})\big)\mathbb{1}_{j\leq\tau_N}\big],$$

where we have used Lemma 4.4 in the last inequality. This implies that

$$\mathbb{E}_{x,\theta}[Y_{\tau_N}] + a\mathbb{E}_{x,\theta}\left[\sum_{j=0}^{\tau_N-1}(1-a)^j V^\beta(X_j)\right] + (1-a)\beta c\mathbb{E}_{x,\theta}\left[\sum_{j=0}^{\tau_N-1}(1-a)^j V^{\beta-\alpha}(X_j)\right]$$
$$\leq V^\beta(x) + \bar{b}(1-a)\mathbb{E}_{x,\theta}\left[\sum_{j=0}^{\tau_N-1}(1-a)^j \mathbb{1}_{\mathcal{C}}(X_j)\right].$$

The result follows when $N \to +\infty$.

(iii) The previous case provides two upper bounds, namely, for $0 < \beta \leq 1 - \alpha$,

$$a\mathbb{E}_{x,\theta}\left[\sum_{j=0}^{\tau-1}(1-a)^j V^{\beta}(X_j)\right] \leq V^{\beta}(x) + \bar{b}(1-a)\mathbb{E}_{x,\theta}\left[\sum_{j=0}^{\tau-1}(1-a)^j \mathbb{1}_{\mathcal{C}}(X_j)\right]$$

and

$$(1-a)\big((\beta+\alpha)c\big)\mathbb{E}_{x,\theta}\left[\sum_{j=0}^{\tau-1}(1-a)^j V^{\beta}(X_j)\right] \leq V^{\beta+\alpha}(x) + \bar{b}\mathbb{E}_{x,\theta}\left[\sum_{j=0}^{\tau-1}(1-a)^j \mathbb{1}_{\mathcal{C}}(X_j)\right].$$

We then use the property $[c \leq c_1 \wedge c_2] \Longrightarrow c \leq c_1^{1/q} c_2^{1-1/q}$ for any $q \in [1, +\infty]$. □

Proposition 4.7. *Assume* (6). *Let* $\{r_n, n \geq 0\}$ *be a non-increasing positive sequence. There exists* $\bar{b}$ *such that for any* $l \geq 0$, $(x,\theta) \in \mathsf{X} \times \Theta$, $0 \leq \beta \leq 1$ *and* $n \geq 0$,

$$\beta c \mathbb{E}_{x,\theta}^{(l)}\left[\sum_{k \geq n} r_{k+1} V^{\beta-\alpha}(X_k)\right] \leq r_n \mathbb{E}_{x,\theta}^{(l)}[V^{\beta}(X_n)] + \bar{b}\mathbb{E}_{x,\theta}^{(l)}\left[\sum_{k \geq n} r_{k+1}\mathbb{1}_{\mathcal{C}}(X_k)\right].$$

The proof follows along the same lines as the proof of Proposition 4.6(ii) and is thus omitted.

4.1.3. *Delayed successive visits to an accessible level set of V*

Let $\mathcal{D} \in \mathcal{X}$ and let $n_\star, N$ be positive integers. Define on $(\Omega, \mathcal{F}, \mathbb{P}_{x,\theta}^{(l)})$ the sequence of $\mathbb{N}$-valued random variables $\{\tau^n, n \geq 1\}$ as

$$\tau^0 \stackrel{\text{def}}{=} \tau_{\mathcal{D}}, \qquad \tau^1 \stackrel{\text{def}}{=} \tau^0 + n_\star + \tau_{\mathcal{D}} \circ \underline{\theta}^{\tau^0+n_\star}, \qquad \tau^{k+1} \stackrel{\text{def}}{=} \tau^k + N + \tau_{\mathcal{D}} \circ \underline{\theta}^{\tau^k+N}, \qquad k \geq 1.$$

Proposition 4.8. *Assume* A3 *and that there exist* $V : \mathsf{X} \to [1, +\infty)$ *and a constant* $b < +\infty$ *such that for any* $\theta \in \Theta$, $P_\theta V \leq V - 1 + b\mathbb{1}_{\mathcal{C}}$. *Let* $\mathcal{D} \in \mathcal{X}$. *Let* $n_\star, N$ *be two non-negative integers. Then*

$$\varepsilon \nu(\mathcal{D}) \mathbb{E}_{x,\theta}^{(l)}\left[\sum_{k=0}^{\tau_{\mathcal{D}}-1} \mathbb{1}_{\mathcal{C}}(X_k)\right] \leq 1$$

and if $\sup_{\mathcal{D}} V < +\infty$ *and* $\nu(\mathcal{D}) > 0$, *then there exists a (finite) constant* C *depending on* $\varepsilon, \nu(\mathcal{D}), \sup_{\mathcal{D}} V, b, n_\star, N$ *such that for any* $l \geq 0$, $(x,\theta) \in \mathsf{X} \times \Theta$ *and* $k \geq 0$,

$$\mathbb{E}_{x,\theta}^{(l)}[\tau^k] \leq kC + V(x).$$

Proof. Since $V \geq 1$, Proposition 4.6(ii), applied with $a = 0$, $\beta = \alpha = 1$, $c = 1$ and $\tau = \tau_{\mathcal{D}}$, implies that

$$\mathbb{E}_{x,\theta}^{(l)}[\tau_{\mathcal{D}}] \leq V(x) + \bar{b}\mathbb{E}_{x,\theta}^{(l)}\left[\sum_{k=0}^{\tau_{\mathcal{D}}-1} \mathbb{1}_{\mathcal{C}}(X_k)\right].$$

By A3, we have $P_\theta(x,\mathcal{D}) \geq [\varepsilon\nu(\mathcal{D})]\mathbb{1}_{\mathcal{C}}(x)$ for any (x,θ) so that

$$\varepsilon\nu(\mathcal{D})\mathbb{E}^{(l)}_{x,\theta}\left[\sum_{k=0}^{\tau_{\mathcal{D}}-1}\mathbb{1}_{\mathcal{C}}(X_k)\right] \leq \mathbb{E}^{(l)}_{x,\theta}\left[\sum_{k=0}^{\tau_{\mathcal{D}}-1}P_{\theta_k}(X_k,\mathcal{D})\right]$$
$$= \mathbb{E}^{(l)}_{x,\theta}\left[\sum_{k=0}^{\tau_{\mathcal{D}}-1}\mathbb{1}_{\mathcal{D}}(X_{k+1})\right] \leq 1.$$

Hence, $\mathbb{E}^{(l)}_{x,\theta}[\tau_{\mathcal{D}}] \leq V(x) + \bar{b}[\varepsilon\nu(\mathcal{D})]^{-1}$. By the Markov property and Proposition 4.6(i), we have

$$\mathbb{E}^{(l)}_{x,\theta}[\tau^1] \leq n_\star + V(x) + \bar{b}[\varepsilon\nu(\mathcal{D})]^{-1} + \mathbb{E}^{(l)}_{x,\theta}\big[\mathbb{E}^{(n_\star+l+\tau_{\mathcal{D}})}_{Z_{n_\star+\tau_{\mathcal{D}}}}[\sigma_{\mathcal{D}}]\big]$$
$$\leq n_\star + 2\bar{b}[\varepsilon\nu(\mathcal{D})]^{-1} + V(x) + \sup_{\mathcal{D}} V + n_\star\bar{b}.$$

The proof is by induction on k. Assume that $\mathbb{E}^{(l)}_{x,\theta}[\tau^k] \leq kC + V(x)$ with $C \geq 2\bar{b}[\varepsilon\nu(\mathcal{D})]^{-1} + \sup_{\mathcal{D}} V + (N \vee n_\star)(1+\bar{b})$. Then, again using the Markov property and Proposition 4.6(i), and upon noting that $\mathbb{P}^{(l)}_{x,\theta}(Z_{\tau^k} \in \mathcal{D}) = 1$,

$$\mathbb{E}^{(l)}_{x,\theta}[\tau^{k+1}] \leq N + \mathbb{E}^{(l)}_{x,\theta}[\tau^k] + \mathbb{E}^{(l)}_{x,\theta}\big[\mathbb{E}^{(\tau^k+N+l)}_{Z_{\tau^k+N}}[\tau_{\mathcal{D}}]\big]$$
$$\leq N + \bar{b}[\varepsilon\nu(\mathcal{D})]^{-1} + \mathbb{E}^{(l)}_{x,\theta}[\tau^k] + \mathbb{E}^{(l)}_{x,\theta}[V(X_{\tau^k+N})]$$
$$\leq N + \bar{b}[\varepsilon\nu(\mathcal{D})]^{-1} + \mathbb{E}^{(l)}_{x,\theta}[\tau^k] + \mathbb{E}^{(l)}_{x,\theta}\big[\mathbb{E}^{(\tau^k+l)}_{Z_{\tau^k}}[V(X_N)]\big]$$
$$\leq N + \bar{b}[\varepsilon\nu(\mathcal{D})]^{-1} + \mathbb{E}^{(l)}_{x,\theta}[\tau^k] + \Big(\sup_{\mathcal{D}} V + N\bar{b}\Big).$$

□

4.1.4. *Generalized Poisson equation*

Assume (6). Let $0 < a < 1$, $l \geq 0$ and $0 \leq \beta \leq 1-\alpha$. For $f \in \mathcal{L}_{V^\beta}$ such that $\pi(|f|) < +\infty$, let us define the function

$$\hat{g}^{(l)}_a(x,\theta) \stackrel{\text{def}}{=} \sum_{j\geq 0}(1-a)^{j+1}\mathbb{E}^{(l)}_{x,\theta}[\bar{f}(X_j)].$$

Proposition 4.9. *Assume* (6). *Let* $0 \leq \beta \leq 1-\alpha$ *and* $f \in \mathcal{L}_{V^\beta}$. *For any* $(x,\theta) \in \mathsf{X}\times\Theta$, $l \geq 0$ *and* $0 < a < 1$, $\hat{g}^{(l)}_a$ *exists, and*

$$\bar{f}(x) = \frac{1}{1-a}\hat{g}^{(l)}_a(x,\theta) - \mathbb{E}^{(l)}_{x,\theta}\big[\hat{g}^{(l+1)}_a(X_1,\theta_1)\big].$$

Proof. By Proposition 4.6(i), $|\mathbb{E}^{(l)}_{x,\theta}[\bar{f}(X_j)]| \leq |\bar{f}|_{V^\beta}(V^\beta(x) + \bar{b}j^\beta)$. Hence, $\hat{g}^{(l)}_a(x,\theta)$ exists for any x,θ,l. Furthermore, $\hat{g}^{(l+1)}_a(X_1,\theta_1)$ is $\mathbb{P}^{(l)}_{x,\theta}$-integrable. By the definition of $\hat{g}^{(l)}_a$ and by the Markov property, we have

$$\begin{aligned}\mathbb{E}^{(l)}_{x,\theta}\big[\hat{g}^{(l+1)}_a(X_1,\theta_1)\big] &= \sum_{j\geq 0}(1-a)^{j+1}\mathbb{E}^{(l)}_{x,\theta}[\bar{f}(X_{j+1})] \\ &= (1-a)^{-1}\sum_{j\geq 1}(1-a)^{j+1}\mathbb{E}^{(l)}_{x,\theta}[\bar{f}(X_j)] \\ &= (1-a)^{-1}\big(\hat{g}^{(l)}_a(x,\theta)-(1-a)\bar{f}(x)\big). \qquad \square\end{aligned}$$

Theorem 4.10. *Assume A3–A5 and B2. Let $0\leq\beta<1-\alpha$. For any $\epsilon>0$, there exists an integer $n\geq 2$ such that for any $0<a<1$, $f\in\mathcal{L}_{V^\beta}$, $l\geq 0$, $(x,\theta)\in\mathsf{X}\times\Theta$ and $q\in[1,+\infty]$,*

$$\begin{aligned}(|\bar{f}|_{V^\beta})^{-1}|\hat{g}^{(l)}_a(x,\theta)| \leq{}& 4\epsilon\big(1-(1-a)^n\big)^{-1}n \\ &+ \frac{V^{\beta+\alpha/q}(x)}{a^{1-1/q}(1-a)^{1/q}}(\alpha c)^{-1/q} \\ &\times\Big(1+\bar{b}[\varepsilon\nu(\mathcal{D})]^{-1}+2(1+\bar{b}n_\star)(1+\bar{b})\sup_{\mathcal{D}}V^{\beta+\alpha/q}\Big).\end{aligned}$$

By convention, $1/q=0$ when $q=+\infty$. In particular, $\lim_{a\to 0}(|\bar{f}|_{V^\beta})^{-1}|a\hat{g}^{(l)}_a(x,\theta)|=0$.

