LIMIT THEOREMS FOR SOME ADAPTIVE MCMC ALGORITHMS WITH SUBGEOMETRIC KERNELS

YVES ATCHADÉ AND GERSENDE FORT

July 2008

Abstract. This paper deals with the ergodicity and the existence of a strong law of large numbers for adaptive Markov Chain Monte Carlo. We show that a diminishing adaptation assumption together with a drift condition for positive recurrence is enough to imply ergodicity. Strengthening the drift condition to a polynomial drift condition yields a strong law of large numbers for possibly unbounded functions. These results broaden considerably 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. (2001) when the target distribution is sub-exponential in the tails.

1. Introduction

This paper deals with the convergence of Adaptive Markov Chain Monte Carlo (AMCMC). Markov Chain Monte Carlo (MCMC) is a well known, widely used method to sample from arbitrary probability distributions. One of the major limitation 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 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 $\{\bar{P}\left(n; (x, \theta); (dx', d\theta')\right), n \geq 0\}$

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

Key words and phrases. Adaptive Markov chain Monte Carlo, Markov chain, Subgeometric ergodicity.

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

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

This work is partly supported by the french National Research Agency (ANR) under the program ANR-05-BLAN-0299.

satisfying $\int_\Theta \bar{P}\left(n; (x,\theta); (\cdot, d\theta')\right) = P_\theta(x,\cdot)$. Often, these transition kernels are of the form $\{P_\theta(x,dy)\delta_{H_n(\theta,y)}(d\theta'), n \geq 0\}$ where $\{H_l,\ l \geq 0\}$ is a family 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 $\int H(\theta,x)\pi(dx) = 0$. Since the pioneer 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 in recent years. But despite many recent works on the topic, the asymptotic behavior of these algorithms is still not completely understood. Almost all previous works on the convergence of AMCMC are limited to the case when each kernel P_θ is geometrically ergodic (see e.g.. Roberts and Rosenthal (2007); Andrieu and Moulines (2006)). In this paper, we weaken this condition and consider the case when each transition kernel is sub-geometrically ergodic.

More specifically, we study the ergodicity of the marginal $\{X_n, n \geq 0\}$ i.e. 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 positive recurrence uniform-in-θ drift condition towards a small set C (assumptions A1-2) and a uniform-in-θ ergodicity condition of the kernels $\{P_\theta, \theta \in \Theta\}$ (assumption A3) are enough to imply the ergodicity of AMCMC. This result solves a problem left open in Roberts and Rosenthal (2007) and we believe is close to be 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 Harris recurrent kernel.

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 towards 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 number 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 (Meyn and Tweedie (1993)).

On a more technical note, this paper makes two key contributions to the analysis of AMCMC. Firstly, to study the ergodicity, we use a more careful coupling technique which extends the coupling approach of Roberts and Rosenthal (2007). Secondly, we tackle the law of large numbers using a resolvent kernel approach together with martingales theory. This approach has a decisive advantage over the more classical Poisson equation approach

(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 all 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 comparison 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 sub-exponential in the tails. This is covered in Section 3 together with a toy example taken from Atchade and Rosenthal (2005). All the proofs are postponed to Section 4.

2. Statement of the results and discussion

2.1. **Notations.** For a transition kernel P on a measurable general state space $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$, denote by P^n, $n \geq 0$, its n-th iterate defined as

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

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

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

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

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

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

Let X, Θ be two general state space resp. endowed with a countably generated σ-field $\mathcal{X}$ 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 $A \in \mathcal{X}$,

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

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 when $\{(X_n, \theta_n), n \geq 0\}$ 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

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

A special case is the case when $p_{n+1} = 1$ 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,

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

Such a situation occurs for example if 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 $\{\bar{P}\left(n; \cdot, \cdot\right), \ n \geq 0\}$ 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 $\bar{P}_l\left(n; \cdot, \cdot\right) \stackrel{\text{def}}{=} \bar{P}\left(l + n; \cdot, \cdot\right)$ and we denote by $\mathbb{P}^{(l)}_{x,\theta}$ and $\mathbb{E}^{(l)}_{x,\theta}$ the probability and expectation 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 notations $\mathbb{P}_{x,\theta}$ and $\mathbb{E}_{x,\theta}$ as shorthand notations for $\mathbb{P}^{(0)}_{x,\theta}$ and $\mathbb{E}^{(0)}_{x,\theta}$.

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. These assumptions imply that each transition kernel P_θ is ϕ-irreducible, aperiodic and possesses an unique invariant probability measure π.

A1 There exist a probability measure ν on X, a positive constant ε and a set $\mathcal{C} \in \mathcal{X}$ such that for any $\theta \in \Theta$, $P_\theta(x, \cdot) \geq \mathbb{1}_{\mathcal{C}}(x) \ \varepsilon\nu(\cdot)$.

A2 There exist a measurable function $V : \mathsf{X} \to [1, +\infty)$ and a positive constant b such that for any $\theta \in \Theta$, $P_\theta V \leq V - 1 + b\mathbb{1}_{\mathcal{C}}$.

A3 There exists a probability measure π such that for any level set $\mathcal{D} \stackrel{\text{def}}{=} \{x \in \mathsf{X}, V(x) \leq d\}$ of V,

$$\lim_{n \to +\infty} \ \sup_{\mathcal{D} \times \Theta} \|P^n_\theta(x, \cdot) - \pi\|_{\mathrm{TV}} = 0 \ .$$

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

Theorem 2.1. *Assume A1-3, B1 and $\xi_1(V) < +\infty$. Then*

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

Assumption A2 is designed to control the behavior of the chain "far from the center". When the state space X is "bounded" so that for example, the conditions A3 and B1 hold with $\mathcal{D} = \mathsf{X}$, assumptions A1-A2 are unnecessary.

Proposition 2.2. *Assume B1 and that there exists a probability measure π such that $\lim_{n\to+\infty} \sup_{\mathsf{X}\times\Theta} \|P_\theta^n(x,\cdot) - \pi\|_{\mathrm{TV}} = 0$. Then*

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

In that case, Proposition 2.2 coincides with (Roberts and Rosenthal, 2007, Theorem 5).

2.3. **Strong law of large numbers.** Assumptions A2-3 and B1 are strengthened as follows

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

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

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

Theorem 2.3. *Assume A1, A4-5 and B2. Then for any measurable function $f : \mathsf{X} \to \mathbb{R}$ in $\mathcal{L}_{V^\beta}$ and any initial distribution ξ_1, ξ_2 resp. on X, Θ such that $\xi_1(V) < +\infty$,*

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

As in the case of the convergence of the marginals, when A5 and B2 hold with $\mathcal{D} = \mathsf{X}$ and $\beta = 0$, A1 and A4 can be omitted. We thus have

Proposition 2.4. *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, ξ_2 resp. on X, Θ*

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

2.4. **Discussion.**

2.4.1. *Non-adaptive case.* We start by comparing our assumptions to assumptions in Markov chain theory under which the ergodicity and 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, A1-A2 are equivalent to the condition that $P_{\theta_\star}$ is an aperiodic positive Harris recurrent transition kernel ((Meyn and Tweedie, 1993, Theorem 11.0.1)), a condition which in turn is known to be minimal and to imply ergodicity in total variation norm ((Meyn and Tweedie, 1993, Theorem 13.0.1)). B1 is trivially true. Finally, A3 is stronger than what we want to prove (since A3 implies the conclusion of Theorem 2.1 in the non-adaptive case); this is indeed due to our technique of proof 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.

For the strong law of large numbers (Theorem 2.3), B2 is still trivially true in the Markovian case and A5 is implied by A1 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, A1 and A4 are known sufficient conditions for a strong law of large numbers for $f \in \mathcal{L}_{V^{1-\alpha}}$, which is a bit stronger than the conclusions of Theorem 2.3. This slight loss of efficiency is due to the technique of proof based on martingale theory (see comments Section 2.4.5). Observe that in the geometric case, there is the same loss of generality in (Andrieu and Moulines, 2006, Theorem 8). More generally, any proof of the law of large numbers based on the martingale theory (through for example the use of the Poisson's equation or of the resolvent kernel) will incur the same loss of efficiency since limit theorems exist only for L^p-martingale with $p > 1$.

2.4.2. *Checking assumptions A3 and A5.* A3 and A5 are the most technical of our assumptions. Contrary to the case of a single kernel, the relations between A3 (resp. A5) and A1-2 (resp. A1, A4) are not completely well understood. Nevertheless these assumptions can be checked under conditions which are essentially of the form A1, A4 plus the assumptions that each transition kernel P_θ is phi-irreducible and aperiodic, as discussed in Appendix A.

Lastly, it is a recognized fact that A4 is equivalent to $\pi(V^{1-\alpha}) < +\infty$ (although, to the best of our knowledge, there exist no rigorous proof in the literature). In that sense, A2-A3 is weaker than the set of conditions "A2 and $\pi(V) < +\infty$".

