STABILITY OF STOCHASTIC APPROXIMATION UNDER VERIFIABLE CONDITIONS

CHRISTOPHE ANDRIEU *, ÉRIC MOULINES †, AND PIERRE PRIOURET ‡

Abstract. In this paper we address the problem of the stability and convergence of the stochastic approximation procedure

$$\theta_{n+1} = \theta_n + \gamma_{n+1}[h(\theta_n) + \xi_{n+1}].$$

The stability of such sequences $\{\theta_n\}$ is known to heavily rely on the behaviour of the mean field h at the boundary of the parameter set and the magnitude of the stepsizes used. The conditions typically required to ensure convergence, and in particular the boundedness or stability of $\{\theta_n\}$, are either too difficult to check in practice or not satisfied at all. This is the case even for very simple models. The most popular technique to circumvent the stability problem consists of constraining $\{\theta_n\}$ to a compact subset $\mathcal{K}$ in the parameter space. This is obviously not a satisfactory solution as the choice of $\mathcal{K}$ is a delicate one. In the present contribution we first prove a "deterministic" stability result which relies on simple conditions on the sequences $\{\xi_n\}$ and $\{\gamma_n\}$. We then propose and analyze an algorithm based on projections on adaptive truncation sets which ensures that the aforementioned conditions required for stability are satisfied. We focus in particular on the case where $\{\xi_n\}$ is a so-called Markov state-dependent noise. We establish both the stability and convergence w.p. 1 of the algorithm under a set of simple and verifiable assumptions. We illustrate our results with an example related to adaptive Markov chain Monte Carlo algorithms.

Key words. Stochastic approximation, state-dependent noise, randomly varying truncation, Adaptive Markov Chain Monte Carlo.

AMS subject classifications. 62L20,90C15

1. Introduction.

In many contexts it is of interest to find the roots of possibly non linear equations of the form

$$h(\theta) = 0, \qquad \theta \in \Theta, \tag{1.1}$$

for some mapping $h : \Theta \to \mathbb{R}^{n_\theta}$, where $\Theta \subset \mathbb{R}^{n_\theta}$ for some integer n_θ. Most of the methods for solving the previous equation are iterative, *i.e.* produce a sequence of iterates $\{\theta_n, n \geq 0\}$ which eventually converges to the set of solutions of Eq. (1.1),

$$\mathcal{S} := \{\theta \in \Theta, h(\theta) = 0\}\,. \tag{1.2}$$

Stochastic Approximation (SA) is a class of algorithms to solve Eq. (1.1) in the situation where only noisy measurements of h are available. In its simplest form, the Robbins-Monro algorithm produces a sequence $\{\theta_n, n \geq 0\}$ defined recursively as follows,

$$\theta_0 \in \Theta, \qquad \theta_{n+1} = \theta_n + \gamma_{n+1}\zeta_{n+1}, \quad n \geq 1, \tag{1.3}$$

where $\{\gamma_n, n \geq 1\}$ is a sequence of stepsizes which satisfies standard conditions (say $\gamma_n \downarrow 0$ and $\sum_{n\geq 1}\gamma_n = \infty$) and for any $n \geq 1$, ζ_n is a noisy measurement of $h(\theta_n)$. It is useful to introduce the sequence

*University of Bristol, School of Mathematics, University Walk, BS8 1TW, UK (`c.andrieu@bris.ac.uk`). C. Andrieu would like to thank the CNRS/Royal Society Partnership and the Nuffield Foundation.

†École Nationale Supérieure des Télécommunications, URA CNRS 820, 46, rue Barrault, F 75634 PARIS Cedex 13 (`moulines@tsi.enst.fr`). É. Moulines would like to thank the CNRS/Royal Society Partnership.