Remark 1. Before launching into the proof of the theorem, we first make two important remarks. First, a simplified restatement of Theorem 4.10 is the following: there exists a finite constant c_0 such that for any $0<a\leq 1/2$, $f\in\mathcal{L}_{V^\beta}$, $l\geq 0$, $(x,\theta)\in\mathsf{X}\times\Theta$ and $q\in[1,+\infty]$,

$$\big|\hat{g}^{(l)}_a(x,\theta)\big|\leq c_0|\bar{f}|_{V^\beta}a^{-1}\big(1+a^{1/q}V^{\beta+\alpha/q}(x)\big). \tag{7}$$

This follows by taking $\epsilon=1$, say, and upon noting that $n(1-(1-a)^n)^{-1}\leq 2^{n-1}/a$. The second point is that if we take $a_1,a_2\in(0,1)$, then we can write

$$\hat{g}^{(l)}_{a_1}(x,\theta)-\hat{g}^{(l)}_{a_2}(x,\theta)=\frac{a_2-a_1}{(1-a_1)(1-a_2)}\times\sum_{k\geq 0}(1-a_1)^{k+1}\mathbb{E}^{(l)}_{x,\theta}\big[\hat{g}^{(l+k)}_{a_2}(X_k,\theta_k)\big].$$

By (7) and Proposition 4.6(iii), it holds that

$$\big|\hat{g}^{(l)}_{a_1}(x,\theta)-\hat{g}^{(l)}_{a_2}(x,\theta)\big|\leq c_1|\bar{f}|_{V^\beta}|a_2-a_1|a_2^{-1}a_1^{-2+1/q}V^{\beta+\alpha/q}(x), \tag{8}$$

for some finite constant c_1, for all $0<a_1,a_2\leq 1/2$, $f\in\mathcal{L}_{V^\beta}$, $l\geq 0$, $(x,\theta)\in\mathsf{X}\times\Theta$ and $q\in[1,+\infty]$.

Proof of Theorem 4.10. Let $\epsilon>0$. Let us consider the sequence of stopping times $\{\tau^k, k\geq 0\}$ defined in Section 4.1.3, where $(\mathcal{D},N,n_\star)$ are defined below.

Choice of $\mathcal{D}, N, n_\star$. Choose a level set $\mathcal{D}$ of V large enough so that $\nu(\mathcal{D}) > 0$. Choose N such that

$$\frac{1}{N}\sum_{j=0}^{N-1} \sup_{\mathcal{D}\times\Theta} \|P_\theta^j(x,\cdot) - \pi(\cdot)\|_{V^\beta} \le \epsilon, \tag{9}$$

the existence of which is given by A5, and such that – since $\alpha+\beta<1$ –

$$(\alpha c)^{-1} N^{-1}\Big(\sup_{\mathcal{D}} V^{\beta+\alpha} + \bar{b} N^{\beta+\alpha} + \bar{b}[\varepsilon\nu(\mathcal{D})]^{-1}\Big) \le \epsilon. \tag{10}$$

Set $\epsilon_N \stackrel{\text{def}}{=} N^{-2}\{\epsilon(\sup_{\mathcal{D}} V^\beta + \bar{b} N^{-1}\sum_{j=1}^{N-1} j^\beta)^{-1}\}^{1/(1-\beta)}$ (which can be assumed to be strictly lower than N^{-2} since $\beta > 0$). By B2, choose $n_\star$ such that for any $q \ge n_\star$, $l \ge 0$, $\sup_{\mathcal{D}\times\Theta} \mathbb{P}_{x,\theta}^{(l)}(D(\theta_q, \theta_{q-1}) \ge \epsilon_N/2) \le \epsilon_N/4$.

By Proposition 4.8, $\mathbb{P}_{x,\theta}^{(l)}(\tau^k < +\infty) = 1$ for any $(x,\theta) \in \mathsf{X}\times\Theta$, $l \ge 0$, $k \ge 0$.

Optimal coupling. With these definitions, $\sup_{i\ge 1}\sup_{k\ge 1} \mathbb{E}_{x,\theta}^{(l)}[\mathbb{E}_{Z_{\tau^k}}^{(\tau^k+l)}[D(\theta_i,\theta_{i-1})]] \le \epsilon_N$, upon noting that $\mathbb{P}_{x,\theta}^{(l)}(n_\star \le \tau^k) = 1$ and $D(\theta,\theta') \le 2$. We apply Proposition 4.3 and set $\mathcal{E}_N \stackrel{\text{def}}{=} \{X_k = \tilde{X}_k, 0 \le k < N\}$. We have, for any $l \ge 0$, $k \ge 1$, $(x,\theta) \in \mathsf{X}\times\Theta$,

$$\mathbb{E}_{x,\theta}^{(l)}\big[\overline{\mathbb{P}}_{Z_{\tau^k}, Z_{\tau^k}}^{(\tau^k+l)}(\mathcal{E}_N^c)\big] \le \sum_{j=1}^{N-1}\sum_{i=1}^{j} \mathbb{E}_{x,\theta}^{(l)}\big[\mathbb{E}_{Z_{\tau^k}}^{(\tau^k+l)}[D(\theta_i,\theta_{i-1})]\big] \le N^2\epsilon_N < 1. \tag{11}$$

Observe that $\mathcal{D}$, N and $n_\star$ do not depend upon a, l, x, θ and f.

Proof of Theorem 4.10. Assume that for any $0 < a < 1$, $l \ge 0$, $(x,\theta) \in \mathsf{X}\times\Theta$ and $k \ge 2$,

$$\left|\mathbb{E}_{x,\theta}^{(l)}\left[\sum_{j=0}^{N-1}(1-a)^{\tau^k+j+1}\bar{f}(X_{\tau^k+j})\right]\right| \le |\bar{f}|_{V^\beta} 3N\epsilon(1-a)^{n_\star+(k-1)N}. \tag{12}$$

We have

$$\hat{g}_a^{(l)}(x,\theta) = \sum_{j\ge 0}(1-a)^{j+1}\Big\{\mathbb{E}_{x,\theta}^{(l)}[\bar{f}(X_j)\mathbb{1}_{j<\tau^1}] + \sum_{k\ge 1}\mathbb{E}_{x,\theta}^{(l)}[\bar{f}(X_j)\mathbb{1}_{\tau^k \le j < \tau^{k+1}}]\Big\}.$$

On one hand, by Proposition 4.6(iii) applied with $\tau = \tau_{\mathcal{D}}$ and Proposition 4.8,

$$\begin{aligned}\left|\sum_{j\ge 0}(1-a)^{j+1}\mathbb{E}_{x,\theta}^{(l)}[\bar{f}(X_j)\mathbb{1}_{j<\tau^0}]\right| &= \left|\mathbb{E}_{x,\theta}^{(l)}\left[\sum_{j=0}^{\tau_{\mathcal{D}}-1}(1-a)^{j+1}\bar{f}(X_j)\right]\right| \\ &\le |\bar{f}|_{V^\beta}\mathbb{E}_{x,\theta}^{(l)}\left[\sum_{j=0}^{\tau_{\mathcal{D}}-1}(1-a)^{j+1}V^\beta(X_j)\right] \\ &\le |\bar{f}|_{V^\beta}\frac{V^{\beta+\alpha/q}(x)}{a^{1-1/q}}\frac{(1+\bar{b}[\varepsilon\nu(\mathcal{D})]^{-1})}{(1-a)^{1/q}}(\alpha c)^{-1/q}.\end{aligned}$$

Applied with $\tau = \tau_{\mathcal{D}}$, Propositions 4.6(i), (iii) and 4.8 yield

$$
\begin{aligned}
&|\bar f|_{V^\beta}^{-1}\left|\sum_{j\ge 0}(1-a)^{j+1}\mathbb{E}_{x,\theta}^{(l)}[\bar f(X_j)\mathbb{1}_{\tau^0\le j<\tau^1}]\right| \\
&\quad = |\bar f|_{V^\beta}^{-1}\left|\mathbb{E}_{x,\theta}^{(l)}\left[\sum_{j=\tau_{\mathcal{D}}}^{\tau_{\mathcal{D}}+n_\star+\tau_{\mathcal{D}}\circ\underline{\theta}^{n_\star+\tau_{\mathcal{D}}}-1}(1-a)^{j+1}\bar f(X_j)\right]\right| \\
&\quad \le \mathbb{E}_{x,\theta}^{(l)}\left[\mathbb{E}_{Z_{\tau_{\mathcal{D}}}}^{(\tau_{\mathcal{D}}+l)}\left[\sum_{j=0}^{n_\star+\tau_{\mathcal{D}}\circ\underline{\theta}^{n_\star}-1}(1-a)^{j+1}V^\beta(X_j)\right]\right] \\
&\quad \le \mathbb{E}_{x,\theta}^{(l)}\left[\mathbb{E}_{Z_{\tau_{\mathcal{D}}}}^{(\tau_{\mathcal{D}}+l)}\left[\sum_{j=0}^{n_\star-1}(1-a)^{j+1}V^\beta(X_j)\right]\right]+\mathbb{E}_{x,\theta}^{(l)}\left[\mathbb{E}_{Z_{\tau_{\mathcal{D}}+n_\star}}^{(\tau_{\mathcal{D}}+n_\star+l)}\left[\sum_{j=0}^{\tau_{\mathcal{D}}-1}(1-a)^{j+1}V^\beta(X_j)\right]\right] \\
&\quad \le 2\frac{(1+\bar b n_\star)(1+\bar b)}{a^{1-1/q}(1-a)^{1/q}}(\alpha c)^{-1/q}\sup_{\mathcal{D}} V^{\beta+\alpha/q}.
\end{aligned}
$$

For $k\ge 1$,

$$
\begin{aligned}
&\left|\sum_{j\ge 0}(1-a)^{j+1}\mathbb{E}_{x,\theta}^{(l)}[\bar f(X_j)\mathbb{1}_{\tau^k\le j<\tau^{k+1}}]\right| \\
&\quad \le \left|\mathbb{E}_{x,\theta}^{(l)}\left[\sum_{j=\tau^k}^{\tau^k+N-1}(1-a)^{j+1}\bar f(X_j)\right]\right| \\
&\qquad + \mathbb{E}_{x,\theta}^{(l)}\left[(1-a)^{\tau^k+N}\mathbb{E}_{Z_{\tau^k+N}}^{(\tau^k+N+l)}\left[\sum_{j=0}^{\tau_{\mathcal{D}}-1}(1-a)^{j+1}|\bar f|(X_j)\right]\right].
\end{aligned}
$$

By Proposition 4.6(i) and (ii) applied with $\tau=\tau_{\mathcal{D}}$, Proposition 4.8 and equation (12), and upon noting that $\tau^k \ge n_\star + (k-1)N$ $\mathbb{P}_{(x,\theta)}^{(l)}$-a.s.,

$$
\begin{aligned}
&\left|\sum_{j\ge 0}(1-a)^{j+1}\mathbb{E}_{x,\theta}^{(l)}[\bar f(X_j)\mathbb{1}_{\tau^k\le j<\tau^{k+1}}]\right| \\
&\quad \le |\bar f|_{V^\beta}\mathbb{E}_{x,\theta}^{(l)}\big[(1-a)^{n_\star+(k-1)N}\big(3N\epsilon+(1-a)^N\{V^{\beta+\alpha}(X_{\tau^k+N})+\bar b[\varepsilon\nu(\mathcal{D})]^{-1}\}(\alpha c)^{-1}\big)\big] \\
&\quad \le |\bar f|_{V^\beta}(1-a)^{n_\star+(k-1)N}\Big(3N\epsilon+(\alpha c)^{-1}\sup_{r,\mathcal{D}\times\Theta}\mathbb{E}_{x,\theta}^{(r)}\big[V^{\beta+\alpha}(X_N)+\bar b[\varepsilon\nu(\mathcal{D})]^{-1}\big]\Big) \\
&\quad \le |\bar f|_{V^\beta}(1-a)^{n_\star+(k-1)N}\Big(3N\epsilon+(\alpha c)^{-1}\Big(\sup_{\mathcal{D}}V^{\beta+\alpha}+\bar b N^{\beta+\alpha}+\bar b[\varepsilon\nu(\mathcal{D})]^{-1}\Big)\Big) \\
&\quad \le 4\epsilon|\bar f|_{V^\beta}(1-a)^{(k-1)N}N,
\end{aligned}
$$

where we used the definition of N (see equation (10)) and Proposition 4.6(i). This yields the desired result.