2.4.3. *On the uniformity in θ in assumptions A1, A2 and A4.* We have formulated A1, A2 and A4 such that all the constants involved are independent of θ, for $\theta \in \Theta$. Intuitively, this corresponds to AMCMC algorithms based on kernels with overall similar ergodicity properties. This uniformity assumption might seem unrealistically strong at first. But the next example shows that when A1, A2 and A4 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. The 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 $\{X_n, n \geq 0\}$ 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, A2 and A4 hold uniformly for all $\theta \in \mathsf{K}$. But these assumption do not hold uniformly for all $\theta \in (0, 1)$. Therefore Theorems 2.1 and 2.3 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 $\mathbb{E}_{x,\theta_0}\left[\left(n^{-1} \sum_{k=1}^n f(X_k) - \pi(f)\right)^2\right]$ do 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, A2 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 in 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 out of the scope of this paper; the use of the Chen's truncation method to weaken our assumption is addressed in Atchade 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

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

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 (2008) gives a theoretical discussion on the efficiency of AMCMC under (2).

Roberts and Rosenthal (2007) improves on the literature by relaxing the convergence rate assumption on the kernels. They prove the convergence of the marginal and a weak law of large numbers for bounded functions. But 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 with (2)) but is very difficult to prove in the polynomial case. We show here that this uniform control is unnecessary.

Yang (2007) tackles some open questions mentioned in Roberts and Rosenthal (2007), by providing sufficient conditions - close to the conditions we give in Theorems 2.1 and 2.3 - 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 (our conditions are weaker than $\pi(V) < +\infty$, see Section 2.4.2) and not easily checked (the proof that $\{V(X_n), n \geq 0\}$ is bounded in probability is not trivial at all in the sub-geometric case). We also observed some skips 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 $P_{\theta_n}^N(X_n, \cdot)$ to be within ϵ to 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 to $P_{\theta_n}^N(X_n, \cdot)$. Combining these two arguments, we can then conclude that the distribution of X_{n+N} is within 2ϵ to π. This is essentially the argument of Roberts and Rosenthal (2007). The difficulty with this argument is that the distance between $P_{\theta_n}^N(x, \cdot)$ and π depends in general on x and can rarely be bounded uniformly in x. We solve this problem here by introducing some very large level set of V - $\mathcal{D} = \{x \in \mathsf{X} : \; V(x) \leq d\}$ - and by using two basic facts: *(i)* under A2, the process cannot wait too long before coming back in $\mathcal{D}$; *(ii)* under A3, a bound on the distance between $P_{\theta_n}^N(x, \cdot)$ and π uniformly in x, for $x \in \mathcal{D}$, is possible.

The proof of Theorem 2.3 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's equation technique which has been used to study adaptive MCMC in Andrieu and Moulines (2006). Under A1 and A4, a solution g_θ to the Poisson's equation with transition kernel P_θ exists for any $f \in \mathcal{L}_{V^\beta}$, $0 \leq \beta \leq 1 - \alpha$ and $g_\theta \in \mathcal{L}_{V^{\beta+\alpha}}$. But in order to use $\{g_\theta, \; \theta \in \Theta\}$ to obtain a 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

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

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 $\{1,\cdots,K\}$, $K\geq 4$. Let $\{P_\theta, \theta\in\{1,\cdots,M\}\}$ be a family of Random Walk Metropolis algorithm with proposal distribution q_θ, the uniform distribution on $\{x-\theta,\cdots,x-1,x+1,\cdots,x+\theta\}$.

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 $\{\bar{P}(n;(x,\theta),(dx',d\theta'), n\geq 0\}$ where

$$\bar{P}(n;(x,\theta),(dx',d\theta') = P_\theta(x,dx')\ \Big(\mathbb{1}_{x=x'}\left\{p_{n+1}\ \delta_{1\vee(\theta-1)}(d\theta')+(1-p_{n+1})\ \delta_\theta(d\theta')\right\} \\ +\mathbb{1}_{x\neq x'}\left\{p_{n+1}\ \delta_{M\wedge(\theta+1)}(d\theta')+(1-p_{n+1})\ \delta_\theta(d\theta')\right\}\Big) \ .$$

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

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

Since Θ is finite, this implies that A3 (resp. A5) hold with $\mathcal{D}=\mathsf{X}$ (resp. $\mathcal{D}=\mathsf{X}$ and $\beta=0$). Furthermore, $\mathbb{E}^{(l)}_{x,\theta}\left[D(\theta_n,\theta_{n+1})\right]\leq 2p_{n+1}$ so that B1 (resp. B2) hold with any probability measures ξ_1,ξ_2 (resp. with $\mathcal{D}=\mathsf{X}$) provided $p_n\to 0$. By Propositions 2.2 and 2.4, we have

Proposition 3.1. *Assume* $\lim_n p_n = 0$. *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

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

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 Hastings (1970); Metropolis et al. (1953). Let a target density π, absolutely continuous w.r.t. the Lebesgue measure μ_{Leb} with density still denoted by π. Choose a proposal distribution with density w.r.t. 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 $q(\cdot)$. Then we either 'accept' Y and set $X_{n+1} = Y$ with probability $\alpha(x, Y) \stackrel{\text{def}}{=} \min\left(1, \pi(Y)/\pi(x)\right)$ 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 super-exponential in the tails with regular contours, then the RWM kernel is typically geometrically ergodic (Jarner and Hansen (2000)). Otherwise, it is typically sub-geometric (Fort and Moulines (2000, 2003); Douc et al. (2004)).

Define

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

resp. the expectation and the covariance matrix of π ($\cdot^T$ denotes the transpose operation). Theoretical results suggest setting the variance-covariance matrix Σ of the proposal distribution $\Sigma = c_\star \Sigma_\star$ 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 e.g. Roberts and Rosenthal (2001) for more details. Haario et al. (2001) have proposed an adaptive algorithm to learn Σ_* adaptively during the simulation. This algorithm has been studied in detail in Andrieu and Moulines (2006) under the assumption that π is super-exponential in the tails. An adaptive algorithm to find the optimal value $c_\star$ has been proposed in Atchade and Rosenthal (2005) (see also Atchade (2006)) and studied under the assumption that π is super-exponential in the tails. We extend these results to cases where π is sub-exponential in the tails.

For $0 < \Lambda_l < \Lambda_u$, let $\Theta_+(\Lambda_l, \Lambda_u)$ be the space of $p \times p$ symmetric positive definite matrices A such that $\Lambda_l \leq |A|_{\mathrm{s}} \leq \Lambda_u$, where $|A|_{\mathrm{s}}$ denotes the spectral norm $|A|_{\mathrm{s}} \stackrel{\text{def}}{=} \sqrt{\lambda_m(AA^T)}$, and for a symmetric square matrix $\lambda_m(M)$ denotes the largest eigenvalue of M. Next, for $-\infty < \kappa_l < \kappa_u < \infty$ and Θ_μ a compact subset of X, we introduce the space $\Theta \stackrel{\text{def}}{=} \Theta_\mu \times \Theta_+(\Lambda_l, \Lambda_u) \times [\kappa_l, \kappa_u]$. 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 $e^c\Sigma$.

Consider the adaptive RWM 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 $Z_{n+1} \sim q_{\theta_n} d\mu_{Leb}$ 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*

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

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

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

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 $\{(X_n, \theta_n), n \geq 0\}$ which is a non-homogeneous Markov chain with successive transitions kernels $\{\bar{P}(n; (x, \theta), (dx', d\theta')), n \geq 0\}$ given by

$$\bar{P}(n; (x, \theta), (dx', d\theta')) = \int q_\theta(z) \ \left\{\alpha(x, x+z)\delta_{x+z}(dx') + (1 - \alpha(x, x+z))\delta_x(dx')\right\} \cdots \\ \left(\mathbb{1}_{\{\phi(\theta, x+z, x') \in \Theta\}}\delta_{\phi(\theta, x+z, x')}(d\theta') + \mathbb{1}_{\{\phi(\theta, x+z, x') \notin \Theta\}}\delta_\theta(d\theta')\right) \ d\mu_{Leb}(dz)$$

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

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

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 $\mu_{Leb}(\mathcal{C}) > 0$ satisfies A1.*

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

Following (Fort and Moulines (2000)), we assume that π is sub-exponential 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 $|x| \geq R$:

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

Examples of target density that satisfies D1-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 $|x| \to +\infty$,

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

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

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

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

Lemma 3.3. *Assume D1-3. For $0 < s \leq s_\star$, define $V_s(x) \stackrel{\text{def}}{=} 1 + \pi^{-s}(x)$. There exist $0 < s \leq s_\star$ 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 A2-5 hold.

Lemma 3.4. *Assume D1-3. B2 holds and B1 holds for any probability measures ξ_1,ξ_2 such that $\int |\ln \pi|^{2/m} d\xi_1 < +\infty$.*

The proof of Lemmas 3.3 and 3.4 are in Appendix C.