‡Université Pierre & Marie Curie, Laboratoire de Probabilités et Modélisation Aléatoire, URA CNRS 224, F 75252 PARIS Cédex 05 (`priouret@ccr.jussieu.fr`)

$\{\xi_n, n \geq 1\}$ defined as

$$\zeta_{n+1} = h(\theta_n) + \xi_{n+1}, \tag{1.4}$$

which will be referred to as the *noise sequence*. Convergence of SA has been studied under various sets of assumptions for the mean field h and the noise sequence $\{\xi_n, n \geq 1\}$ since the early work by [23]; see *e.g.* [5],[18], [24], [16] and the references therein. Essentially, convergence of the SA sequence can be established toward an *attractive* subset provided that the sequence $\{\theta_n, n \geq 0\}$ is with probability 1 (hereafter w.p. 1) in a compact subset of Θ and is w.p. 1 infinitely often in the *domain of attraction* of this attractive subset. Showing in practice that $\{\theta_n, n \geq 0\}$ satisfies these boundedness and recurrence conditions proves to be a difficult task. The available results hold under conditions which are still restrictive, despite recent advances (see [1], [7], [6] and references therein). This major drawback has motivated the design of modified Robbins-Monro recursions. Probably the most widely used method in practice consists of constraining the sequence $\{\theta_n, n \geq 0\}$ to some compact set $\mathcal{K} \subset \Theta$ by means of a reprojection onto $\mathcal{K}$. This method has been thoroughly investigated in [24] (see also [8] and the references therein). Although relatively easy to implement, and sound when constraints about the system considered are available *a priori*, this approach becomes impractical and questionable in many situations of interest.

Our contributions to solve the stability and convergence problems are here twofold:

First we establish and prove in Section 2 a general result of stability, Theorem 2.2, for deterministic sequences of the form given by Eqs. (1.3)-(1.4). This key deterministic result assumes the existence of a global Lyapunov function for the mean field h and mild general assumptions about the noise and stepsize sequences. In contrast with previous results, the conditions required on the growth of the Lyapunov functions and the mean field h when θ approaches the boundaries of the parameter set Θ are minimal. As a consequence the result is applicable to quite general settings. We then show that, under the conditions that guarantee stability, the convergence of the deterministic sequence Eq. (1.3)-(1.4) is ensured (see Theorem 2.3).

Our second contribution here consists of proposing a SA algorithm (Section 3) for which the aforementioned noise and stepsize conditions are satisfied w.p. 1. There are many different applications of stochastic approximations which imply markedly different types of assumptions on the noise sequence $\{\xi_n\}$. Whereas our deterministic stability and convergence results mentioned above can be applied quite generally, we focus in this paper on the subtle Markov state dependent noise (see [24, Chapter 6, Section 6.6] and Section 3 in this paper), for which the availability of algorithms whose convergence can be established under general but nevertheless verifiable assumptions is still missing. The proposed algorithm is a modification of the classical Robbins-Monro procedure described in Eq. (1.3)-(1.4), based on truncations on adaptive truncation sets, in the spirit of the seminal works [11] and [10].

The convergence of SA with adaptive truncation sets has been considered under various conditions on the noise sequence $\{\xi_n\}$. These include state-independent noise conditions (see for example [12, Section 2.4, pp. 42-44]) but also state-dependent martingale differences ([32], [14], [9], [12, Section 2.5, pp. 49-57]) or state-dependent ϕ-mixing processes ([9], [12, Section 2.5, pp. 49]). However the application of this strategy to the Markovian state dependent case requires even more care, and it is therefore not surprising to find that the results on the topic are scarce, and have been obtained under conditions that are more stringent than those considered in the present paper; see [33], [13] and for the special case of ARMAX models, [12, Chapter 6]. As we shall see our procedure differs in some respects from the original procedure proposed by [11] and [10], and offers additional degrees of freedom. Our technique of proof for the stability relies on a novel approach and offers as a byproduct an explicit bound for the tail probability of the number of reprojections, which is found to be super-exponential under mild technical conditions.

In order to illustrate our findings and their applicability, we propose (see Section 7) to analyse the convergence of an adaptive Markov chain Monte Carlo (MCMC) algorithm recently proposed in [20] and analysed under more stringent conditions than those considered here. Other examples can be found in [2].