Proof of Equation (12). By the strong Markov property and since $\tau^k \geq n_\star + N(k-1)$ $\mathbb{P}^{(l)}_{x,\theta}$-a.s.,

$$\begin{aligned}&\left|\mathbb{E}^{(l)}_{x,\theta}\left[\sum_{j=0}^{N-1}(1-a)^{\tau^k+j+1}\bar{f}(X_{\tau^k+j})\right]\right|\\&\quad\leq(1-a)^{n_\star+N(k-1)}\mathbb{E}^{(l)}_{x,\theta}\left[\left|\mathbb{E}^{(\tau^k+l)}_{Z_{\tau^k}}\left[\sum_{j=0}^{N-1}(1-a)^{j+1}\bar{f}(X_j)\right]\right|\right].\end{aligned}$$

Furthermore, by Proposition 4.3,

$$\begin{aligned}&\mathbb{E}^{(\tau^k+l)}_{Z_{\tau^k}}\left[\sum_{j=0}^{N-1}(1-a)^{j+1}\bar{f}(X_j)\right]\\&\quad=\overline{\mathbb{E}}^{(\tau^k+l)}_{Z_{\tau^k},Z_{\tau^k}}\left[\sum_{j=0}^{N-1}(1-a)^{j+1}\bar{f}(X_j)\right]\\&\quad=\overline{\mathbb{E}}^{(\tau^k+l)}_{Z_{\tau^k},Z_{\tau^k}}\left[\sum_{j=0}^{N-1}(1-a)^{j+1}\bar{f}(\tilde{X}_j)\right]+\overline{\mathbb{E}}^{(\tau^k+l)}_{Z_{\tau^k},Z_{\tau^k}}\left[\sum_{j=0}^{N-1}(1-a)^{j+1}\{\bar{f}(X_j)-\bar{f}(\tilde{X}_j)\}\mathbb{1}_{\mathcal{E}^c_N}\right].\end{aligned}$$

On one hand, we have $\mathbb{P}^{(l)}_{x,\theta}$-a.s.,

$$\begin{aligned}&\left|\overline{\mathbb{E}}^{(\tau^k+l)}_{Z_{\tau^k},Z_{\tau^k}}\left[\sum_{j=0}^{N-1}(1-a)^{j+1}\bar{f}(\tilde{X}_j)\right]\right|\\&\quad\leq|\bar{f}|_{V^\beta}\sum_{j=0}^{N-1}(1-a)^{j+1}\sup_{\mathcal{D}\times\Theta}\|P^j_\theta(x,\cdot)-\pi(\cdot)\|_{V^\beta}\leq|\bar{f}|_{V^\beta}N\epsilon,\end{aligned}$$

by (9). On the other hand, $\mathbb{P}^{(l)}_{x,\theta}$-a.s.,

$$\begin{aligned}&\left|\overline{\mathbb{E}}^{(\tau^k+l)}_{Z_{\tau^k},Z_{\tau^k}}\left[\sum_{j=0}^{N-1}(1-a)^{j+1}\{\bar{f}(X_j)-\bar{f}(\tilde{X}_j)\}\mathbb{1}_{\mathcal{E}^c_N}\right]\right|\\&\quad\leq|\bar{f}|_{V^\beta}\overline{\mathbb{E}}^{(\tau^k+l)}_{Z_{\tau^k},Z_{\tau^k}}\left[\sum_{j=0}^{N-1}(1-a)^{j+1}\{V^\beta(X_j)+V^\beta(\tilde{X}_j)\}\mathbb{1}_{\mathcal{E}^c_N}\right]\\&\quad\leq|\bar{f}|_{V^\beta}\overline{\mathbb{E}}^{(\tau^k+l)}_{Z_{\tau^k},Z_{\tau^k}}\left[\left(\sum_{j=0}^{N-1}(1-a)^{j+1}\{V^\beta(X_j)+V^\beta(\tilde{X}_j)\}\right)^{\beta^{-1}}\right]^\beta\left(\overline{\mathbb{P}}^{(\tau^k+l)}_{Z_{\tau^k},Z_{\tau^k}}(\mathcal{E}^c_N)\right)^{1-\beta},\end{aligned}$$

by using Jensen's inequality ($\beta < 1$). By the Minkowski inequality, Proposition 4.6(i) and iterating the drift inequality A4, we have

$$\overline{\mathbb{E}}^{(\tau^k+l)}_{Z_{\tau^k},Z_{\tau^k}}\left[\left(\sum_{j=0}^{N-1}(1-a)^{j+1}\{V^\beta(X_j)+V^\beta(\tilde{X}_j)\}\right)^{\beta^{-1}}\right]^\beta$$
$$\le \sum_{j=0}^{N-1}(1-a)^{j+1}\{\overline{\mathbb{E}}^{(\tau^k+l)}_{Z_{\tau^k},Z_{\tau^k}}[V(X_j)]^\beta+\overline{\mathbb{E}}^{(\tau^k+l)}_{Z_{\tau^k},Z_{\tau^k}}[V(\tilde{X}_j)]^\beta\}$$
$$\le \sum_{j=0}^{N-1}(1-a)^{j+1}\left\{\sup_{l,\mathcal{D}\times\Theta}\left(\mathbb{E}^{(l)}_{x,\theta}[V(X_j)]\right)^\beta+\left(\sup_{\mathcal{D}\times\Theta}P^j_\theta V(x)\right)^\beta\right\}$$
$$\le 2\sum_{j=0}^{N-1}(1-a)^{j+1}\left(\sup_{\mathcal{D}} V+j\bar{b}\right)^\beta \le 2N\left(\sup_{\mathcal{D}} V^\beta+\bar{b}N^{-1}\sum_{j=1}^{N-1} j^\beta\right).$$

Finally,

$$\mathbb{E}^{(l)}_{x,\theta}\left[\left(\overline{\mathbb{P}}^{(\tau^k+l)}_{Z_{\tau^k},Z_{\tau^k}}(\mathcal{E}^c_N)\right)^{1-\beta}\right] \le \left(\mathbb{E}^{(l)}_{x,\theta}\left[\overline{\mathbb{P}}^{(\tau^k+l)}_{Z_{\tau^k},Z_{\tau^k}}(\mathcal{E}^c_N)\right]\right)^{1-\beta} \le (N^2\epsilon_N)^{1-\beta},$$

where we used (11) in the last inequality. To conclude the proof, we use the definition of ϵ_N. □

4.2. Proof of Theorem 2.1

Let $\epsilon > 0$. We prove that there exists n_ϵ such that for any $n \ge n_\epsilon$, $\sup_{\{f,|f|_1\le 1\}}|\mathbb{E}_{\xi_1,\xi_2}[\bar{f}(X_n)]| \le \epsilon$.

4.2.1. *Definition of $\mathcal{D}$, N, Q and $n_\star$*

By A1(i), choose Q such that

$$\sup_l \sup_{(x,\theta)\in\mathcal{C}\times\Theta} \mathbb{E}^{(l)}_{x,\theta}[\mathbf{r}(\tau_{\mathcal{C}})]\sum_{k\ge Q}\frac{1}{\mathbf{r}(k)} \le \epsilon. \tag{13}$$

By A1(ii), choose N such that

$$\sup_{(x,\theta)\in\mathcal{C}\times\Theta} V^{-1}(x)\|P^N_\theta(x,\cdot)-\pi(\cdot)\|_{\mathrm{TV}} \le \frac{\epsilon}{Q}. \tag{14}$$

By B1, choose $n_\star$ such that, for any $n \ge n_\star$,

$$\mathbb{P}_{\xi_1,\xi_2}\left(D(\theta_n,\theta_{n-1}) \ge \epsilon/\left(2(N+Q-1)^2Q\right)\right) \le \frac{\epsilon}{4(N+Q-1)^2Q}. \tag{15}$$

4.2.2. *Optimal coupling*

We apply Proposition 4.3 with $l = 0$ and $N \leftarrow N + Q$. Set $\mathcal{E}_{N+Q} \overset{\text{def}}{=} \{X_k = \tilde{X}_k, 0 \le k \le N + Q\}$. It holds for any $r \ge n_\star$ that

$$\begin{aligned}
&\mathbb{E}_{\xi_1,\xi_2}\big[\mathbb{1}_{X_r\in\mathcal{C}}\overline{\mathbb{P}}^{(r)}_{Z_r,Z_r}(\mathcal{E}^c_{N+Q})\big] \\
&\quad \le \sum_{j=1}^{N+Q-1}\sum_{i=1}^{j}\mathbb{E}_{\xi_1,\xi_2}\big[\mathbb{1}_{X_r\in\mathcal{C}}\mathbb{E}^{(r)}_{Z_r}[D(\theta_i,\theta_{i-1})]\big] \\
&\quad \le \sum_{j=1}^{N+Q-1}\sum_{i=1}^{j}\mathbb{E}_{\xi_1,\xi_2}[D(\theta_{i+r},\theta_{i+r-1})] \le \epsilon Q^{-1},
\end{aligned} \tag{16}$$

where, in the last inequality, we use the fact that $D(\theta, \theta') \le 2$ and the definition of $n_\star$ (see equation (15)).

4.2.3. *Proof*

Let $n \ge N + Q + n_\star$. We consider the partition given by the last exit from the set $\mathcal{C}$ before time $n - N$. We use the notation $\{X_{n:m} \notin \mathcal{C}\}$ as shorthand for $\bigcap_{k=n}^{m}\{X_k \notin \mathcal{C}\}$, with the convention that $\{X_{m+1:m} \notin \mathcal{C}\} = \Omega$. We write

$$\mathbb{E}_{\xi_1,\xi_2}[\bar{f}(X_n)] = \mathbb{E}_{\xi_1,\xi_2}[\bar{f}(X_n)\mathbb{1}_{X_{0:n-N}\notin\mathcal{C}}] + \sum_{k=0}^{n-N}\mathbb{E}_{\xi_1,\xi_2}[\bar{f}(X_n)\mathbb{1}_{X_k\in\mathcal{C}}\mathbb{1}_{X_{k+1:n-N}\notin\mathcal{C}}].$$

Since $\bar{f}$ is bounded on X by $|\bar{f}|_1$, we have

$$\mathbb{E}_{\xi_1,\xi_2}[\bar{f}(X_n)\mathbb{1}_{X_{0:n-N}\notin\mathcal{C}}] \le |\bar{f}|_1\mathbb{P}_{\xi_1,\xi_2}(\tau_{\mathcal{C}} \ge n-N) \le |\bar{f}|_1\mathbb{E}_{\xi_1,\xi_2}\left[\frac{\tau_{\mathcal{C}}}{n-N}\wedge 1\right].$$

The right-hand side is upper bounded by $|\bar{f}|_1\epsilon$ for n large enough. By definition of Q in (13),

$$\begin{aligned}
&\sum_{k=0}^{n-(N+Q)}\mathbb{E}_{\xi_1,\xi_2}[\bar{f}(X_n)\mathbb{1}_{X_k\in\mathcal{C}}\mathbb{1}_{X_{k+1:n-N}\notin\mathcal{C}}] \\
&\le |\bar{f}|_1\sum_{k=0}^{n-(N+Q)}\mathbb{E}_{\xi_1,\xi_2}\big[\mathbb{1}_{X_k\in\mathcal{C}}\mathbb{P}^{(k)}_{X_k,\theta_k}(\tau_{\mathcal{C}} \ge n-N-k)\big] \\
&\le |\bar{f}|_1\sup_l\sup_{\mathcal{C}\times\Theta}\mathbb{E}^{(l)}_{x,\theta}[\mathbf{r}(\tau_{\mathcal{C}})]\sum_{k\ge Q}\frac{1}{\mathbf{r}(k)} \le |\bar{f}|_1\epsilon.
\end{aligned} \tag{17}$$