Proposition 3.5. *Assume D1-3. Consider the sequence $\{X_n, n \geq 0\}$ given by the algorithm 3.1.*

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

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

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

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

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 sub-exponential moments (see Proposition 3.5(ii)), which always holds with $\xi_1 = \delta_x$, $x \in \mathsf{X}$.

3.2.2. *Law of large numbers for polynomially increasing functions.* Proposition 3.5 also addresses the case when 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 the Jensen's inequality to the drift inequality (18) with the concave function $\phi(t) \sim [\ln t]^{(r+\iota+2)/m-1}$; see (Jarner and Roberts, 2002, Lemma 3.5) for similar calculations). Hence, the condition on ξ_1 gets into $\xi_1(|x|^{r+\iota+2-m}) < +\infty$ for some $\iota > 0$.

Drift inequalities with $V \sim (-\ln \pi)^s$ for some $s > 2/m - 1$, can also be derived by direct computations: 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 return-time to $\mathcal{C} \times \Theta$ of $\{Z_n, n \geq 0\}$ and its hitting-time on $\mathcal{C} \times \Theta$ by

$$\tau_{\mathcal{C}} \stackrel{\text{def}}{=} \inf\{n \geq 1, Z_n \in \mathcal{C} \times \Theta\} , \qquad \sigma_{\mathcal{C}} \stackrel{\text{def}}{=} \inf\{n \geq 0, Z_n \in \mathcal{C} \times \Theta\} .$$

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 gives 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 $(x,\theta) \in \mathsf{X} \times \Theta$,*

$$\Delta \stackrel{\text{def}}{=} \left| \mathbb{E}^{(l)}_{x,\theta}\left[f(X_1, \cdots, X_N)\right] - \int_{\mathsf{X}^N} P_\theta(x, dx_1) \prod_{k=2}^{N} P_\theta(x_{k-1}, dx_k) f(x_1, \cdots, x_n) \right|$$
$$\leq |f|_1 \sum_{j=1}^{N-1} \sum_{i=1}^{j} \mathbb{E}^{(l)}_{x,\theta}\left[D(\theta_i, \theta_{i-1})\right] .$$

Proof. We can assume w.l.g. that $|f|_1 \leq 1$. 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 $\int_{\mathsf{X}} P_\theta(x, dx') f(x') = \int_{\mathsf{X}\times\Theta} \bar{P}_l(0; (x,\theta), (dx', d\theta')) f(x')$,

$$\Delta = \Bigg| \int_{(\mathsf{X}\times\Theta)^N} \sum_{j=1}^{N-1} \bar{P}_l(0; (x,\theta), dz_1) \prod_{k=2}^{j} \bar{P}_l(k-1; z_{k-1}, dz_k) \cdots$$
$$\left\{\bar{P}_l(j; z_j, dz_{j+1}) - \bar{P}_l(0; (x_j,\theta), dz_{j+1})\right\} \prod_{k=j+2}^{N} \bar{P}_l(0; (x_{k-1},\theta), dz_k) f(x_1, \cdots, x_N) \Bigg|$$
$$\leq \sum_{j=1}^{N-1} \int_{\mathsf{X}^j} \bar{P}_l(0; (x,\theta), dz_1) \prod_{k=2}^{j} \bar{P}_l(k-1; z_{k-1}, dz_k) \sup_{x\in\mathsf{X}} \|P_{t_j}(x,\cdot) - P_\theta(x,\cdot)\|_{\mathrm{TV}}$$

where we used that

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

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

$$\int_{\mathsf{X}\times\Theta} \left\{\bar{P}_l(j; z_j, dz_{j+1}) - \bar{P}_l(0; (x_j,\theta), dz_{j+1})\right\} \Xi(x_1, \cdots, x_{j+1})$$
$$= \int_{\mathsf{X}} \left\{P_{t_j}(x_j, dx_{j+1}) - P_\theta(x_j, dx_{j+1})\right\} \Xi(x_1, \cdots, x_{j+1}) \leq \sup_{x\in\mathsf{X}} \|P_{t_j}(x,\cdot) - P_\theta(x,\cdot)\|_{\mathrm{TV}} \, |\Xi|_1 .$$

Hence

$$\Delta \leq \sum_{j=1}^{N-1} \mathbb{E}^{(l)}_{x,\theta}\left[\sup_{x\in\mathsf{X}} \|P_{\theta_j}(x,\cdot) - P_{\theta_0}(x,\cdot)\|_{\mathrm{TV}}\right]$$
$$\leq \sum_{j=1}^{N-1} \mathbb{E}^{(l)}_{x,\theta}\left[\sum_{i=1}^{j} \sup_{x\in\mathsf{X}} \|P_{\theta_i}(x,\cdot) - P_{\theta_{i-1}}(x,\cdot)\|_{\mathrm{TV}}\right] = \sum_{j=1}^{N-1} \sum_{i=1}^{j} \mathbb{E}^{(l)}_{x,\theta}\left[D(\theta_i, \theta_{i-1})\right] .$$

□

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 e.g. in (Roberts and Rosenthal, 2004, Proposition 3). As a consequence of Lemmas 4.1 and 4.2, we have

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*

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

$(X_0, \cdots, X_N)$ has the X-marginal distribution of $\mathbb{P}^{(l)}_z$ restricted to the time-interval $\{0, \cdots, N\}$, and $(\tilde{X}_0, \cdots, \tilde{X}_N)$ 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 $\theta \in \Theta$,

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

Under A2, (6) holds with $\alpha = 1$ and $c = 1$; under A4, (6) holds with $\alpha < 1$. We establish a set of results that hold under (6) and thus, under A2 or A4.

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 (Jarner and Roberts, 2002, Lemma 3.5). □

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

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

Proof. The proof can be adapted from (Douc et al., 2004, Proposition 2.1) and (Meyn and Tweedie, 1993, Proposition 11.3.2)and is 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 $(x, \theta) \in \mathsf{X} \times \Theta$

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

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

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

(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 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,

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

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

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

The results 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

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

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 $q \in [1, +\infty]$. $\square$

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 $n \geq 0$,*

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

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

4.1.3. *Delayed successive visits to an accessible level set of V.* Let $\mathcal{D} \in \mathcal{X}$ and two positive integers $n_\star, N$. Define on $(\Omega, \mathcal{F}, \mathbb{P}_{x,\theta}^{(l)})$ the sequence of $\mathbb{N}$-valued random variables $\{\tau^n, n \geq 1\}$ as

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

Proposition 4.8. *Assume A1 and A2. Let $\mathcal{D} \in \mathcal{X}$. Let $n_\star, N$ be two positive 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* $\nu(\mathcal{D}) > 0$*, there exists a (finite) constant C depending upon* $\varepsilon, \nu(\mathcal{D}), \sup_{\mathcal{D}} V, b, n_\star, N$ *such that for any $l \geq 0$, $(x, \theta) \in \mathsf{X} \times \Theta$ and $k \geq 0$,*

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

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

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

By A1, we have $P_\theta(x, \mathcal{D}) \geq [\varepsilon \nu(\mathcal{D})] \ \mathbb{1}_{\mathcal{C}}(x)$ for any (x, θ) so that

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

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

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

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

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

□

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*

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

Proof. By Proposition 4.6(i), $\left|\mathbb{E}^{(l)}_{x,\theta}\left[\bar{f}(X_j)\right]\right| \leq |\bar{f}|_{V^\beta}\ \left(V^\beta(x) + \bar{b}j^\beta\right)$. 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 definition of $\hat{g}^{(l)}_a$ and by the Markov property,

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

□

Theorem 4.10. *Assume A1, A4-5 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 $q \in [1,+\infty]$,*

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

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

Proof. 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

$$\frac{1}{N} \sum_{j=0}^{N-1} \sup_{\mathcal{D}\times\Theta} \|P_\theta^j(x,\cdot) - \pi(\cdot)\|_{V^\beta} \leq \epsilon \;, \tag{7}$$

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

$$(\alpha c)^{-1}\; N^{-1} \left(\sup_{\mathcal{D}} V^{\beta+\alpha} + \bar{b} N^{\beta+\alpha} + \bar{b}[\varepsilon\nu(\mathcal{D})]^{-1} \right) \leq \epsilon \;. \tag{8}$$

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

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

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

$$\mathbb{E}_{x,\theta}^{(l)}\left[\overline{\mathbb{P}}_{Z_{\tau^k},Z_{\tau^k}}^{(\tau^k+l)}\left(\mathcal{E}_N^c\right)\right] \leq \sum_{j=1}^{N-1}\sum_{i=1}^{j} \mathbb{E}_{x,\theta}^{(l)}\left[\mathbb{E}_{Z_{\tau^k}}^{(\tau^k+l)}\left[D(\theta_i,\theta_{i-1})\right]\right] \leq N^2\epsilon_N < 1 \;. \tag{9}$$

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 $0 < a < 1$, $l \geq 0$, $(x,\theta) \in \mathsf{X}\times\Theta$ and $k \geq 2$,

$$\left|\mathbb{E}_{x,\theta}^{(l)}\left[\sum_{j=0}^{N-1}(1-a)^{\tau^k+j+1}\;\bar{f}\left(X_{\tau^k+j}\right)\right]\right| \leq |\bar{f}|_{V^\beta}\; 3N\epsilon\; (1-a)^{n_\star+(k-1)N} \;. \tag{10}$$

We have

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

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

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

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

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

For $k \geq 1$,

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

By Proposition 4.6(i) and (ii) applied with $\tau = \tau_{\mathcal{D}}$, Proposition 4.8 and Eq. (10), and upon noting that $\tau^k \geq n_\star + (k-1)N \ \mathbb{P}_{(x,\theta)}^{(l)}$-a.s. ,

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

where we used the definition of N (see Eq. (8)) and Proposition 4.6(i). This yields the desired result.