2. Key deterministic results. In this section we establish both stability and convergence results for deterministic recursions of the type described in Eqs. (1.3)-(1.4). Before stating our first assumptions, some definitions and notation are needed. Let d be a positive integer. An element v of $\mathbb{R}^d$ is denoted by its column vector v and its transpose is denoted by v^{T}. For elements v, w of $\mathbb{R}^d$, we denote $< v, w >$ their inner product, so that $|v| = \sqrt{< v, v >}$ denotes the norm of v. Our first assumption is the existence of a global Lyapunov function w for the mean field h. Denoting $\mathcal{W}_M := \{\theta \in \Theta, w(\theta) \leq M\} \subset \Theta$ we assume,

(**A1**) Θ is an open subset of $\mathbb{R}^{n_\theta}$, $h : \Theta \to \mathbb{R}^{n_\theta}$ is continuous and there exists a continuously differentiable function $w : \Theta \to [0, \infty)$ such that
   (i) There exists $M_0 > 0$ such that

$$\mathcal{L} := \left\{\theta \in \Theta, \left\langle \nabla w(\theta), h(\theta) \right\rangle = 0\right\} \subset \{\theta \in \Theta, w(\theta) < M_0\},$$

   (ii) There exists $M_1 \in (M_0, \infty]$ such that $\mathcal{W}_{M_1}$ is a compact set,
   (iii) For any $\theta \in \Theta \setminus \mathcal{L}$, $\left\langle \nabla w(\theta), h(\theta) \right\rangle < 0$,
   (iv) The closure of $w(\mathcal{L})$ has an empty interior.

If h is a gradient field, *i.e.* $h = -\nabla J$ for some lower bounded real valued and differentiable function $\theta \mapsto J(\theta)$, then the choice $w = J$ is appropriate, provided that J is continuously differentiable. Note that in situations where the set of stationary points cannot be characterized explicitly, one might use Sard's theorem from differential geometry in order to check (A1-iv). Indeed, Sard's theorem states that if w is n_θ-times continuously differentiable, then $w(\{\nabla w = 0\})$ has an empty interior.

Our approach to prove our stability and convergence results can be decomposed into two distinct steps. In the first step (this section), we establish deterministic conditions on a noise sequence $\{\xi_n\}$ and a stepsize sequence $\{\rho_n\}$ upon which a deterministic sequence $\{\theta_n\}$ defined as

$$\theta_0 \in \Theta \quad \theta_{n+1} = \theta_n + \rho_{n+1}[h(\theta_n) + \xi_{n+1}] \quad \text{for} \quad n \geq 0, \tag{2.1}$$

has the following properties: (i) it remains in a compact subset of Θ (see Theorem 2.2) and (ii) provided that $\{\theta_n\}$ remains in a compact subset of Θ, converges to $\mathcal{L}$ (Theorem 2.3). In a second step - which is probabilistic in nature and depends on how the noise is generated - we develop a general algorithm for the case where $\{\xi_n\}$ follows a Markovian state-dependent dynamic which allows one to show that the required condition on $\{\xi_n\}$ is satisfied w.p. 1 (Sections 3-6).

Before proving Theorem 2.2 and Theorem 2.3 we prove in the following lemma a fundamental contraction property of the Lyapunov function w. This result is the crux to both the proof of stability and convergence.

LEMMA 2.1. *Assume (A1).Then*

(i) Let $\mathcal{K} \subset \Theta$ *be a compact subset such that* $0 < \inf_{\theta \in \mathcal{K}} \left|\left\langle \nabla w, h \right\rangle\right|$. *For any* $0 < \delta < \inf_{\theta \in \mathcal{K}} \left|\left\langle \nabla w, h \right\rangle\right|$, *there exist* $\lambda > 0$ *and* $\beta > 0$, *such that, for any* ρ, $0 \leq \rho \leq \lambda$, ζ, $|\zeta| \leq \beta$, *and* $\theta \in \mathcal{K}$, $w(\theta + \rho h(\theta) + \rho\zeta) \leq w(\theta) - \rho\delta$.

(ii) For any $M \in (M_0, M_1]$ *(where* M_0 *is defined in (A1-i) and* M_1 *is defined in (A1-ii)), there exist* $\lambda > 0$ *and* $\beta > 0$ *such that for any* ρ, $0 \leq \rho \leq \lambda$, ζ, $|\zeta| \leq \beta$, *and* $\theta \in \mathcal{W}_M$, $\theta + \rho h(\theta) + \rho\zeta \in \mathcal{W}_M$.