Let $k \in \{n-(N+Q)+1, \dots, n-N\}$. By definition of N and $n_\star$ (see equations (14) and (15)), upon noting that $k \geq n-(N+Q) \geq n_\star$, we have

$$\begin{aligned}
&\mathbb{E}_{\xi_1,\xi_2}[\bar{f}(X_n)\mathbb{1}_{X_k \in \mathcal{C}}\mathbb{1}_{X_{k+1:n-N}\notin \mathcal{C}}] - |\bar{f}|_1 \mathbb{E}_{\xi_1,\xi_2}\big[\mathbb{1}_{X_k\in\mathcal{C}}\overline{\mathbb{P}}^{(k)}_{Z_k,Z_k}(\mathcal{E}^c_{N+Q})\big] \\
&\quad \leq \mathbb{E}_{\xi_1,\xi_2}\big[\mathbb{1}_{X_k\in\mathcal{C}}\overline{\mathbb{E}}^{(k)}_{Z_k,Z_k}[\bar{f}(X_{n-k})\mathbb{1}_{X_{1:n-N-k}\notin\mathcal{C}}\mathbb{1}_{\mathcal{E}_{N+Q}}]\big] \\
&\quad \leq \mathbb{E}_{\xi_1,\xi_2}\big[\mathbb{1}_{X_k\in\mathcal{C}}\overline{\mathbb{E}}^{(k)}_{Z_k,Z_k}[\bar{f}(\tilde{X}_{n-k})\mathbb{1}_{\tilde{X}_{1:n-N-k}\notin\mathcal{C}}\mathbb{1}_{\mathcal{E}_{N+Q}}]\big] \\
&\quad \leq \mathbb{E}_{\xi_1,\xi_2}\big[\mathbb{1}_{X_k\in\mathcal{C}}\overline{\mathbb{E}}^{(k)}_{Z_k,Z_k}[\bar{f}(\tilde{X}_{n-k})\mathbb{1}_{\tilde{X}_{1:n-N-k}\notin\mathcal{C}}]\big] + |\bar{f}|_1\mathbb{E}_{\xi_1,\xi_2}\big[\mathbb{1}_{X_k\in\mathcal{C}}\overline{\mathbb{P}}^{(k)}_{Z_k,Z_k}(\mathcal{E}^c_{N+Q})\big] \\
&\quad \leq \mathbb{E}_{\xi_1,\xi_2}\big[\mathbb{1}_{X_k\in\mathcal{C}}\overline{\mathbb{E}}^{(k)}_{Z_k,Z_k}[\mathbb{1}_{\tilde{X}_{1:n-N-k}\notin\mathcal{C}}P^N_{\theta_k}\bar{f}(\tilde{X}_{n-N-k})]\big] + |\bar{f}|_1\epsilon Q^{-1} \\
&\quad \leq |\bar{f}|_1\epsilon Q^{-1}\mathbb{E}_{\xi_1,\xi_2}\big[\mathbb{1}_{X_k\in\mathcal{C}}\overline{\mathbb{E}}^{(k)}_{Z_k,Z_k}[\mathbb{1}_{\tilde{X}_{1:n-N-k}\notin\mathcal{C}}V(\tilde{X}_{n-N-k})]\big] + |\bar{f}|_1\epsilon Q^{-1} \\
&\quad \leq |\bar{f}|_1\epsilon Q^{-1}\Big\{\sup_{(x,\theta)\in\mathcal{C}\times\Theta} P_\theta V(x) + \sup_{\mathcal{C}} V\Big\} + |\bar{f}|_1\epsilon Q^{-1},
\end{aligned}$$

where we used A1(iii) in the last inequality. Hence,

$$\sum_{k=n-(N+Q)+1}^{n-N} \mathbb{E}_{\xi_1,\xi_2}[\bar{f}(X_n)\mathbb{1}_{X_k\in\mathcal{C}}\mathbb{1}_{X_{k+1:n-N}\notin\mathcal{C}}] \leq \Big(1+\sup_{(x,\theta)\in\mathcal{C}\times\Theta} P_\theta V(x) + \sup_{\mathcal{C}} V\Big)\epsilon|\bar{f}|_1.$$

This concludes the proof.

Remark 2. In the case where the process is non-adaptive, we can assume without loss of generality that it possesses an atom α; in that case, the lines (17) can be modified so that the assumptions $\sum_n\{1/\mathbf{r}(n)\} < +\infty$ can be removed. In the case of an atomic chain, we can indeed apply the above computations with $\mathcal{C}$ replaced by α and write

$$\begin{aligned}
\sum_{k=0}^{n-(N+Q)} \mathbb{E}_{\xi_1}[\bar{f}(X_n)\mathbb{1}_{X_k\in\alpha}\mathbb{1}_{X_{k+1:n-N}\notin\alpha}] &\leq |\bar{f}|_1 \sum_{k=0}^{n-(N+Q)} \mathbb{P}_\alpha(\tau_\alpha \geq n-N-k) \\
&\leq |\bar{f}|_1 \sum_{k\geq Q}\mathbb{P}_\alpha(\tau_\alpha \geq k).
\end{aligned}$$

The right-hand side is small for convenient Q, provided that $\mathbb{E}_\alpha[\mathbf{r}(\tau_\alpha)] < +\infty$ with $\mathbf{r}(n) = n$. Unfortunately, the adaptive chain $\{(X_n,\theta_n), n\geq 0\}$ does not possess an atom, thus explaining the condition on $\mathbf{r}$.

4.3. Proof of Corollary 2.2

The condition A1(ii) is established in Appendix A. Let $\mathcal{D}$ be a level set large enough such that $\nu(\mathcal{D}) > 0$; Proposition 4.8 then implies that there exists a constant $c < \infty$ such that for any $l \geq 0$,

$\mathbb{E}^{(l)}_{x,\theta}[\tau_{\mathcal{D}}] \le cV(x)$. This implies that for $0 < \eta \le 1-\alpha$,

$$\mathbb{E}^{(l)}_{x,\theta}\left[\sum_{k=0}^{\tau_{\mathcal{D}}}(k+1)^{\eta}\right] \le \mathbb{E}^{(l)}_{x,\theta}\left[\sum_{k=0}^{\tau_{\mathcal{D}}}\bigl(\mathbb{E}^{(k+l)}_{X_k,\theta_k}[\tau_{\mathcal{D}}]\bigr)^{\eta}\right] \le c^{\eta}\mathbb{E}^{(l)}_{x,\theta}\left[\sum_{k=0}^{\tau_{\mathcal{D}}} V^{1-\alpha}(X_k)\right]$$
$$\le C\bigl(V(x) + b\mathbb{E}^{(l)}_{x,\theta}[\tau_{\mathcal{D}}]\bigr) \le C'V(x),$$

for some finite constants C, C' independent of θ. Hence, A1(i) holds with $\mathbf{r}(n) \sim n^{1+\eta}$. Finally, $P_\theta V \le V - cV^{1-\alpha} + b\mathbb{1}_{\mathcal{C}}$ implies that $P_\theta V \le V - c\gamma V^{1-\alpha} + b\mathbb{1}_{\mathcal{D}}$ for any $\gamma \in (0,1)$ and the level set $\mathcal{D} \stackrel{\text{def}}{=} \{x, V^{1-\alpha} \le b[c(1-\gamma)]^{-1}\}$. This yields A1(iii).

4.4. Proof of Proposition 2.4

Under A2, there exists a constant C – that does not depend on θ – such that for any $(x,\theta) \in \mathsf{X} \times \Theta$, $n \ge 0$ and $\kappa \in [1, \alpha^{-1}]$,

$$\|P^n_\theta(x,\cdot) - \pi(\theta)\|_{\mathrm{TV}} \le C\frac{V^{\kappa\alpha}(x)}{(n+1)^{\kappa-1}};$$

see Appendix A. To apply Theorem 13 of Roberts and Rosenthal (2007), we only have to prove that there exists $\kappa \in [1, \alpha^{-1}]$ such that the sequence $\{V^{\kappa\alpha}(X_n); n \ge 0\}$ is bounded in probability, which is equivalent to proving that $\{V^{\beta}(X_n); n \ge 0\}$ is bounded in probability for some (and thus any) $\beta \in (0,1]$. This is a consequence of Lemma 4.11 applied with $W = V^{\beta}$ for some $\beta \in (0,1]$ and $\mathbf{r}(n) = (n+1)^{1+\eta}$ for some $\eta > 0$ (see the proof of Corollary 2.2 for similar computations).

Lemma 4.11. *Assume that there exist a set $\mathcal{C}$ and functions $W : \mathsf{X} \to (0,+\infty)$ and $\mathbf{r} : \mathbb{N} \to (0,+\infty)$ such that $\mathbf{r}$ is non-decreasing, $P_\theta W \le W$ on $\mathcal{C}^c$ and*

$$\sup_{\mathcal{C}\times\Theta} P_\theta W < +\infty, \qquad \sup_{l}\sup_{\mathcal{C}\times\Theta} \mathbb{E}^{(l)}_{x,\theta}[\mathbf{r}(\tau_{\mathcal{C}})] < +\infty, \qquad \sum_k \{1/\mathbf{r}(k)\} < +\infty.$$

For any probability distributions ξ_1, ξ_2 on X, Θ, respectively, $\{W(X_n), n \ge 0\}$ is bounded in probability for the probability $\mathbb{P}_{\xi_1,\xi_2}$.

Proof. Let $\epsilon > 0$. We prove that there exist M_ϵ, N_ϵ such that for any $M \ge M_\epsilon$ and $n \ge N_\epsilon$, $\mathbb{P}_{x,\theta}(W(X_n) \ge M) \le \epsilon$. Choose N_ϵ such that for any $n \ge N_\epsilon$

$$\mathbb{E}_{\xi_1,\xi_2}\left[\frac{\tau_{\mathcal{C}}}{n} \wedge 1\right] \le \epsilon/3, \qquad \sup_{l}\sup_{\mathcal{C}\times\Theta} \mathbb{E}^{(l)}_{x,\theta}[\mathbf{r}(\tau_{\mathcal{C}})] \sum_{k\ge n}\{1/\mathbf{r}(k)\} \le \epsilon/3,$$

and choose M_ϵ such that for any $M \ge M_\epsilon$, $N_\epsilon \sup_{\mathcal{C}\times\Theta} P_\theta W \le \epsilon M/3$. We write

$$\mathbb{P}_{\xi_1,\xi_2}\bigl(W(X_n) \ge M\bigr) = \sum_{k=0}^{n-1}\mathbb{P}_{\xi_1,\xi_2}\bigl(W(X_n) \ge M, X_k \in \mathcal{C}, X_{k+1:n} \notin \mathcal{C}\bigr)$$
$$+ \mathbb{P}_{\xi_1,\xi_2}\bigl(W(X_n) \ge M, X_{0:n} \notin \mathcal{C}\bigr).$$

By the Markov inequality, for $n \geq N_\epsilon$,

$$\mathbb{P}_{\xi_1,\xi_2}\big(W(X_n) \geq M, X_{0:n} \notin \mathcal{C}\big) \leq \mathbb{P}_{\xi_1,\xi_2}(X_{0:n} \notin \mathcal{C}) \leq \mathbb{P}_{\xi_1,\xi_2}(\tau_\mathcal{C} > n) \leq \mathbb{E}_{\xi_1,\xi_2}\left[\frac{\tau_\mathcal{C}}{n} \wedge 1\right] \leq \epsilon/3.$$