`Proof of Eq.(10).` By the strong Markov property and since $\tau^k \geq n_\star + N(k-1) \ \mathbb{P}_{x,\theta}^{(l)}$-a.s.

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

Furthermore, by Proposition 4.3,

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

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

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

by (7). On the other hand, $\mathbb{P}^{(l)}_{x,\theta}-$ a.s.,

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

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

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

Finally,

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

where we used (9) in the last inequality. To conclude the proof, 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 $n \geq n_\epsilon$, $\sup_{\{f,|f|_1 \leq 1\}} \left|\mathbb{E}_{\xi_1,\xi_2}\left[\bar{f}(X_n)\right]\right| \leq \epsilon$. To that goal, let us introduce the sequence of random times $\{\tau^k, k \geq 0\}$ as defined in Section 4.1.3 for $(\mathcal{D}, N, n_\star)$ defined as follows.

4.2.1. *Definition of $\mathcal{D}$, N, Q and $n_\star$.* There exists a level set $\mathcal{D}$ of V large enough such that $\nu(\mathcal{D}) > 0$ and $\pi(\mathcal{D}) \geq 1 - \epsilon$.

By A3, choose $N \geq 2$ such that $\sup_{\mathcal{D}\times\Theta} \|P_\theta^N(x,\cdot) - \pi(\cdot)\|_{\mathrm{TV}} \leq \epsilon$. This implies that

$$\sup_{n \geq N} \sup_{\mathcal{D}\times\Theta} P_\theta^n(x, \mathcal{D}^c) \leq 2\epsilon \; . \tag{11}$$

Choose $Q \geq N$ such that

$$\sup_{l \geq 0} \sup_{\mathcal{D}\times\Theta} \; \mathbb{P}_{x,\theta}^{(l)}\left(\tau_{\mathcal{D}} \circ \underline{\theta}^N \geq Q\right) \leq \epsilon/N \; , \tag{12}$$

the existence of which is given by the Markov inequality, the Markov property, Propositions 4.6(i) and 4.8. Finally, by B1, choose $n_\star$ such that for any $n \geq n_\star$,

$$\mathbb{P}_{\xi_1,\xi_2}\left(D(\theta_n, \theta_{n-1}) \geq \epsilon/(2(2N+Q)^2(Q+N))\right) \leq \frac{\epsilon}{4(2N+Q)^2(Q+N)} \; . \tag{13}$$

4.2.2. *Optimal coupling.* Let us define the sequence of variables $\{d_n, n \geq 0\}$ on $(\Omega, \mathcal{F})$ as follows : $d_n = -1$ for $0 \leq n < \tau^1$ and

$$d_{\tau^k} = 0 \qquad k \geq 1 \; , \qquad\qquad d_{n+1} = d_n + 1 \;\; \text{otherwise} \; .$$

We apply Proposition 4.3 with $l = 0$ and $N \leftarrow 2N + Q$. Set $\mathcal{E}_{2N+Q} \stackrel{\text{def}}{=} \{X_k = \tilde{X}_k, 0 \leq k < 2N+Q\}$. By definition of $n_\star$ (see Eq. (13)) and of the random variables $\{d_n, n \geq 0\}$, it holds for any $r \geq n_\star$,

$$\begin{aligned}
\mathbb{E}_{\xi_1,\xi_2}\left[\mathbb{1}_{d_r=0}\, \overline{\mathbb{P}}_{Z_r,Z_r}^{(r)}\left(\mathcal{E}_{2N+Q}^c\right)\right] &\leq \sum_{j=1}^{2N+Q} \sum_{i=1}^{j} \mathbb{E}_{\xi_1,\xi_2}\left[\mathbb{1}_{d_r=0}\; \mathbb{E}_{Z_r}^{(r)}\left[D(\theta_i, \theta_{i-1})\right]\right] \\
&\leq \sum_{j=1}^{2N+Q} \sum_{i=1}^{j} \mathbb{E}_{\xi_1,\xi_2}\left[D(\theta_{i+r}, \theta_{i+r-1})\right] \leq \epsilon (Q+N)^{-1} \; .
\end{aligned} \tag{14}$$

4.2.3. *Proof.* We write

$$\mathbb{E}_{\xi_1,\xi_2}\left[\bar{f}(X_n)\right] = \mathbb{E}_{\xi_1,\xi_2}\left[\bar{f}(X_n)\left\{\mathbb{1}_{n \leq \tau^2} + \mathbb{1}_{n > \tau^2}\left(\mathbb{1}_{0 \leq d_n \leq N-1} + \mathbb{1}_{N \leq d_n \leq N+Q} + \mathbb{1}_{d_n > N+Q}\right)\right\}\right] \; ,$$

and consider each term in turn. We use the notation $\{X_{n:m} \notin \mathcal{D}\}$ as a shorthand notation for $\bigcap_{l=n}^m \{X_l \notin \mathcal{D}\}$.

`Convergence of Term 1.` By Proposition 4.8, there exists a constant C such that $\mathbb{E}_{\xi_1,\xi_2}\left[\tau^2\right] \leq 2C + V(x)$ which implies

$$\lim_n \left|\mathbb{E}_{\xi_1,\xi_2}\left[\bar{f}(X_n)\mathbb{1}_{n \leq \tau^2}\right]\right| \leq |\bar{f}|_1 \, n^{-1}\, \mathbb{E}_{\xi_1,\xi_2}\left[\tau^2\right] = 0 \; .$$

`Convergence of Term 2.` Let $n \geq 2N+Q+n_\star$. On the set $\{d_n = q\}$, there exists $k \geq 2$ such that $\tau^k = d_{n-q}$. We decompose the event $\{d_n = q\}$ based on the length between τ^{k-1} and τ^k. We write

$$
\begin{aligned}
|\bar{f}|_1^{-1} \left|\mathbb{E}_{\xi_1,\xi_2}\left[\bar{f}(X_n)\mathbb{1}_{0\leq d_n<N, n>\tau^2}\right]\right| \leq \left|\sum_{q=0}^{N-1} \sum_{l \leq n-q-(N+Q)} \mathbb{P}_{\xi_1,\xi_2}\left(d_n = q, d_l = 0, X_{l+N+1:n-q-1} \notin \mathcal{D}\right)\right| \\
+ \left|\sum_{q=0}^{N-1} \sum_{l=n-q-(N+Q)+1}^{n-q-N-2} \mathbb{P}_{\xi_1,\xi_2}\left(d_n = q, d_l = 0, X_{l+N+1:n-q-1} \notin \mathcal{D}\right)\right| \\
+ |\bar{f}|_1^{-1} \sum_{q=0}^{N-1} \left|\mathbb{E}_{\xi_1,\xi_2}\left[\bar{f}(X_n)\mathbb{1}_{d_n=q}\mathbb{1}_{d_{n-q-(N+1)}=0}\right]\right| \; .
\end{aligned}
$$