Proof. We first prove (i). For any $0 < \delta < \inf_{\theta \in \mathcal{K}} |\left\langle \nabla w, h \right\rangle|$, there exist $\lambda > 0$ and $\beta > 0$ such that for all ρ, $0 \leq \rho \leq \lambda$, ζ, $|\zeta| \leq \beta$ and t, $0 \leq t \leq 1$, we have for all $\theta \in \mathcal{K}$, $\theta + \rho t h(\theta) + \rho t \zeta \in \Theta$ and

$$\left| \left\langle \nabla w(\theta), h(\theta) \right\rangle - \left\langle \nabla w(\theta + \rho t h(\theta) + \rho t \zeta), h(\theta) + \zeta \right\rangle \right| \leq \inf_{\theta \in \mathcal{K}} |\left\langle \nabla w, h \right\rangle| - \delta.$$

Then for any ρ, $0 \leq \rho \leq \lambda$ and ζ, $|\zeta| \leq \beta$,

$$\begin{aligned} w(\theta+\rho h(\theta)+\rho\zeta)-w(\theta) &= \rho\left\langle \nabla w(\theta), h(\theta) \right\rangle + \rho \int_0^1 \left(\left\langle \nabla w(\theta + t\rho h(\theta) + t\rho\zeta), h(\theta) + \zeta \right\rangle - \left\langle \nabla w(\theta), h(\theta) \right\rangle \right) dt \\ &\leq -\rho \inf_{\theta \in \mathcal{K}} |\left\langle \nabla w, h \right\rangle| + \rho \left(\inf_{\theta \in \mathcal{K}} |\left\langle \nabla w, h \right\rangle| - \delta \right) = -\rho\delta. \end{aligned}$$

We now prove (ii). Consider $M' \in (M_0, M)$. Since $\mathcal{W}_{M'}$ is compact and w continuous, there exists $\lambda_0 > 0$ and $\beta_0 > 0$ such that, for all $0 \leq \rho \leq \lambda_0$, $|\zeta| \leq \beta_0$ and $\theta \in \mathcal{W}_{M'}$ then $\theta + \rho h(\theta) + \rho\zeta \in \mathcal{W}_M$. We can apply (i) to the set $\mathcal{K} = \{\theta \in \Theta, M' \leq w(\theta) \leq M\}$ to show that there exists λ_1, β_1 such that, for all ρ, $0 \leq \rho \leq \lambda_1$, ζ, $|\zeta| \leq \beta_1$ and $\theta \in \mathcal{K}$, $w(\theta + \rho h(\theta) + \rho\zeta) \leq w(\theta) \leq M$, showing that $\theta + \rho h(\theta) + \rho\zeta \in \mathcal{W}_M$. □

2.1. Boundedness. In this section, we show that under (A1) and mild additional conditions on $\{\xi_n\}$ and $\{\rho_n\}$, then the sequence defined in Eq. (2.1) remains in a compact subset of Θ.

THEOREM 2.2. *Assume (A1). For any $M \in (M_0, M_1]$ there exist $\delta_0 > 0$ and $\lambda_0 > 0$ such that, for all $n \geq 1$, all $\theta_0 \in \mathcal{W}_{M_0}$, all sequences $\{\rho_k\}$ of non negative integers and all sequences $\{\xi_k\}$ of n_θ-dimensional vectors satisfying*

$$\sup_{1 \leq k \leq n} \rho_k \leq \lambda_0 \quad \textit{and} \sup_{1 \leq k \leq n} \left| \sum_{j=1}^{k} \rho_j \xi_j \right| \leq \delta_0,$$

we have for $k \in \{1, \dots, n\}$, $w(\theta_k) \leq M$, where $\theta_k = \theta_{k-1} + \rho_k h(\theta_{k-1}) + \rho_k \xi_k$.