Furthermore, for $n \geq N_\epsilon$,

$$\begin{aligned}
&\sum_{k=0}^{n-N_\epsilon} \mathbb{P}_{\xi_1,\xi_2}\big(W(X_n) \geq M, X_k \in \mathcal{C}, X_{k+1:n} \notin \mathcal{C}\big) \\
&\quad \leq \sum_{k=0}^{n-N_\epsilon} \mathbb{P}_{\xi_1,\xi_2}(X_k \in \mathcal{C}, X_{k+1:n} \notin \mathcal{C}) \\
&\quad \leq \sum_{k=0}^{n-N_\epsilon} \mathbb{E}_{\xi_1,\xi_2}\Big[\mathbb{1}_\mathcal{C}(X_k) \sup_l \sup_{\mathcal{C}\times\Theta} \mathbb{P}^{(l)}_{x,\theta}(X_{1:n-k} \notin \mathcal{C})\Big] \leq \sum_{k=0}^{n-N_\epsilon} \sup_l \sup_{\mathcal{C}\times\Theta} \mathbb{P}^{(l)}_{x,\theta}(\tau_\mathcal{C} \geq n-k) \\
&\quad \leq \sum_{k=N_\epsilon}^{n} \frac{1}{\mathbf{r}(k)} \sup_l \sup_{\mathcal{C}\times\Theta} \mathbb{E}^{(l)}_{x,\theta}[\mathbf{r}(\tau_\mathcal{C})] \leq \epsilon/3.
\end{aligned}$$

Finally, for $n \geq N_\epsilon$, we write

$$\begin{aligned}
&\sum_{k=n-N_\epsilon+1}^{n} \mathbb{P}_{x,\theta}\big(W(X_n) \geq M, X_k \in \mathcal{C}, X_{k+1:n} \notin \mathcal{C}\big) \\
&\quad \leq \sum_{k=n-N_\epsilon+1}^{n} \mathbb{E}_{x,\theta}\big[\mathbb{1}_\mathcal{C}(X_k)\mathbb{P}^{(k)}_{X_k,\theta_k}\big(W(X_{n-k}) \geq M, X_{1:n-k} \notin \mathcal{C}\big)\big].
\end{aligned}$$

We have, for any $k \in \{n - N_\epsilon + 1, \dots, n\}$ and $(x, \theta) \in \mathcal{C} \times \Theta$,

$$\mathbb{P}^{(k)}_{x,\theta}\big(W(X_{n-k}) \geq M, X_{1:n-k} \notin \mathcal{C}\big) \leq \frac{1}{M}\mathbb{E}^{(k)}_{x,\theta}[W(X_{n-k})\mathbb{1}_{\mathcal{C}^c}(X_{1:n-k-1})] \leq \frac{1}{M}\mathbb{E}^{(k)}_{x,\theta}[W(X_1)],$$

where, in the last inequality, we used the drift inequality on W outside $\mathcal{C}$. Hence,

$$\sum_{k=n-N_\epsilon+1}^{n} \mathbb{P}_{x,\theta}\big(W(X_n) \geq M, X_k \in \mathcal{C}, X_{k+1:n} \notin \mathcal{C}\big) \leq \frac{N_\epsilon}{M} \sup_{\mathcal{C}\times\Theta} P_\theta W(x) \leq \epsilon/3.$$

The proof is thus complete. $\square$

4.5. Proof of Theorem 2.5

By using the function $\hat{g}_a^{(l)}$ introduced in Section 4.1.4 and by Proposition 4.9, we write, $\mathbb{P}_{x,\theta}$-a.s.,

$$\begin{aligned}
n^{-1}\sum_{k=1}^{n}\bar{f}(X_k) &= n^{-1}\sum_{k=1}^{n}\Big((1-a)^{-1}\hat{g}_a^{(k)}(X_k,\theta_k)-\mathbb{E}^{(k)}_{X_k,\theta_k}\big[\hat{g}_a^{(k+1)}(X_1,\theta_1)\big]\Big)\\
&= n^{-1}(1-a)^{-1}\sum_{k=1}^{n}\big\{\hat{g}_a^{(k)}(X_k,\theta_k)-\mathbb{E}_{x,\theta}\big[\hat{g}_a^{(k)}(X_k,\theta_k)|\mathcal{F}_{k-1}\big]\big\}\\
&\quad + n^{-1}(1-a)^{-1}\sum_{k=1}^{n}\big\{\mathbb{E}_{x,\theta}\big[\hat{g}_a^{(k)}(X_k,\theta_k)|\mathcal{F}_{k-1}\big]\\
&\qquad\qquad - (1-a)\mathbb{E}_{x,\theta}\big[\hat{g}_a^{(k+1)}(X_{k+1},\theta_{k+1})|\mathcal{F}_k\big]\big\}\\
&= n^{-1}(1-a)^{-1}\sum_{k=1}^{n}\big\{\hat{g}_a^{(k)}(X_k,\theta_k)-\mathbb{E}_{x,\theta}\big[\hat{g}_a^{(k)}(X_k,\theta_k)|\mathcal{F}_{k-1}\big]\big\}\\
&\quad + n^{-1}(1-a)^{-1}\big\{\mathbb{E}_{x,\theta}\big[\hat{g}_a^{(1)}(X_1,\theta_1)|\mathcal{F}_0\big]-\mathbb{E}_{x,\theta}\big[\hat{g}_a^{(n+1)}(X_{n+1},\theta_{n+1})|\mathcal{F}_n\big]\big\}\\
&\quad + an^{-1}(1-a)^{-1}\sum_{k=1}^{n}\mathbb{E}_{x,\theta}\big[\hat{g}_a^{(k+1)}(X_{k+1},\theta_{k+1})|\mathcal{F}_k\big].
\end{aligned}$$

We apply the above inequalities with $a=a_n$ and consider the different terms in turn. We show that they tend $\mathbb{P}_{x,\theta}$-a.s. to zero when the deterministic sequence $\{a_n, n\ge 1\}$ satisfies conditions which are verified, for example, with $a_n=(n+1)^{-\zeta}$ for some ζ such that

$$\zeta>0,\qquad 2\zeta<1-\big(0.5\vee\beta(1-\alpha)^{-1}\big),\qquad \zeta<1-\beta(1-\alpha)^{-1}.$$

To prove that each term converges a.s. to zero, we use the following characterization:

$$\Big[\forall\epsilon>0,\ \lim_{n\to+\infty}\mathbb{P}\Big(\sup_{m\ge n}|X_m|\ge\epsilon\Big)\Big]\quad\Longleftrightarrow\quad [\{X_n,n\ge 0\}\to 0\ \mathbb{P}\text{-a.s.}].$$

Hereafter, we assume that $|f|_{V^\beta}=1$. In the following, c (and, below, c_1, c_2) is a constant, the value of which may vary from appearance to appearance.

Convergence of Term 1. *Set $p\stackrel{\text{def}}{=}(1-\alpha)/\beta$. We prove that*

$$n^{-1}(1-a_n)^{-1}\sum_{k=1}^{n}\big\{\hat{g}_{a_n}^{(k)}(X_k,\theta_k)-\mathbb{E}_{\xi_1,\xi_2}\big[\hat{g}_{a_n}^{(k)}(X_k,\theta_k)|\mathcal{F}_{k-1}\big]\big\}\longrightarrow 0\qquad \mathbb{P}_{\xi_1,\xi_2}\text{-a.s.},$$

provided that the sequence $\{a_n, n\ge 0\}$ is non-increasing, $\lim_{n\to\infty} n^{\max(1/p,1/2)-1}/a_n=0$, $\sum_n n^{-1}[n^{\max(1/p,1/2)-1}/a_n]^p<+\infty$ and $\sum_n |a_n-a_{n-1}|a_{n-1}^{-2}[n^{\max(1/p,1/2)-1}/a_n]<+\infty$.

Proof. Define $D_{n,k} \stackrel{\text{def}}{=} \hat{g}_{a_n}^{(k)}(X_k,\theta_k) - \mathbb{E}_{\xi_1,\xi_2}[\hat{g}_{a_n}^{(k)}(X_k,\theta_k)|\mathcal{F}_{k-1}]$; $S_{n,k} \stackrel{\text{def}}{=} \sum_{j=1}^k D_{n,j}$ if $k \le n$ and $S_{n,k} \stackrel{\text{def}}{=} \sum_{j=1}^n D_{n,j} + \sum_{j=n+1}^k D_{j,j}$ if $k > n$; and $R_n \stackrel{\text{def}}{=} \sum_{j=1}^{n-1} D_{n,j} - D_{n-1,j}$. Then, for each n, $\{(S_{n,k},\mathcal{F}_k), k \ge 1\}$ is a martingale. For $k > n$ and by Lemma B.1, there exists a universal constant C such that

$$\begin{aligned}\mathbb{E}_{\xi_1,\xi_2}[|S_{n,k}|^p] &\le Ck^{\max(p/2,1)-1}\left(\sum_{j=1}^n \mathbb{E}_{\xi_1,\xi_2}[|D_{n,j}|^p] + \sum_{j=n+1}^k \mathbb{E}_{\xi_1,\xi_2}[|D_{j,j}|^p]\right)\\ &\le c_1|\bar{f}|_{V^\beta} k^{\max(p/2,1)-1} a_k^{-p} \sum_{j=1}^k \mathbb{E}_{\xi_1,\xi_2}[V(X_j)] \qquad (18)\\ &\le c_1|\bar{f}|_{V^\beta} k^{\max(p/2,1)} a_k^{-p}\xi_1(V),\end{aligned}$$

where we used (7) and Proposition 4.6(ii). It follows that for any $n \ge 1$, $\lim_{N\to\infty} N^{-p}\mathbb{E}_{\xi_1,\xi_2}(|S_{n,N}|^p) \le c_1 \lim_{N\to\infty}(N^{\max(1/p,1/2)-1}/a_N)^p = 0$. Then, by the martingale array extension of the Chow–Birnbaum–Marshall inequality (Lemma B.2),

$$\begin{aligned}&2^{-p}\delta^p \mathbb{P}_{\xi_1,\xi_2}\left(\sup_{m\ge n} m^{-1}(1-a_m)^{-1}\left|\sum_{j=1}^n D_{n,j}\right| > \delta\right)\\ &\quad\le \sum_{k=n}^\infty \bigl(k^{-p}-(k+1)^{-p}\bigr)\mathbb{E}_{\xi_1,\xi_2}[|S_{n,k}|^p] + \left(\sum_{k=n+1}^\infty k^{-1}\mathbb{E}_{\xi_1,\xi_2}^{1/p}[|R_k|^p]\right)^p.\end{aligned}$$

Under the assumptions on the sequence $\{a_n, n \ge 0\}$ and given the bound (18), the first term in the right-hand side tends to zero as $n \to +\infty$. To bound the second term, we first note that $\{(\sum_{j=1}^k D_{n,j} - D_{n-1,j}, \mathcal{F}_k), k \ge 1\}$ is a martingale for each n. Therefore, by Lemma B.1 and the definition of $D_{n,j}$,

$$\begin{aligned}\mathbb{E}_{\xi_1,\xi_2}[|R_n|^p] &\le Cn^{\max(p/2,1)-1}\sum_{j=1}^{n-1}\mathbb{E}_{\xi_1,\xi_2}[|D_{n,j}-D_{n-1,j}|^p]\\ &\le 2Cn^{\max(p/2,1)-1}\sum_{j=1}^{n-1}\mathbb{E}_{\xi_1,\xi_2}\bigl[\bigl|\hat{g}_{a_n}^{(j)}(X_j,\theta_j) - \hat{g}_{a_{n-1}}^{(j)}(X_j,\theta_j)\bigr|^p\bigr].\end{aligned}$$

Then, using (8) (with $q = \infty$) and the usual argument for bounding moments of $V^\beta(X_j)$, we get

$$\mathbb{E}_{\xi_1,\xi_2}^{1/p}[|R_n|^p] \le c_1|\bar{f}|_{V^\beta} n^{\max(1/2,1/p)}|a_n - a_{n-1}|a_n^{-1}a_{n-1}^{-2}\xi_1(V).$$

Under the assumptions, $\sum_n n^{-1}\mathbb{E}_{\xi_1,\xi_2}^{1/p}[|R_n|^p] < +\infty$ and this completes the proof. □

Convergence of Term 2. *We prove that*

$$n^{-1}(1-a_n)^{-1}\mathbb{E}_{\xi_1,\xi_2}\bigl[\hat{g}_{a_n}^{(1)}(X_1,\theta_1)|\mathcal{F}_0\bigr] \longrightarrow 0,$$

provided that $\lim_n na_n = +\infty$ *and* $\lim_n a_n = 0$.