For the first term, we have

$$
\begin{aligned}
&\sum_{q=0}^{N-1} \sum_{l \leq n-q-(N+Q)} \mathbb{P}_{\xi_1,\xi_2}\left(d_n = q, d_l = 0, X_{l+N+1:n-q-1} \notin \mathcal{D}\right) \\
&\leq \sum_{q=0}^{N-1} \sum_{l \leq n-q-(N+Q)} \mathbb{E}_{\xi_1,\xi_2}\left[\mathbb{1}_{d_l=0}\, \mathbb{P}^{(l)}_{Z_l}\left(d_{n-l} = q, X_{N+1:n-q-1-l} \notin \mathcal{D}\right)\right] \\
&\leq \sum_{q=0}^{N-1} \sum_{l \leq n-q-(N+Q)} \mathbb{E}_{\xi_1,\xi_2}\left[\mathbb{1}_{d_l=0}\, \mathbb{P}^{(l)}_{Z_l}\left(d_{n-l} = q, X_{N+1:N+Q-1} \notin \mathcal{D}\right)\right] \\
&\leq \sum_{q=0}^{N-1} \sum_{l \leq n-q-(N+Q)} \mathbb{E}_{\xi_1,\xi_2}\left[\mathbb{1}_{d_l=0}\, \mathbb{P}^{(l)}_{Z_l}\left(\tau_{\mathcal{D}} \circ \underline{\theta}^N \geq Q\right)\right] \\
&\leq \sum_{q=0}^{N-1} \mathbb{P}_{\xi_1,\xi_2}\left(\exists j \geq 1, \tau^j \leq n-q-(N+Q)\right) \sup_{l\geq 0, \mathcal{D}\times\Theta} \mathbb{P}^{(l)}_{z}\left(\tau_{\mathcal{D}} \circ \underline{\theta}^N \geq Q\right) \leq \epsilon \; ,
\end{aligned}
$$

by definition of Q (see Eq. (12)). For the second term, we have

$$
\begin{aligned}
&\sum_{q=0}^{N-1} \sum_{l=n-q-(N+Q)}^{n-q-N-2} \mathbb{P}_{\xi_1,\xi_2}\left(d_n = q, d_l = 0, X_{l+N+1:n-q-1} \notin \mathcal{D}\right) \\
&\qquad \le \sum_{l=n-2N-Q+1}^{n-N-Q-1} \sum_{q=n-N-Q-l}^{N-1} \mathbb{P}_{\xi_1,\xi_2}\left(d_l = 0, d_n = q, X_{l+N+1:n-q-1} \notin \mathcal{D}\right) \\
&\qquad\quad + \sum_{l=n-N-Q}^{n-2N-1} \sum_{q=0}^{N-1} \mathbb{P}_{\xi_1,\xi_2}\left(d_l = 0, d_n = q, X_{l+N+1:n-q-1} \notin \mathcal{D}\right) \\
&\qquad\quad + \sum_{l=n-2N}^{n-N-2} \sum_{q=0}^{n-N-l-2} \mathbb{P}_{\xi_1,\xi_2}\left(d_l = 0, d_n = q, X_{l+N+1:n-q-1} \notin \mathcal{D}\right) \\
&\le \sum_{l=n-2N-Q+1}^{n-N-2} \mathbb{P}_{\xi_1,\xi_2}\left(d_l = 0, X_{l+N+1} \notin \mathcal{D}\right) = \sum_{l=n-2N-Q+1}^{n-N-2} \mathbb{E}_{\xi_1,\xi_2}\left[\mathbb{1}_{d_l=0}\, \mathbb{P}^{(l)}_{Z_l}\left(X_{N+1} \notin \mathcal{D}\right)\right] \\
&\le \sum_{l=n-2N-Q+1}^{n-N-2} \left\{\mathbb{E}_{\xi_1,\xi_2}\left[\mathbb{1}_{d_l=0}\, \overline{\mathbb{P}}^{(l)}_{Z_l,Z_l}\left(\tilde{X}_{N+1} \notin \mathcal{D}\right)\right] + \mathbb{E}_{\xi_1,\xi_2}\left[\mathbb{1}_{d_l=0}\, \overline{\mathbb{P}}^{(l)}_{Z_l,Z_l}\left(\mathcal{E}^c_{2N+Q}\right)\right]\right\} \\
&\qquad \le \sup_{\mathcal{D}\times\Theta} P^N_\theta(x, \mathcal{D}^c) + (N+Q) \sup_{r \ge n_\star} \mathbb{E}_{\xi_1,\xi_2}\left[\mathbb{1}_{d_r=0}\, \overline{\mathbb{P}}^{(r)}_{Z_r,Z_r}\left(\mathcal{E}^c_{2N+Q}\right)\right] \le 2\epsilon + \epsilon ,
\end{aligned}
$$

by Eqs. (11) and (14). For the third term, we have for any $0 \le q \le N-1$,

$$
\begin{aligned}
&\left|\mathbb{E}_{\xi_1,\xi_2}\left[\bar{f}(X_n)\mathbb{1}_{d_n=q}\mathbb{1}_{d_{n-q-(N+1)}=0}\right]\right| = \left|\mathbb{E}_{\xi_1,\xi_2}\left[\mathbb{1}_{d_{n-q-(N+1)}=0}\mathbb{1}_{X_{n-q}\in\mathcal{D}}\bar{f}(X_n)\right]\right| \\
&\qquad \le \mathbb{E}_{\xi_1,\xi_2}\left[\mathbb{1}_{d_{n-q-(N+1)}=0}\left|\mathbb{E}^{(n-q-(N+1))}_{Z_{n-q-(N+1)}}\left[\mathbb{1}_{X_{N+1}\in\mathcal{D}}\bar{f}(X_{q+N+1})\right]\right|\right] \\
&\qquad \le \mathbb{E}_{\xi_1,\xi_2}\left[\mathbb{1}_{d_{n-q-(N+1)}=0}\left|\overline{\mathbb{E}}^{(n-q-(N+1))}_{Z_{n-q-(N+1)},Z_{n-q-(N+1)}}\left[\bar{f}(\tilde{X}_{q+N+1})\right]\right|\right] \\
&\qquad\quad + |\bar{f}|_1\, \mathbb{E}_{\xi_1,\xi_2}\left[\mathbb{1}_{d_{n-q-(N+1)}=0}\, \overline{\mathbb{P}}^{(n-q-(N+1))}_{Z_{n-q-(N+1)},Z_{n-q-(N+1)}}\left(\tilde{X}_{N+1} \in \mathcal{D}^c\right)\right] \\
&\qquad\qquad + |\bar{f}|_1 \sup_{r \ge n_\star} \mathbb{E}_{\xi_1,\xi_2}\left[\mathbb{1}_{d_r=0}\, \overline{\mathbb{P}}^{(r)}_{Z_r,Z_r}\left(\mathcal{E}^c_{2N+Q}\right)\right] \\
&\le \mathbb{P}_{\xi_1,\xi_2}\left(d_{n-q-(N+1)} = 0\right)\ |\bar{f}|_1 \left\{\sup_{\mathcal{D}\times\Theta} \|P^N_\theta(x,\cdot) - \pi(\cdot)\|_{\mathrm{TV}} + \sup_{\mathcal{D}\times\Theta} P^{N+1}_\theta(x, \mathcal{D}^c)\right\} + |\bar{f}|_1\, \epsilon (N+Q)^{-1} ,
\end{aligned}
$$

by Eq. (14). The rhs is lower than $5\epsilon|\bar{f}|_1$ by definition of N and Eq. (11).

`Convergence of Term 3.` Let $n \ge N + Q + n_\star$. We write

$$
\begin{aligned}
\mathbb{E}_{\xi_1,\xi_2}\left[\bar{f}(X_n)\mathbb{1}_{N\le d_n \le N+Q, n>\tau^2}\right] &= \sum_{l=N}^{N+Q} \mathbb{E}_{\xi_1,\xi_2}\left[\bar{f}(X_n)\mathbb{1}_{d_n=l,n>\tau^2}\right] \\
&= \mathbb{E}_{\xi_1,\xi_2}\left[\bar{f}(X_n)\mathbb{1}_{d_{n-N}=0,n>\tau^2}\right] + \sum_{l=N+1}^{N+Q} \mathbb{E}_{\xi_1,\xi_2}\left[\bar{f}(X_n)\mathbb{1}_{d_{n-l}=0,n>\tau^2}\mathbb{1}_{X_{n-l+N+1:n}\notin\mathcal{D}}\right] .
\end{aligned}
$$

We have

$$
\begin{aligned}
&\left|\mathbb{E}_{\xi_1,\xi_2}\left[\bar{f}(X_n)\mathbb{1}_{d_{n-N}=0,n>\tau^2}\right]\right| \le \mathbb{E}_{\xi_1,\xi_2}\left[\mathbb{1}_{d_{n-N}=0,n-N\ge\tau^2}\left|\overline{\mathbb{E}}^{(n-N)}_{Z_{n-N},Z_{n-N}}\left[\bar{f}(X_N)\right]\right|\right] \\
&\quad\le \mathbb{E}_{\xi_1,\xi_2}\left[\mathbb{1}_{d_{n-N}=0,n-N\ge\tau^2}\left|\overline{\mathbb{E}}^{(n-N)}_{Z_{n-N},Z_{n-N}}\left[\bar{f}(\tilde{X}_N)\right] + 2\,|\bar{f}|_1\,\overline{\mathbb{P}}^{(n-N)}_{Z_{n-N},Z_{n-N}}\left(\mathcal{E}^c_{2N+Q}\right)\right|\right] \\
&\le |\bar{f}|_1 \sup_{\mathcal{D}\times\Theta}\|P^N_\theta(x,\cdot)-\pi(\cdot)\|_{\mathrm{TV}} + 2|\bar{f}|_1 \sup_{r\ge n_\star}\mathbb{E}_{\xi_1,\xi_2}\left[\mathbb{1}_{d_r=0}\,\overline{\mathbb{P}}^{(r)}_{Z_r,Z_r}\left(\mathcal{E}^c_{2N+Q}\right)\right] \le |\bar{f}|_1\, 2\epsilon\,,
\end{aligned}
$$