Proof. Let M' be such that $M' \in (M_0, M)$. Lemma 2.1 shows that there exists $\lambda_0 > 0$, $\beta_0 > 0$ such that, for all θ, ρ and ζ satisfying $w(\theta) \leq M'$, $0 \leq \rho \leq \lambda_0$ and $|\zeta| \leq \beta_0$,

$$w(\theta + \rho h(\theta) + \rho\zeta) \leq M'. \tag{2.2}$$

By continuity of h and w there exists $\delta_0 \in (0, \beta_0]$ such that for all $(\theta, \bar{\theta}) \in \Theta \times \Theta$ satisfying $w(\theta) \leq M$ and $|\theta - \bar{\theta}| \leq \delta_0$, we have

$$|h(\bar{\theta}) - h(\theta)| \leq \beta_0 \quad \text{and} \quad |w(\bar{\theta}) - w(\theta)| \leq M - M'. \tag{2.3}$$

We will now prove by induction that, for all $k \in \{1, \dots, n\}$, $w(\bar{\theta}_k) \leq M'$ and $w(\theta_k) \leq M$, where the sequence $\{\bar{\theta}_k\}$ is defined recursively as follows: $\bar{\theta}_0 = \theta_0$ and for all $k \in \{1, \dots, n\}$,

$$\bar{\theta}_k = \bar{\theta}_{k-1} + \rho_k h(\theta_{k-1}).$$

Under the stated assumptions $w(\theta_0) = w(\bar{\theta}_0) \leq M_0$ and since $0 \leq \rho_1 \leq \lambda_0$ and $|\theta_1 - \bar{\theta}_1| = |\rho_1 \xi_1| \leq \delta_0$, on the one hand Lemma 2.1 shows that $w(\bar{\theta}_1) = w(\bar{\theta}_0 + \rho_1 h(\bar{\theta}_0)) \leq M'$ and on the other hand $w(\theta_1) = w(\theta_0 + \rho_1 h(\theta_0) + \rho_1 \xi_1) \leq M$, which proves the result for $n = 1$. Assume now that the result holds up to $1 \leq k \leq n-1$ for $n > 1$. By construction, for $j \in \{1, \dots, k\}$, $\theta_j - \bar{\theta}_j = \theta_{j-1} - \bar{\theta}_{j-1} + \rho_j \xi_j$, which implies that

$$\theta_j - \bar{\theta}_j = \sum_{i=1}^{j} \rho_i \xi_i. \tag{2.4}$$

Under the stated assumptions and Eq. (2.3), for $j \in \{1, \ldots, k\}$, $|\theta_j - \bar{\theta}_j| \leq \delta_0$ and $|h(\theta_j) - h(\bar{\theta}_j)| \leq \beta_0$. On the other hand,

$$\bar{\theta}_{k+1} = \bar{\theta}_k + \rho_{k+1} h(\theta_k) = \bar{\theta}_k + \rho_{k+1} h(\bar{\theta}_k) + \rho_{k+1}\left(h(\theta_k) - h(\bar{\theta}_k)\right).$$

Since $0 \leq \rho_{k+1} \leq \lambda_0$ and $w(\bar{\theta}_k) \leq M'$, Lemma 2.1 shows that $w(\bar{\theta}_{k+1}) \leq M'$. Using again that $|\theta_{k+1} - \bar{\theta}_{k+1}| \leq \delta_0$, Eq. (2.3) implies that $w(\theta_{k+1}) \leq M$, which concludes the proof. □

2.2. Convergence. Theorem 2.2 provides us with conditions on $\{\xi_n\}$ and $\{\rho_n\}$ upon which a sequence as defined in Eq. (2.1) stays within a compact subset of Θ. In the next proposition we show that whenever $\{\theta_k\}$ stays in a compact subset of Θ, then under mild additional assumptions it converges to $\mathcal{L}$. The key result of this section is the following theorem, adapted here from [14, Theorem 2] (see [12] for a similar result). For an integer d and A a subset of $\mathbb{R}^d$, we define $d(x, A) = \inf\{y \in A, |x-y|\}$. For any set $A \subset \Theta$ and any $\delta > 0$, we define $A_\delta := \{\theta \in \Theta, d(\theta, A) \leq \delta\}$; for any function $\phi : \Theta \to \mathbb{R}$, we define $\|\phi\|_A := \sup_{\theta \in A} |\phi(\theta)|$.