Proof. By Theorem 4.10 applied with $q = +\infty$, it may be proven that there exist constants c, N such that

$$\left|\mathbb{E}_{\xi_1,\xi_2}\left[\hat{g}^{(1)}_{a_n}(X_1,\theta_1)|\mathcal{F}_0\right]\right| \leq c a_n^{-1}\xi_1(V) + c\left(1-(1-a_n)^N\right)^{-1} N.$$

Divided by $n^{-1}(1-a_n)$, the right-hand side tends to zero as $n \to +\infty$. □

Convergence of Term 3. *We prove that*

$$n^{-1}(1-a_n)^{-1}\mathbb{E}_{\xi_1,\xi_2}\left[\hat{g}^{(n+1)}_{a_n}(X_{n+1},\theta_{n+1})|\mathcal{F}_n\right] \longrightarrow 0 \qquad \mathbb{P}_{\xi_1,\xi_2}\text{-}a.s.,$$

provided that the sequence $\{n^{-1}a_n^{-1}, n \geq 1\}$ *is non-increasing*, $\lim_n n^{1-\beta(1-\alpha)^{-1}} a_n = +\infty$, $\sum_n (na_n)^{-(1-\alpha)\beta^{-1}} < +\infty$ *and* $\lim_n a_n = 0$.

Proof. There exist constants c_1, c_2, N such that for any n large enough (i.e., such that $1 - a_n \geq 1/2$) and $p \stackrel{\text{def}}{=} (1-\alpha)\beta^{-1} > 1$,

$$\begin{aligned}
&\mathbb{P}_{\xi_1,\xi_2}\left(\sup_{m\geq n} m^{-1}(1-a_m)^{-1}\left|\mathbb{E}_{\xi_1,\xi_2}\left[\hat{g}^{(m+1)}_{a_m}(X_{m+1},\theta_{m+1})|\mathcal{F}_m\right]\right| \geq \delta\right) \\
&\quad \leq 2^p\delta^{-p}\mathbb{E}_{\xi_1,\xi_2}\left[\sup_{m\geq n} m^{-p}\left|\mathbb{E}_{\xi_1,\xi_2}\left[\hat{g}^{(m+1)}_{a_m}(X_{m+1},\theta_{m+1})|\mathcal{F}_m\right]\right|^p\right] \\
&\quad \leq 2^p\delta^{-p}\sum_{m\geq n} m^{-p}\mathbb{E}_{\xi_1,\xi_2}\left[\left|\mathbb{E}_{\xi_1,\xi_2}\left[\hat{g}^{(m+1)}_{a_m}(X_{m+1},\theta_{m+1})|\mathcal{F}_m\right]\right|^p\right] \\
&\quad \leq 2^p\delta^{-p}\sum_{m\geq n} m^{-p}\mathbb{E}_{\xi_1,\xi_2}\left[\left|\hat{g}^{(m+1)}_{a_m}(X_{m+1},\theta_{m+1})\right|^p\right] \\
&\quad \leq 2^{2p-1}\delta^{-p}\sum_{m\geq n} m^{-p}\left\{\frac{c_1}{a_m^p}\mathbb{E}_{\xi_1,\xi_2}[V^{\beta p}(X_{m+1})] + c_2\left(\frac{N}{(1-(1-a_m)^N)}\right)^p\right\},
\end{aligned}$$

where we used Theorem 4.10 with $q = +\infty$. Furthermore, by Propositions 4.6(i), 4.7 and the drift inequality,

$$\begin{aligned}
&\mathbb{P}_{\xi_1,\xi_2}\left(\sup_{m\geq n} m^{-1}(1-a_m)^{-1}\left|\mathbb{E}_{\xi_1,\xi_2}\left[\hat{g}^{(n+1)}_{a_m}(X_{m+1},\theta_{m+1})|\mathcal{F}_m\right]\right| \geq \delta\right) \\
&\quad \leq \frac{2^p c_3}{\delta^p}\left\{n^{-p}a_n^{-p}\mathbb{E}_{\xi_1,\xi_2}[V(X_n)] + \sum_{m\geq n} m^{-p}a_m^{-p} + \sum_{m\geq n} m^{-p}\left(\frac{N}{(1-(1-a_m)^N)}\right)^p\right\} \\
&\quad \leq \frac{2^p c_3}{\delta^p}\left\{n^{-p}a_n^{-p}\left(\xi_1(V) + n\bar{b}\right) + \bar{b}\sum_{m\geq n} m^{-p}a_m^{-p} + \sum_{m\geq n} m^{-p}\left(\frac{N}{(1-(1-a_m)^N)}\right)^p\right\}.
\end{aligned}$$

Under the stated conditions on $\{a_n, n \geq 1\}$, the right-hand side tends to zero as $n \to +\infty$. □

Convergence of Term 4. *We prove that*

$$a_n n^{-1}(1-a_n)^{-1}\sum_{k=1}^{n}\mathbb{E}_{\xi_1,\xi_2}\big[\hat{g}_{a_n}^{(k+1)}(X_{k+1},\theta_{k+1})|\mathcal{F}_k\big]\longrightarrow 0 \qquad \mathbb{P}_{\xi_1,\xi_2}\text{-}a.s.,$$

provided that $\{a_n^{1\wedge[(1-\alpha-\beta)/\alpha]}n^{-1}, n\geq 1\}$ *is non-increasing,* $\sum_n a_n^{1\wedge[(1-\alpha-\beta)/\alpha]}n^{-1}<+\infty$ *and* $\lim_n a_n = 0$.

Proof. Choose $q\geq 1$ such that $\beta+\alpha/q\leq 1-\alpha$. Fix $\epsilon>0$. From Theorem 4.10, there exist constants C, N such that for any $n\geq 1$, $l\geq 0$, $(x,\theta)\in\mathsf{X}\times\Theta$,

$$\big|\hat{g}_{a_n}^{(l)}(x,\theta)\big|\leq Ca_n^{1/q-1}V^{\beta+\alpha/q}(x)+4\epsilon N\big(1-(1-a_n)^N\big)^{-1}.$$

Hence, for n large enough such that $(1-a_n)\geq 1/2$, we have

$$\begin{aligned}&\left|a_n n^{-1}(1-a_n)^{-1}\sum_{k=1}^{n}\mathbb{E}_{\xi_1,\xi_2}\big[\hat{g}_{a_n}^{(k+1)}(X_{k+1},\theta_{k+1})|\mathcal{F}_k\big]\right|\\ &\quad\leq 8a_n\epsilon N\big(1-(1-a_n)^N\big)^{-1}+2Ca_n^{1/q}n^{-1}\sum_{k=1}^{n}\mathbb{E}_{\xi_1,\xi_2}[V^{\beta+\alpha/q}(X_{k+1})|\mathcal{F}_k]\\ &\quad\leq 8a_n\epsilon N\big(1-(1-a_n)^N\big)^{-1}+2Ca_n^{1/q}n^{-1}\sum_{k=1}^{n}V^{1-\alpha}(X_k)+2Ca_n^{1/q}\bar{b},\end{aligned}$$

where we used $\beta+\alpha/q\leq 1-\alpha$ and Proposition 4.6(i) in the last inequality. Since $\lim_n a_n=0$ and $\lim_n a_n\epsilon N(1-(1-a_n)^N)^{-1}=\epsilon$, we only have to prove that $a_n^{1/q}n^{-1}\sum_{k=1}^{n}V^{1-\alpha}(X_k)$ converges to zero $\mathbb{P}_{\xi_1,\xi_2}$-a.s. By Kronecker's lemma (see, e.g., Section 2.6 of Hall and Heyde (1980)), this amounts to proving that $\sum_{k\geq 1}a_k^{1/q}k^{-1}V^{1-\alpha}(X_k)$ is finite a.s. This property holds upon noting that, by Proposition 4.7 and Proposition 4.6(i),

$$\begin{aligned}\mathbb{E}_{\xi_1,\xi_2}\left[\sum_{k\geq n}a_k^{1/q}k^{-1}V^{1-\alpha}(X_k)\right]&\leq a_n^{1/q}n^{-1}\mathbb{E}_{\xi_1,\xi_2}[V(X_n)]+\sum_{k\geq n}a_k^{1/q}k^{-1}\\ &\leq a_n^{1/q}n^{-1}\big(\xi_1(V)+\bar{b}n\big)+\sum_{k\geq n}a_k^{1/q}k^{-1}\end{aligned}$$

and the right-hand side tends to zero under the stated assumptions. □

4.6. Proof of Proposition 2.6

We only give the sketch of the proof since the proof is very similar to that of Theorem 2.5. We start by proving a result similar to Theorem 4.10. Since $\mathcal{D}=\mathsf{X}$, the sequence $\{\tau^k, k\geq 0\}$ is deterministic and $\tau^{k+1}=\tau^k+N+1$. By adapting the proof of Theorem 4.10 (f is bounded and

$\mathcal{D} = \mathsf{X}$), we establish that for any $\epsilon > 0$, there exists an integer $n \geq 2$ such that for any $0 < a < 1$, any bounded function f, $l \geq 0$ and $(x, \theta) \in \mathsf{X} \times \Theta$,

$$(|\bar{f}|_1)^{-1} \bigl|\hat{g}_a^{(l)}(x, \theta)\bigr| \leq n + \epsilon\bigl(1 - (1-a)^n\bigr)^{-1} n.$$

We then introduce the martingale decomposition as in the proof of Theorem 2.5 and follow along the same lines (with any $p > 1$).

Appendix A: Explicit control of convergence

We provide sufficient conditions for the assumptions A1(ii) and A5. The technique relies on the explicit control of convergence of a transition kernel P on a general state space $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ to its stationary distribution π.

Proposition A.1. *Let P be a ϕ-irreducible and aperiodic transition kernel on $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$.*

(i) *Assume that there exist a probability measure ν on $\mathbb{T}$, positive constants ε, b, c, a measurable set $\mathcal{C}$, a measurable function $V : \mathbb{T} \to [1, +\infty)$ and $0 < \alpha \leq 1$ such that*

$$P(x, \cdot) \geq \mathbb{1}_{\mathcal{C}}(x) \varepsilon \nu(\cdot), \qquad PV \leq V - cV^{1-\alpha} + b\mathbb{1}_{\mathcal{C}}. \tag{19}$$

Then P possesses an invariant probability measure π and $\pi(V^{1-\alpha}) < +\infty$.

(ii) *Assume, in addition, that $c \inf_{\mathcal{C}^c} V^{1-\alpha} \geq b$, $\sup_{\mathcal{C}} V < +\infty$ and $\nu(\mathcal{C}) > 0$. There then exists a constant C depending on $\sup_{\mathcal{C}} V$, $\nu(\mathcal{C})$ and $\varepsilon, \alpha, b, c$, such that for any $0 \leq \beta \leq 1 - \alpha$ and $1 \leq \kappa \leq \alpha^{-1}(1 - \beta)$,*

$$(n+1)^{\kappa - 1} \|P^n(x, \cdot) - \pi(\cdot)\|_{V^\beta} \leq C V^{\beta + \alpha\kappa}(x). \tag{20}$$

Proof. Conditions (19) imply that V is unbounded off petite set and P is recurrent. They also imply that $\{V < +\infty\}$ is full and absorbing. Hence, there exists a level set $\mathcal{D}$ of V large enough so that $\nu(\mathcal{D}) > 0$. Following along the same lines as in the proof of Proposition 4.8, we prove that $\sup_{\mathcal{D}} \mathbb{E}_x[\tau_{\mathcal{D}}] < +\infty$. The proof of (i) in concluded by Theorems 8.4.3 and 10.0.1 of Meyn and Tweedie (1993). The proof of (ii) is given in, for example, Fort and Moulines (2003) (see also Andrieu and Fort (2005); Douc *et al.* (2007)). □

When $b \leq c$, $c \inf_{\mathcal{C}^c} V^{1-\alpha} \geq b$. Otherwise, it is easy to deduce the conditions of (ii) from conditions of the form (i).