by definition of N and Eq. (14). Furthermore, for any $N+1\le l\le N+Q$,

$$
\begin{aligned}
&|\bar{f}|_1^{-1}\left|\mathbb{E}_{\xi_1,\xi_2}\left[\bar{f}(X_n)\mathbb{1}_{d_{n-l}=0,n>\tau^2}\mathbb{1}_{X_{n-l+N+1:n}\notin\mathcal{D}}\right]\right| \le \mathbb{E}_{\xi_1,\xi_2}\left[\mathbb{1}_{d_{n-l}=0}\mathbb{1}_{X_{n-l+N+1:n}\notin\mathcal{D}}\right] \\
&\quad\le \mathbb{E}_{\xi_1,\xi_2}\left[\mathbb{1}_{d_{n-l}=0}\,\mathbb{E}^{(n-l)}_{Z_{n-l}}\left[\mathbb{1}_{X_{N+1}\notin\mathcal{D}}\right]\right] \le \mathbb{E}_{\xi_1,\xi_2}\left[\mathbb{1}_{d_{n-l}=0}\,\overline{\mathbb{P}}^{(n-l)}_{Z_{n-l},Z_{n-l}}\left(\tilde{X}_{N+1}\notin\mathcal{D}\right)\right] \\
&\qquad+\mathbb{E}_{\xi_1,\xi_2}\left[\mathbb{1}_{d_{n-l}=0}\,\bar{\mathbb{E}}^{(n-l)}_{Z_{n-l},Z_{n-l}}\left[\left|\mathbb{1}_{X_{N+1}\notin\mathcal{D}}-\mathbb{1}_{\tilde{X}_{N+1}\notin\mathcal{D}}\right|\mathbb{1}_{\mathcal{E}^c_{2N+Q}}\right]\right] \\
&\qquad\quad\le \sup_{\mathcal{D}\times\Theta}P^{N+1}_\theta(x,\mathcal{D}^c)\;\mathbb{P}_{\xi_1,\xi_2}\left(d_{n-l}=0\right)+\sup_{r\ge n_\star}\mathbb{E}_{\xi_1,\xi_2}\left[\mathbb{1}_{d_r=0}\,\bar{\mathbb{P}}^{(r)}_{Z_r,Z_r}\left(\mathcal{E}^c_{2N+Q}\right)\right]\,.
\end{aligned}
$$

Hence,

$$
|\bar{f}|_1^{-1}\sum_{l=N+1}^{N+Q}\left|\mathbb{E}_{\xi_1,\xi_2}\left[\bar{f}(X_n)\mathbb{1}_{d_{n-l}=0,n>\tau^2}\mathbb{1}_{X_{n-l+N+1:n}\notin\mathcal{D}}\right]\right| \le 3\epsilon\,,
$$

by Eqs. (11) and (14).

`Convergence of Term 4`. Let $n\ge N+Q+1$. We write (upon noting that $\mathbb{P}_{\xi_1,\xi_2}(d_n\le n)=1$)

$$
\begin{aligned}
\mathbb{E}_{\xi_1,\xi_2}\left[\bar{f}(X_n)\mathbb{1}_{N+Q<d_n,n>\tau^2}\right] &= \sum_{l=N+Q+1}^{n}\mathbb{E}_{\xi_1,\xi_2}\left[\bar{f}(X_n)\mathbb{1}_{d_n=l,n>\tau^2}\right] \\
&= \sum_{l=N+Q+1}^{n}\mathbb{E}_{\xi_1,\xi_2}\left[\bar{f}(X_n)\mathbb{1}_{d_{n-l}=0,n>\tau^2}\mathbb{1}_{X_{n-l+N+1:n}\notin\mathcal{D}}\right]\,.
\end{aligned}
$$

Furthermore, for any $N+Q+1\le l\le n$,

$$
\begin{aligned}
&|\bar{f}|_1^{-1}\left|\mathbb{E}_{\xi_1,\xi_2}\left[\bar{f}(X_n)\mathbb{1}_{d_{n-l}=0,n>\tau^2}\mathbb{1}_{X_{n-l+N+1:n}\notin\mathcal{D}}\right]\right| \\
&\quad\le \mathbb{E}_{\xi_1,\xi_2}\left[\mathbb{1}_{d_{n-l}=0}\,\mathbb{E}^{(n-l)}_{Z_{n-l}}\left[\mathbb{1}_{X_{N+1:l}\notin\mathcal{D}}\right]\right] \le \mathbb{E}_{\xi_1,\xi_2}\left[\mathbb{1}_{d_{n-l}=0}\,\mathbb{E}^{(n-l)}_{Z_{n-l}}\left[\mathbb{1}_{X_{N+1:N+1+Q}\notin\mathcal{D}}\right]\right] \\
&\qquad\le \mathbb{E}_{\xi_1,\xi_2}\left[\mathbb{1}_{d_{n-l}=0}\,\mathbb{P}^{(n-l)}_{Z_{n-l}}\left(\tau_{\mathcal{D}}\circ\underline{\theta}^N>Q\right)\right] \le \epsilon N^{-1}\,\mathbb{P}_{\xi_1,\xi_2}\left(d_{n-l}=0\right)\,,
\end{aligned}
$$

by Eq. (12). This concludes the proof upon noting that $\sum_{l=N+Q+1}^{n}\mathbb{P}_{\xi_1,\xi_2}\left(d_{n-l}=0\right)\le 1$.

4.3. **Proof of Proposition 2.2.** We only give the sketch of the proof since the proof is very similar to the proof of Theorem 2.1. In the present case, $\mathcal{D}=\mathsf{X}$. Let $\epsilon>0$. Choose $N, Q=2, n_\star$ as in the proof of Theorem 2.1. Since $\mathcal{D}=\mathsf{X}$, the sequence $\{\tau^k, k\ge 0\}$ is deterministic and for any $k\ge 1$, $\tau^{k+1}=\tau^k+N+1$. Finally, the bell variables $\{d_n, n\ge 0\}$ take the values $\{0,\cdots,N\}$. The proof is then similar to that of Theorem 2.1 upon noting that *(i)* Term 1 is similar ; *(ii)* Term 2 reduces to the third part only i.e. the term relative to $\bigcup_{q=0}^{N-1}\{d_n = q, d_{n-q-(N+1)} = 0\}$; *(iii)* Term 3 reduces to the first part i.e. the term relative to $\{d_{n-N} = 0\}$; *(iv)* Term 4 does not exist when $\mathcal{D} = \mathsf{X}$.

4.4. **Proof of Theorem 2.3.** 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.

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

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 $\{a_n, n \geq 1\}$ satisfies conditions which are verified e.g. with $a_n = (n+1)^{-\zeta}$ for some ζ such that

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

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

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

Hereafter, we assume that $|f|_{V^\beta} = 1$.

L^p-`martingale`. By definition, $\sum_{k=1}^{n}\left\{\hat{g}_{a_n}^{(k)}(X_k,\theta_k) - \mathbb{E}_{x,\theta}\left[\hat{g}_{a_n}^{(k)}(X_k,\theta_k)|\mathcal{F}_{k-1}\right]\right\}$ is a $\mathcal{F}_k$-martingale under $\mathbb{P}_{x,\theta}$. Let $p \stackrel{\text{def}}{=} (1-\alpha)\beta^{-1} > 1$. There exists a constant C such that for any n large enough,

$$
\mathbb{E}_{x,\theta}\left[\left|\sum_{k=1}^{n}\left\{\hat{g}_{a_n}^{(k)}(X_k,\theta_k) - \mathbb{E}_{x,\theta}\left[\hat{g}_{a_n}^{(k)}(X_k,\theta_k)|\mathcal{F}_{k-1}\right]\right\}\right|^p\right] \leq \frac{C\, n^{[p/2\vee 1]}}{a_n^p}V(x)\,. \tag{15}
$$

Indeed, the Burkholder's inequality (Hall and Heyde, 1980, Theorem 2.10) yields

$$
\mathbb{E}_{x,\theta}\left[\left|\sum_{k=1}^{n}\left\{\hat{g}_{a_n}^{(k)}(X_k,\theta_k) - \mathbb{E}_{x,\theta}\left[\hat{g}_{a_n}^{(k)}(X_k,\theta_k)|\mathcal{F}_{k-1}\right]\right\}\right|^p\right] \leq c\,\mathbb{E}_{x,\theta}\left[\left(\sum_{k=1}^{n}\left|\hat{g}_{a_n}^{(k)}(X_k,\theta_k)\right|^2\right)^{p/2}\right]\,,
$$

where c (and below, c_1, c_2) is a constant the value of which may vary upon each appearance. Assume first that $1 < p \leq 2$. By a convexity inequality,

$$\mathbb{E}_{x,\theta}\left[\left|\sum_{k=1}^{n}\left\{\hat{g}_{a_n}^{(k)}(X_k,\theta_k) - \mathbb{E}_{x,\theta}\left[\hat{g}_{a_n}^{(k)}(X_k,\theta_k)|\mathcal{F}_{k-1}\right]\right\}\right|^p\right] \leq c \sum_{k=1}^{n}\mathbb{E}_{x,\theta}\left[\left|\hat{g}_{a_n}^{(k)}(X_k,\theta_k)\right|^p\right] .$$