Corollary A.2. *Let P be a ϕ-irreducible and aperiodic transition kernel on $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$. Assume that there exist positive constants b, c, a measurable set $\mathcal{C}$, an unbounded measurable function $V : \mathbb{T} \to [1, +\infty)$ and $0 < \alpha \leq 1$ such that $PV \leq V - cV^{1-\alpha} + b\mathbb{1}_{\mathcal{C}}$. Assume, in addition, that the level sets of V are 1-small. There then exist a level set $\mathcal{D}$ of V, positive constants $\varepsilon_{\mathcal{D}}, c_{\mathcal{D}}$ and a probability measure $\nu_{\mathcal{D}}$ such that*

$$P(x, \cdot) \geq \mathbb{1}_{\mathcal{D}}(x) \varepsilon_{\mathcal{D}} \nu_{\mathcal{D}}(\cdot), \qquad PV \leq V - c_{\mathcal{D}} V^{1-\alpha} + b\mathbb{1}_{\mathcal{D}},$$

$\sup_{\mathcal{D}} V < +\infty$, $\nu_{\mathcal{D}}(\mathcal{D}) > 0$, *and* $c_{\mathcal{D}} \inf_{\mathcal{D}^c} V^{1-\alpha} \geq b$.

Proof. For any $0 < \gamma < 1$, $PV \leq V - \gamma c V^{1-\alpha} + b\mathbb{1}_{\mathcal{D}_\gamma}$ with $\mathcal{D}_\gamma \stackrel{\text{def}}{=} \{V^{1-\alpha} \leq b[c(1-\gamma)]^{-1}\}$. Hence, $\sup_{\mathcal{D}_\gamma} V < +\infty$ and for γ close to 1, we have $\gamma c \inf_{\mathcal{D}_\gamma^c} V^{1-\alpha} \geq b$. Finally, the drift condition (19) implies that the set $\{V < +\infty\}$ is full and absorbing and thus the level sets $\{V \leq d\}$ are accessible for any d large enough. □

The 1-smallness assumption is usually done for convenience and is not restrictive. In the case where the level sets are petite (and thus m-small for some $m \geq 1$), the explicit upper bounds get intricate and are never given a detailed treatment in the literature (at least not in the polynomial case). Nevertheless, it is a recognized fact that the bounds derived in the case $m = 1$ can be extended to the case $m > 1$.

Appendix B: L^p-martingales and the Chow–Birnbaum–Marshall inequality

We deal with martingales and martingale arrays in the paper using the following two results.

Lemma B.1. *Let* $\{(D_k, \mathcal{F}_k), 1 \leq k \geq 1\}$ *be a martingale difference sequence and* $M_n = \sum_{k=1}^n D_k$. *For any* $p > 1$,

$$\mathbb{E}[|M_n|^p] \leq C n^{\max(p/2,1)-1} \sum_{k=1}^n \mathbb{E}(|D_k|^p), \tag{21}$$

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

Proof. By Burkholder's inequality (Hall and Heyde (1980), Theorem 2.10) applied to the martingale $\{(M_n, \mathcal{F}_n), n \geq 1\}$, we get

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

where $C = (18pq^{1/2})^p$, $p^{-1} + q^{-1} = 1$. The proof follows by noting that

$$\left(\sum_{k=1}^n |D_k|^2\right)^{p/2} \leq n^{\max(p/2,1)-1} \sum_{k=1}^n |D_k|^p. \tag{22}$$

To prove (22), note that if $1 < p \leq 2$, then the convexity inequality $(a+b)^\alpha \leq a^\alpha + b^\alpha$, which holds true for all $a, b \geq 0$ and $0 \leq \alpha \leq 1$, implies that $(\sum_{n=1}^n |D_k|^2)^{p/2} \leq \sum_{k=1}^n |D_k|^p$. If $p > 2$, then Hölder's inequality gives $(\sum_{k=1}^n |D_k|^2)^{p/2} \leq n^{p/2-1}(\sum_{k=1}^n |D_k|^p)$. □

Lemma B.2 can be found in Atchade (2009) and provides a generalization to the classical Chow–Birnbaum–Marshall inequality.

Lemma B.2. *Let* $\{D_{n,i}, \mathcal{F}_{n,i}, 1 \leq i \leq n\}, n \geq 1$, *be a martingale-difference array and* $\{c_n, n \geq 1\}$ *a non-increasing sequence of positive numbers. Assume that* $\mathcal{F}_{n,i} = \mathcal{F}_i$ *for all* i, n. *Define*

$$S_{n,k} \stackrel{\text{def}}{=} \sum_{i=1}^{k} D_{n,i}, \qquad \text{if } 1 \leq k \leq n \quad \text{and} \quad S_{n,k} \stackrel{\text{def}}{=} \sum_{i=1}^{n} D_{n,i} + \sum_{j=n+1}^{k} D_{j,j}, \qquad \text{if } k > n;$$

$$R_n \stackrel{\text{def}}{=} \sum_{j=1}^{n-1} (D_{n,j} - D_{n-1,j}).$$

For $n \leq m \leq N$, $p \geq 1$ *and* $\lambda > 0$,

$$\begin{aligned}2^{-p}\lambda^p \mathbb{P}\Big(\max_{n \leq m \leq N} c_m |M_{m,m}| > \lambda\Big) \leq c_N^p \mathbb{E}(|S_{n,N}|^p) + \sum_{j=n}^{N-1} (c_j^p - c_{j+1}^p) \mathbb{E}(|S_{n,j}|^p) \\ + \mathbb{E}\left[\left(\sum_{j=n+1}^{N} c_j |R_j|\right)^p\right].\end{aligned} \tag{23}$$

Appendix C: Proofs for Section 3.2

In the proofs, C will denote a generic finite constant whose actual value might change from one appearance to the next. The proofs below differ from those in earlier works (see, e.g., Fort and Moulines (2000); Douc *et al.* (2004)) since q is not assumed to be compactly supported.

C.1. Proof of Lemma 3.3

Lemma C.1. *Assume* D1 *and* D2. *For all* x *large enough and* $|z| \leq \eta |x|^{\upsilon}$, $t \mapsto V_s(x + tz)$ *is twice continuously differentiable on* $[0, 1]$. *There exist a constant* $C < +\infty$ *and a positive function* ε *such that* $\lim_{|x|\to\infty} \varepsilon(x) = 0$, *such that for all* x *large enough,* $|z| \leq \eta|x|^{\upsilon}$ *and* $s \leq s_\star$,

$$\sup_{t\in[0,1]} |\nabla^2 V_s(x + tz)| \leq C s V_s(x) |x|^{2(m-1)} \big(s + \varepsilon(x)\big).$$

Proof. $|x + z| \geq |x| - \eta|x|^{\upsilon} \geq (1 - \eta)|x|^{\upsilon}$ so that $t \mapsto V_s(x + tz)$ is twice continuously differentiable on $[0, 1]$ for $|x|$ large enough. We have

$$\begin{aligned}|\nabla^2 V_s(x + tz)| \leq s V_s(x) \frac{V_s(x + tz)}{V_s(x)} |\nabla \ln \pi(x + tz) \nabla \ln \pi(x + tz)^{\mathrm{T}}| \cdots \\ \times \left(s + \frac{|\nabla^2 \ln \pi(x + tz)|}{|\nabla \ln \pi(x + tz) \nabla \ln \pi(x + tz)^{\mathrm{T}}|}\right).\end{aligned}$$

Under the stated assumptions, there exists a constant C such that for any x large enough and $|z| \le \eta |x|^{\upsilon}$,

$$\sup_{t\in[0,1]} \left(s + \frac{|\nabla^2 \ln \pi(x+tz)|}{|\nabla \ln \pi(x+tz) \nabla \ln \pi(x+tz)^{\mathrm{T}}|} \right) \le s + \frac{D_2}{d_1^2(1-\eta)} |x|^{-m\upsilon}$$

and

$$\sup_{t\in[0,1]} |\nabla \ln \pi(x+tz) \nabla \ln \pi(x+tz)^{\mathrm{T}}| \le |x|^{2(m-1)} D_1^2 (1-\eta|x|^{\upsilon-1})^{2(m-1)}.$$

Finally,

$$\sup_{t\in[0,1], s\le s_\star} \left(\frac{\pi(x+tz)}{\pi(x)} \right)^{-s} \le 1 + s_\star D_1 |z| \sup_{t\in[0,1]} |x+tz|^{m-1} \sup_{t\in[0,1], s\le s_\star} \left(\frac{\pi(x+tz)}{\pi(x)} \right)^{-s},$$

which yields the desired result upon noting that $|z||x+tz|^{m-1} \le \eta |x|^{\upsilon+m-1}(1-\eta|x|^{\upsilon-1})$ is arbitrarily small for x large enough. □

We now turn to the proof of Lemma 3.3. For $x \in \mathsf{X}$, define $R(x) := \{y \in \mathsf{X} : \pi(y) < \pi(x)\}$ and $R(x) - x \stackrel{\text{def}}{=} \{y - x : y \in R(x)\}$. We have

$$\begin{aligned} P_\theta V_s(x) - V_s(x) &= \int \big(V_s(x+z) - V_s(x)\big) q_\theta(z) \mu_{\text{Leb}}(\mathrm{d}z) \\ &\quad + \int_{R(x)-x} \big(V(x+z) - V(x)\big) \left(\frac{\pi(x+z)}{\pi(x)} - 1 \right) q_\theta(z) \mu_{\text{Leb}}(\mathrm{d}z). \end{aligned}$$

If x remains in a compact set $\mathcal{C}$, then using D2(ii) and the continuity of $x \mapsto V_s(x)$, we have $V_s(x+z) \le C(1 + \exp(s D_0 |z|^m))$. It follows that

$$\sup_{\theta\in\Theta} \sup_{x\in\mathcal{C}} \{P_\theta V_s(x) - V_s(x)\} \le C \sup_{\theta\in\Theta} \int_{R(x)-x} \big(1 + \exp(s D_0 |z|^m)\big) q_\theta(z) \mu_{\text{Leb}}(\mathrm{d}z) < +\infty.$$

More generally, let x be sufficiently large. Define $l(x) \stackrel{\text{def}}{=} \log \pi(x)$, $R_V(x,z) \stackrel{\text{def}}{=} V_s(x+z) - V_s(x) + s V_s(x) \langle z, \nabla l(x) \rangle$, $R_\pi(x,z) \stackrel{\text{def}}{=} \pi(x+z)(\pi(x))^{-1} - 1 - \langle z, \nabla l(x) \rangle$. Using the fact that the mean of q_θ is zero, we can write $P_\theta V_s(x) - V_s(x) = I_1(x,\theta,s) + I_2(x,\theta,s) + I_3(x,\theta,s)$, where

$$I_1(x,\theta,s) \stackrel{\text{def}}{=} -s V_s(x) \int_{R(x)-x} \langle z, \nabla l(x) \rangle^2 q_\theta(z) \mu_{\text{Leb}}(\mathrm{d}z),$$

$$I_2(x,\theta,s) \stackrel{\text{def}}{=} \int R_V(x,z) q_\theta(z) \mu_{\text{Leb}}(\mathrm{d}z) + \int_{R(x)-x} R_V(x,z) \left(\frac{\pi(x+z)}{\pi(x)} - 1 \right) q_\theta(z) \mu_{\text{Leb}}(\mathrm{d}z)$$

and

$$I_3(x,\theta,s) \stackrel{\text{def}}{=} -s V_s(x) \int_{R(x)-x} R_\pi(x,z) \langle z, \nabla l(x) \rangle q_\theta(z) \mu_{\text{Leb}}(\mathrm{d}z).$$

C.1.1. *First term*

It follows from (Fort and Moulines (2000), Lemma B.3 and proof of Proposition 3) that, under D2(i), there exists $b > 0$ such that for all $\theta \in \Theta$,

$$\int_{R(x)-x} \langle z, \nabla l(x) \rangle^2 q_\theta(z) \mu_{\mathrm{Leb}}(\mathrm{d}z) \geq b |\nabla l(x)|^2.$$

Hence, $\sup_{\theta\in\Theta} I_1(x,\theta,s) \leq -s V_s(x) b d_1^2 |x|^{2(m-1)}$.