By Theorem 4.10 applied with $q = +\infty$, there exists N such that for any $n \geq 1$ and $(x,\theta) \in \mathsf{X} \times \Theta$,

$$\sup_{k\geq 1} |\hat{g}_{a_n}^{(k)}|^p(x,\theta) \leq c a_n^{-p}\ V^{\beta p}(x) + \left(\frac{N4\epsilon}{1-(1-a_n)^N}\right)^p .$$

This implies

$$\begin{aligned}\mathbb{E}_{x,\theta}&\left[\left|\sum_{k=1}^{n}\left\{\hat{g}_{a_n}^{(k)}(X_k,\theta_k) - \mathbb{E}_{x,\theta}\left[\hat{g}_{a_n}^{(k)}(X_k,\theta_k)|\mathcal{F}_{k-1}\right]\right\}\right|^p\right] \\ &\leq c a_n^{-p}\ \mathbb{E}_{x,\theta}\left[\sum_{k=1}^{n} V^{1-\alpha}(X_k)\right] + n\ c\ \left(\frac{4\epsilon N}{1-(1-a_n)^N}\right)^p \\ &\leq c a_n^{-p}\ \left(V(x) + \bar{b}n\right) + n\ c\ \left(\frac{4\epsilon N}{1-(1-a_n)^N}\right)^p ,\end{aligned}$$

where we used in the last inequality, Proposition 4.6(ii) with $a = 0$, $\beta = 1-\alpha$ and $l = 0$. As $n \to +\infty$,

$$c a_n^{-p}\ \left(V(x) + bn\right) + c\ n\ \left(\frac{4\epsilon N}{1-(1-a_n)^N}\right)^p \sim \frac{n}{a_n^p}\left(c_1 V(x) + c_2\right) .$$

We now consider the case $p \geq 2$. Applying the Minkowski's inequality,

$$\begin{aligned}\mathbb{E}_{x,\theta}\left[\left|\sum_{k=1}^{n}\left\{\hat{g}_{a_n}^{(k)}(X_k,\theta_k) - \mathbb{E}_{x,\theta}\left[\hat{g}_{a}^{(k)}(X_k,\theta_k)|\mathcal{F}_{k-1}\right]\right\}\right|^p\right] &\leq c\ \left(\sum_{k=1}^{n}\mathbb{E}_{x,\theta}\left[\left|\hat{g}_{a_n}^{(k)}(X_k,\theta_k)\right|^p\right]^{2/p}\right)^{p/2} \\ &\leq c\ n^{p/2-1}\ \sum_{k=1}^{n}\mathbb{E}_{x,\theta}\left[\left|\hat{g}_{a_n}^{(k)}(X_k,\theta_k)\right|^p\right] .\end{aligned}$$

Following the same lines as in the case $1 < p \leq 2$, we prove that there exists N such that for any $n \geq 1$ and $(x,\theta) \in \mathsf{X} \times \Theta$,

$$\begin{aligned}|\bar{f}|_{V^\beta}^{-p}\ \mathbb{E}_{x,\theta}&\left[\left|\sum_{k=1}^{n}\left\{\hat{g}_{a_n}^{(k)}(X_k,\theta_k) - \mathbb{E}_{x,\theta}\left[\hat{g}_{a_n}^{(k)}(X_k,\theta_k)|\mathcal{F}_{k-1}\right]\right\}\right|^p\right] \\ &\leq c a_n^{-p}\ n^{p/2-1}\ \left(V(x) + bn\right) + c\ n^{p/2}\ \left(\frac{4\epsilon N}{1-(1-a_n)^N}\right)^p .\end{aligned}$$

As $n \to +\infty$, the rhs is equivalent to $n^{p/2}/a_n^p\,(c_1 V(x) + c_2)$.

`Convergence of Term 1.` We prove that

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

provided $\sum_n n^{-1}\ \left(n^{1-[0.5\vee(\beta/(1-\alpha))]}a_n\right)^{-(1-\alpha)/\beta} < +\infty$, $\lim_n n^{1-[0.5\vee(\beta/(1-\alpha))]}a_n = +\infty$ and $\lim_n a_n = 0$.

Proof. $\sum_{k=1}^{n}\left\{\hat{g}_{a_n}^{(k)}(X_k,\theta_k)-\mathbb{E}_{\xi_1,\xi_2}\left[\hat{g}_{a_n}^{(k)}(X_k,\theta_k)|\mathcal{F}_{k-1}\right]\right\}$ is a martingale which possesses a L^p-moment (see Eq. (15)) with $p=(1-\alpha)\beta^{-1}>1$. By applying Lemma B.1 and (15), there exists a constant C such that for any n large enough (i.e. such that $1-a_n\geq 1/2$)

$$\begin{aligned}
&\mathbb{P}_{\xi_1,\xi_2}\left(\sup_{m\geq n} m^{-1}(1-a_m)^{-1}\left|\sum_{k=1}^{m}\left\{\hat{g}_{a_m}^{(k)}(X_k,\theta_k)-\mathbb{E}_{\xi_1,\xi_2}\left[\hat{g}_{a_m}^{(k)}(X_k,\theta_k)|\mathcal{F}_{k-1}\right]\right\}\right|\geq\delta\right)\\
&\quad\leq \mathbb{P}_{\xi_1,\xi_2}\left(\sup_{m\geq n} m^{-1}\left|\sum_{k=1}^{m}\left\{\hat{g}_{a_m}^{(k)}(X_k,\theta_k)-\mathbb{E}_{\xi_1,\xi_2}\left[\hat{g}_{a_m}^{(k)}(X_k,\theta_k)|\mathcal{F}_{k-1}\right]\right\}\right|\geq\delta/2\right)\\
&\qquad\leq C\xi_1(V)\sum_{m\geq n}\left(m^{-p}-(m+1)^{-p}\right)\ \frac{m^{[p/2\vee 1]}}{a_m^p}+C\xi_1(V)\lim_{n\to+\infty}\frac{1}{n^p}\frac{n^{[p/2\vee 1]}}{a_n^p}.
\end{aligned}$$

Under the stated assumptions, the rhs tends to zero. □

`Convergence of Term 2.` We prove that

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

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

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

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

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

`Convergence of Term 3.` We prove that

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

provided 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 $p \stackrel{\text{def}}{=} (1-\alpha)\beta^{-1} > 1$

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

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

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

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

`Convergence of Term 4.` We prove that

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

provided $\{a_n^{1\wedge[(1-\alpha-\beta)/\alpha]}\ n^{-1}, n\geq 1\}$ 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 $n \geq 1$, $l \geq 0$, $(x,\theta) \in \mathsf{X}\times\Theta$,

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

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

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

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 the Kronecker Lemma (see e.g (Hall and Heyde, 1980, Section 2.6)), this amounts to prove 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)

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

and the rhs tends to zero under the stated assumptions. □

4.5. **Proof of Proposition 2.4.** We only give the sketch of the proof since the proof is very similar to that of Theorem 2.3. We start with 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, $l \geq 0$, $(x,\theta) \in \mathsf{X}\times\Theta$

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

We then introduce the martingale decomposition as in the proof of Theorem 2.3 and follow the same lines (with any $p > 1$).

Appendix A. Explicit control of convergence

We provide sufficient conditions for the assumptions A3 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

$$P(x, \cdot) \geq \mathbb{1}_{\mathcal{C}}(x) \, \varepsilon \, \nu(\cdot) \, , \qquad PV \leq V - c \, V^{1-\alpha} + b \, \mathbb{1}_{\mathcal{C}} \, . \tag{16}$$

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 $\nu(\mathcal{C}) > 0$. Then there exists a constant C depending upon $\sup_{\mathcal{C}} V$, $\nu(\mathcal{C})$ and $\varepsilon, \alpha, b, c$ such that for any $0 \leq \beta \leq 1 - \alpha$ and $0 \leq \kappa \leq \alpha^{-1}(1 - \beta)$,

$$(n + 1)^{\kappa} \, \| P^n(x, \cdot) - \pi(\cdot) \|_{V^{\beta}} \leq C \ V^{\beta + \alpha\kappa}(x). \tag{17}$$

The proof of (i) can be found in (Meyn and Tweedie, 1993, Theorems 8.4.3., 10.0.1). The proof of (ii) is given in e.g. 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 phi-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. Then there 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*

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

and $\sup_{\mathcal{D}} V < +\infty$, $\nu_{\mathcal{D}}(\mathcal{D}) > 0$, 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 $\mathcal{D}_\gamma \stackrel{\mathrm{def}}{=} \{V^{1-\alpha} \leq b[c(1-\gamma)]^{-1}\}$. 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 (16) 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 the level sets are petite (and thus m-small for some $m \geq 1$), the explicit upper bounds get intricate and are never detailed in the literature (at least 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. Birnbaum-Marshall's inequality

Lemma B.1 is a generalization of the Birnbaum-Marshall's inequality, which can be found in (Andrieu and Moulines, 2006, Proposition 22).