C.1.2. *Second term*

For $z \in R(x)-x$, $\pi(x+z) < \pi(x)$. Therefore, $|I_2(x,\theta,s)| \leq 2\int |R_V(x,z)| q_\theta(z)\mu_{\mathrm{Leb}}(\mathrm{d}z)$. By Lemma C.1, there exists $C < +\infty$ – independent of s for $s \leq s_\star$ – such that for any $|z| \leq \eta |x|^\upsilon$,

$$|R_V(x,z)| \leq C s V_s(x) |x|^{2(m-1)} |z|^2 \bigl(s + \varepsilon(x)\bigr).$$

This implies that there exists a constant $C < +\infty$ – independent of s for $s \leq s_\star$ – such that

$$\begin{aligned}
\int |R_V(x,z)| q_\theta(z)\mu_{\mathrm{Leb}}(\mathrm{d}z) &\leq C s V_s(x) |x|^{2(m-1)} \bigl(s+\varepsilon(x)\bigr) \int |z|^2 q_\theta(z)\mu_{\mathrm{Leb}}(\mathrm{d}z) \\
&\quad + V_s(x) \int_{\{z, |z|\geq \eta |x|^\upsilon\}} \frac{V_s(x+z)}{V_s(x)} q_\theta(z)\mu_{\mathrm{Leb}}(\mathrm{d}z) \\
&\quad + C V_s(x) |x|^{m-1} \int_{\{z, |z|\geq \eta |x|^\upsilon\}} |z| q_\theta(z)\mu_{\mathrm{Leb}}(\mathrm{d}z).
\end{aligned}$$

There exists a constant C such that for $\theta \in \Theta$ and $s \leq s_\star$, the first term in the right-hand side is upper bounded by $C s V_s(x) |x|^{2(m-1)}(s + \varepsilon(x))$. Under D3, the second term is upper bounded by $V_s(x)|x|^{2(m-1)}\varepsilon(x)$ with $\lim_{|x|\to+\infty} \varepsilon(x) = 0$ uniformly in θ for $\theta \in \Theta$, and in s for $s \leq s_\star$. Since q_θ is a multivariate Gaussian distribution, there exists $\lambda_\star > 0$ such that $\sup_{\theta\in\Theta} \int \exp(\lambda_\star |z|^2) q_\theta(z)\mu_{\mathrm{Leb}}(\mathrm{d}z) < +\infty$. Under D3, the third term is upper bounded by $C V_s(x)|x|^{2(m-1)} \exp(-\lambda\eta^2 |x|^{2\upsilon})$ for some $\lambda \in (0, \lambda_\star)$, uniformly in θ for $\theta \in \Theta$, and in s for $s \leq s_\star$. Hence, we have proved that there exists $C_\star < \infty$ such that for any $s \leq s_\star$,

$$\sup_{\theta\in\Theta} |I_2(x,\theta,s)| \leq C_\star V_s(x) |x|^{2(m-1)} \bigl(s^2 + \varepsilon(x)\bigr)$$

for a positive function ε independent of s and such that $\lim_{|x|\to+\infty} \varepsilon(x) = 0$.

C.1.3. *Third term*

Following along the same lines as in the control of $I_2(x,\theta,s)$, it may be proven that

$$\begin{aligned}
I_3(x,\theta,s) &\leq s V_s(x) D_1 |x|^{m-1} \int_{\{z, |z|\geq \eta |x|^\upsilon\}} |z|(1 + D_1 |z||x|^{m-1}) q_\theta(z)\mu_{\mathrm{Leb}}(\mathrm{d}z) \\
&\quad + C V_s(x) |x|^{3(m-1)} \int_{\{z, |z|\leq \eta |x|^\upsilon\}} |z|^3 q_\theta(z)\mu_{\mathrm{Leb}}(\mathrm{d}z) \leq C V_s(x) |x|^{2(m-1)} \varepsilon(x)
\end{aligned}$$

for a positive function ε independent of s, θ and such that $\lim_{|x|\to+\infty} \varepsilon(x) = 0$.

C.1.4. Conclusion

Let $\alpha \in (0, 1)$. By combining the above calculations, we prove that by choosing s small enough such that $c_\star \stackrel{\text{def}}{=} bd_1^2 - C_\star s > 0$, we have

$$\sup_{\theta\in\Theta} P_\theta V_s(x) \leq V_s(x) - c_\star V_s(x)|x|^{2(m-1)} + b_\star \mathbb{1}_{\mathcal{C}}(x) \tag{24}$$

$$\leq V_s(x) - 0.5 c_\star V_s^{1-\alpha}(x) + b_\star \mathbb{1}_{\mathcal{C}}(x) \tag{25}$$

for a compact set $\mathcal{C}$. This proves A2(ii) and A4. A5 follows from the results of Appendix A. A2(iii) and A3 follow from Lemma 3.2.

C.2. Proof of Lemma 3.4

An easy modification in the proof of Proposition 11 in Andrieu and Moulines (2006) (to adjust for the difference in the drift function) shows that $D(\theta, \theta') \leq 2\int_{\mathsf{X}} |q_{e^c\Sigma}(x) - q_{e^{c'}\Sigma'}(x)|\mu_{\text{Leb}}(\mathrm{d}x)$. We then apply Lemma 12 from Andrieu and Moulines (2006) to obtain that $D(\theta, \theta') \leq C|e^c\Sigma - e^{c'}\Sigma'|_{\mathrm{s}}$, where C is a finite constant depending on the compact Θ. Hereafter, C is finite and its value may change from one appearance to the next. For any $l, n \geq 0$, $\epsilon > 0$, $x \in \mathbb{R}^p$ and $\theta \in \Theta$, we have

$$\begin{aligned}\mathbb{P}^{(l)}_{x,\theta}\big(D(\theta_n, \theta_{n+1}) \geq \epsilon\big) &\leq \epsilon^{-1}\mathbb{E}^{(l)}_{x,\theta}[D(\theta_n, \theta_{n+1})] \\ &\leq C\mathbb{E}^{(l)}_{x,\theta}[|c_{n+1} - c_n| + |\Sigma_{n+1} - \Sigma_n|_{\mathrm{s}}] \\ &\leq C(l+n+1)^{-1}\Big(1 + \mathbb{E}^{(l)}_{x,\theta}[|X_{n+1}|^2] + \sqrt{\mathbb{E}^{(l)}_{x,\theta}[|X_{n+1}|^2]}\Big).\end{aligned}$$

D2(ii) implies that we can find $C < \infty$ such that $|x|^2 \leq C\phi(V_s(x))$ for all $x \in \mathsf{X}$, where $\phi(t) = [\ln t]^{2/m}$. From the drift condition (Lemma 3.3), Proposition 4.6(i) and the concavity of ϕ, we deduce that there exists C such that $\mathbb{E}^{(l)}_{x,\theta}[|X_n|^2] \leq C[\ln V_s(x)]^{2/m}[\ln n]^{2/m}$. We conclude that for any probability ξ_1 such that $\xi_1([\ln V_s]^{2/m}) < +\infty$, $\lim_n \mathbb{P}_{\xi_1,\xi_2}(D(\theta_n, \theta_{n+1}) \geq \epsilon) = 0$ and for any level set $\mathcal{D}$ of V_s,

$$\lim_{n\to\infty} \sup_{l\geq 0} \sup_{\mathcal{D}\times\Theta} \mathbb{P}^{(l)}_{x,\theta}\big(D(\theta_n, \theta_{n+1}) \geq \epsilon\big) = 0.$$

Acknowledgements

We would like to thank Michael Woodroofe for helpful discussions on the resolvent approach to limit theorems and Prof. Pierre Priouret and Christophe Andrieu for helpful discussions. We also thank M. Vihola for helpful comments. This work was partly supported by the French National Research Agency (ANR) under the program ANR-05-BLAN-0299.

References

Andrieu, C. and Atchade, Y.F. (2007). On the efficiency of adaptive MCMC algorithms. *Electron. Commun. Probab.* **12** 336–349. MR2350572

Andrieu, C. and Fort, G. (2005). Explicit control of subgeometric ergodicity. Technical report, Univ. Bristol, 05:17. Available at http://www.tsi.enst.fr/~gfort/biblio.html.

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

Andrieu, C. and Robert, C.P. (2001). Controlled MCMC for optimal sampling. Technical report, Univ. Paris Dauphine, Ceremade 0125.

Andrieu, C. and Tadic, V. (2008). General result for the stability of controlled MCMC. Technical report, Bristol University. (Personal communication).

Atchade, Y. and Fort, G. (2008). Limit theorems for some adaptive MCMC algorithms with subgeometric kernels (II). Technical report. To appear.

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

Atchade, Y.F. (2009). A strong law of large numbers for martingale arrays. Technical report, Univ. of Michigan. Available at http://www.stat.lsa.umich.edu/~yvesa/.

Atchade, Y.F. and Rosenthal, J.S. (2005). On adaptive Markov chain Monte Carlo algorithm. *Bernoulli* **11** 815–828. MR2172842

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

Benveniste, A., Métivier, M. and Priouret, P. (1987). *Adaptive Algorithms and Stochastic Approximations*. Springer.

Chen, H., Guo, L. and Gao, A. (1988). Convergence and robustness of the Robbins–Monro algorithm truncated at randomly varying bounds. *Stochastic Process. Appl.* **27** 217–231. MR0931029

Chen, H. and Zhu, Y. (1986). Stochastic approximation procedures with randomly varying truncations. *Sci. Sinica. Ser. A* **29** 914–926. MR0869196

Douc, R., Fort, G., Moulines, E. and Soulier, P. (2004). Practical drift conditions for subgeometric rates of convergence. *Ann. Appl. Probab.* **14** 1353–1377. MR2071426

Douc, R., Moulines, E. and Soulier, P. (2007). Computable convergence rates for sub-geometric ergodic Markov chains. *Bernoulli* **13** 831–848. MR2348753

Fort, G. and Moulines, E. (2000). V-subgeometric ergodicity for a Hastings–Metropolis algorithm. *Statist. Probab. Lett.* **49** 401–410. MR1796485

Fort, G. and Moulines, E. (2003). Polynomial ergodicity of Markov transition kernels. *Stochastic Process. Appl.* **103** 57–99. MR1947960

Gilks, W.R., Roberts, G.O. and Sahu, S.K. (1998). Adaptive Markov chain Monte Carlo through regeneration. *J. Amer. Statist. Assoc.* **93** 1045–1054. MR1649199

Haario, H., Saksman, E. and Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli* **7** 223–242. MR1828504

Hall, P. and Heyde, C.C. (1980). *Martingale Limit Theory and Its Application*. New York: Academic Press. MR0624435

Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their application. *Biometrika* **57** 97–109.

Holden, L. (1998). Adaptive chains. Technical report. Norwegian Computing Center.

Jarner, S.F. and Hansen, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.* **85** 341–361. MR1731030

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

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

Merlevede, F., Peligrad, M. and Utev, S. (2006). Recent advances in invariances principles for stationary sequences. *Probab. Surv.* **3** 1–36. MR2206313

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953). Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087–1092.

Meyn, S.P. and Tweedie, R.L. (1993). *Markov Chains and Stochastic Stability*. London: Springer. MR1287609

Roberts, G. and Rosenthal, J. (2004). General state space Markov chains and MCMC algorithms. *Probab. Surv.* **1** 20–71. MR2095565

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

Roberts, G.O. and Rosenthal, J.S. (2007). Coupling and ergodicity of adaptive MCMC. *J. Appl. Probab.* **44** 458–475. MR2340211

Roberts, G.O. and Tweedie, R.L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83** 95–110. MR1399158

Winkler, G. (2003). *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods*, 2nd ed. *Applications of Mathematics (New York)* **27**. Berlin: Springer. MR1950762

Yang, C. (2007). Recurrent and ergodic properties of Adaptive MCMC. Technical report, Univ. Toronto, Canada. Available at http://probability.ca/jeff/ftpdir/chao3.pdf.

Received July 2007 and revised February 2009