Lemma B.1. *Let* $\{S_k, k \geq 1\}$ *be a* $\mathcal{F}_k$*-sub-martingale and* $\{b_k, k \geq 1\}$ *be a non-increasing real-valued sequence. If* $p \geq 1$ *is such that* $\mathbb{E}[|S_k|^p] < +\infty$ *for* $k \leq N$ *then for* $n \leq N$

$$\mathbb{P}\left(\sup_{n \leq m \leq N} b_m |S_m| \geq 1\right) \leq \sum_{m=n}^{N} (b_m^p - b_{m+1}^p)\ \mathbb{E}[|S_m|^p] + b_N^p \mathbb{E}[|S_N|^p].$$

Appendix C. Proofs of 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 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 D1-2. 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* $s \leq s_\star$,

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

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

$$|\nabla^2 V_s(x + tz)| \leq sV_s(x) \quad \frac{V_s(x + tz)}{V_s(x)} |\nabla \ln \pi(x + tz)\nabla \ln \pi(x + tz)^T| \cdots$$
$$\left(s + \frac{|\nabla^2 \ln \pi(x + tz)|}{|\nabla \ln \pi(x + tz)\nabla \ln \pi(x + tz)^T|}\right)$$

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

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

and

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

Finally,

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

which yields the desired result upon noting that $|z||x + tz|^{m-1} \leq \eta|x|^{\upsilon+m-1}(1 - \eta|x|^{\upsilon-1})$ 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:

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

If x remains in a compact set $\mathcal{C}$, using D2(ii) and the continuity of $x \mapsto V_s(x)$, we have $V_s(x+z) \leq C(1 + \exp(sD_0|z|^m))$. It follows that

$$\sup_{\theta\in\Theta} \sup_{x \in \mathcal{C}} \{P_\theta V_s(x) - V_s(x)\} \leq C \ \sup_{\theta \in \Theta} \int_{R(x)-x} (1 + \exp(sD_0|z|^m))\ q_\theta(z)\ \mu_{Leb}(dz) < +\infty\ .$$

More generally, let x large enough. Define $l(x) \stackrel{\text{def}}{=} \log \pi(x)$, $R_V(x,z) \stackrel{\text{def}}{=} V_s(x+z) - V_s(x) + sV_s(x)\langle z, \nabla l(x)\rangle$, $R_\pi(x,z) \stackrel{\text{def}}{=} \pi(x+z)(\pi(x))^{-1} - 1 - \langle z, \nabla l(x)\rangle$. 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

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

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

and

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

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 $\theta \in \Theta$,

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

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

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

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

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

There exists a constant C such that for $\theta \in \Theta$ and $s \leq s_\star$, the first term in the rhs 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 $\sup_{\theta \in \Theta} \int \exp(\lambda_\star |z|^2) q_\theta(z) \mu_{Leb}(dz) < +\infty$. Under D3, the third term is upper bounded by $C\ V_s(x)\ |x|^{2(m-1)}\ \exp(-\lambda \eta^2 |x|^{2v})$ for some $\lambda \in (0, \lambda_\star)$, uniformly in θ for $\theta \in \Theta$, and in s for $s \leq s_\star$. Hence, we proved that there exists $C_\star < \infty$ such that for any $s \leq s_\star$,

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

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

C.1.3. *Third term.* Following the same lines as in the control of $I_2(x, \theta, s)$, it may be proved that

$$I_3(x, \theta, s) \leq s V_s(x) D_1 |x|^{m-1} \int_{\{z, |z| \geq \eta |x|^{\upsilon}\}} |z| \left(1 + D_1 |z| |x|^{m-1}\right) q_\theta(z) \mu_{Leb}(dz)$$
$$+ C\ V_s(x) |x|^{3(m-1)} \int_{\{z, |z| \leq \eta |x|^{\upsilon}\}} |z|^3\ q_\theta(z) \mu_{Leb}(dz) \leq C\ V_s(x) |x|^{2(m-1)} \varepsilon(x)$$

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 $c_\star \stackrel{\text{def}}{=} b d_1^2 - C_\star s > 0$, we have

$$\sup_{\theta \in \Theta} P_\theta V_s(x) \leq V_s(x) - c_\star V_s(x) |x|^{2(m-1)} + b_\star \mathbb{1}_{\mathcal{C}}(x) \tag{18}$$
$$\leq V_s(x) - 0.5 c_\star V_s^{1-\alpha}(x) + b_\star \mathbb{1}_{\mathcal{C}}(x) \tag{19}$$

for a compact set $\mathcal{C}$. This proves A2 and A4. A3 and A5 follow from the results of Appendix A.

C.2. **Proof of Lemma 3.4.** An easy modification in the proof of (Andrieu and Moulines, 2006, Proposition 11) (to adjust for the difference in the drift function) shows that $D(\theta, \theta') \leq 2 \int_{\mathsf{X}} |q_{e^c \Sigma}(x) - q_{e^{c'} \Sigma'}(x)| \mu_{Leb}(dx)$. We then apply (Andrieu and Moulines, 2006, Lemma 12) to obtain that

$$D(\theta, \theta') \leq 2p\ \Lambda_l^{-1} \sqrt{p} |e^c \Sigma - e^{c'} \Sigma'|_{\mathrm{s}} \leq 2 p^{3/2}\ \Lambda_l^{-1} e^{\kappa_u}\ \left(|\Sigma - \Sigma'|_{\mathrm{s}} + \Lambda_u |c - c'|\right)\ .$$

For any $l, n \geq 0$, $\epsilon > 0$, $x \in \mathbb{R}^p$ and $\theta \in \Theta$, we have

$$\begin{aligned} \mathbb{P}_{x,\theta}^{(l)} \left(D(\theta_n, \theta_{n+1}) \geq \epsilon\right) &\leq \epsilon^{-1} \mathbb{E}_{x,\theta}^{(l)} \left[D(\theta_n, \theta_{n+1})\right] \\ &\leq C_0 (l + n + 1)^{-1} \Big(2\Lambda_u + \mathbb{E}_{x,\theta}^{(l)} \left[|X_{n+1}|^2\right] + C^2 \\ &\qquad + 2C \sqrt{\mathbb{E}_{x,\theta}^{(l)} \left[|X_{n+1}|^2\right]}\Big)\ , \end{aligned}$$

where $C_0 = 2p^{3/2}e^{\kappa_u}\Lambda_l^{-1}\epsilon^{-1}$. 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 $\mathbb{E}^{(l)}_{x,\theta}\left[|X_n|^2\right] \leq C\ [\ln V_s(x)]^{2/m}\ [\ln n]^{2/m}$. We conclude that for any probability ξ_1 such that $\xi_1([\ln V_s]^{2/m}) < +\infty$, $\lim_n \mathbb{P}_{\xi_1,\xi_2}\left(D(\theta_n, \theta_{n+1}) \geq \epsilon\right) = 0$ and for any level set $\mathcal{D}$ of V_s,

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

Acknowledgment: 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.

References

Andrieu, C. and Atchade, Y. F. (2008). On the efficiency of adaptive mcmc algorithms. *Electronic Communications in Probability* (to appear).

Andrieu, C. and Fort, G. (2005). Explicit control of subgeometric ergodicity. Tech. rep., University of Bristol, 05:17. Available from 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.

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

Andrieu, C. and Tadic, V. (2008). General result for the stability of controlled mcmc. Tech. rep., Bristol University. (personal communication).

Atchade, Y. and Fort, G. (2008). Limit theorems for some adaptive MCMC algorithms with subgeometric kernels (II). Tech. rep., Work in progress.

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

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

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

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.

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

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

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

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

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

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.

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

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

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

Holden, L. (1998). Adaptive chains. *Technical Report* .

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

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

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

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

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.* Springer-Verlag London Ltd., London.

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

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

ROBERTS, G. O. and ROSENTHAL, J. S. (2007). Coupling and ergodicity of adaptive MCMC. *Journal of Applied Probablity* **44** 458–475.

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

WINKLER, G. (2003). *Image analysis, random fields and Markov chain Monte Carlo methods*, vol. 27 of *Applications of Mathematics (New York)*. 2nd ed. Springer-Verlag, Berlin.

YANG, C. (2007). Recurrent and ergodic properties of Adaptive MCMC. Tech. rep., University of Toronto, Canada. Available from http://probability.ca/jeff/ftpdir/chao3.pdf.