THEOREM 2.3. *Assume (A1). Let $\mathcal{K}$ be a compact subset of Θ such that $\mathcal{L} \cap \mathcal{K} \neq \emptyset$. Let $\{\rho_k\}$ be a monotone non-increasing sequence of positive numbers such that $\rho_0 \leq \lambda_0$ (where λ_0 is given in Theorem 2.2),*

$$\sum_{k=1}^{\infty} \rho_k = \infty \quad \textit{and} \quad \lim_{k \to \infty} \rho_k = 0.$$

Let $\{\xi_n\}$ be a sequence in $\mathbb{R}^{n_\theta}$ satisfying $\limsup_{k\to\infty} \sup_{l \geq k} \left|\sum_{i=k}^{l} \rho_i \xi_i\right| = 0$. Assume that the sequence defined by $\theta_k = \theta_{k-1} + \rho_k h(\theta_{k-1}) + \rho_k \xi_k$, is such that $\{\theta_k\} \subset \mathcal{K}$. Then, $\limsup_{k\to\infty} d(\theta_k, \mathcal{L} \cap \mathcal{K}) = 0$.

We preface the proof of this theorem with two lemmas. For both lemmas, the assumptions of Theorem 2.3 are assumed to hold.

LEMMA 2.4. *Let $\mathcal{N} \subset \Theta$ be an open neighbourhood of $\mathcal{L} \cap \mathcal{K}$. There exist positive constants δ, ε and λ (depending only on the sets $\mathcal{N}$ and $\mathcal{K}$), such that for any $\delta' \in (0, \delta]$, $\lambda' \in (0, \lambda]$ and $\eta > 0$, one can find an integer N and a sequence $\{\bar{\theta}_j\}_{j \geq N}$ satisfying*

$$\sup_{j \geq N} |\theta_j - \bar{\theta}_j| \leq \delta', \quad \sup_{j \geq N} \rho_j \leq \lambda', \quad \textit{and} \quad \sup_{j \geq N} |w(\theta_j) - w(\bar{\theta}_j)| \leq \eta, \tag{2.5}$$

$$w(\bar{\theta}_j) \leq w(\bar{\theta}_{j-1}) - \rho_j \varepsilon + (\eta + \rho_j \varepsilon)\, \mathbb{1}_{\mathcal{N}}(\bar{\theta}_{j-1}) \quad \textit{for} \quad j \geq N+1. \tag{2.6}$$

Proof. Let us choose $\delta_0 > 0$ such that the compact set $\mathcal{K}_{\delta_0} \subset \Theta$. The set $\mathcal{K}_{\delta_0} \setminus \mathcal{N}$ is compact and $\sup_{\mathcal{K}_{\delta_0} \setminus \mathcal{N}} \left\langle \nabla w, h \right\rangle < 0$. By Lemma 2.1, for any $\varepsilon > 0$ such that $\sup_{\theta \in \mathcal{K}_{\delta_0} \setminus \mathcal{N}} \left\langle \nabla w(\theta), h(\theta) \right\rangle < -\varepsilon$, one may choose $\lambda > 0$ and $\beta > 0$ small enough so that for any $\rho \in [0, \lambda]$, $|\zeta| \leq \beta$ and $\theta \in \mathcal{K}_{\delta_0} \setminus \mathcal{N}$,

$$w(\theta + \rho h(\theta) + \rho \zeta) \leq w(\theta) - \rho \varepsilon. \tag{2.7}$$

Using the uniform continuity of continuous functions on compact sets, for any $\eta > 0$ one may choose $\delta \in (0, \lambda \|h\|_{\mathcal{K}}]$ small enough so that for all $(\theta, \bar{\theta}) \in \mathcal{K}_{\delta_0} \times \mathcal{K}_{\delta_0}$ satisfying $|\theta - \bar{\theta}| \leq \delta \leq \lambda \|h\|_{\mathcal{K}}$,

$$|h(\theta) - h(\bar{\theta})| \leq \beta \quad \text{and} \quad |w(\theta) - w(\bar{\theta})| \leq \eta. \tag{2.8}$$

Under the stated conditions for all $\delta' \in (0, \delta]$ and $\lambda' \in (0, \lambda]$ there exists an integer N such that for any $n \geq N+1$, $\rho_n \leq \lambda'$ and $\left|\sum_{k=N+1}^{n} \rho_k \xi_k\right| \leq \delta'$. Define recursively for $j \geq N$ the sequence $\{\bar{\theta}_j\}_{j \geq N}$ as follows: $\bar{\theta}_N := \theta_N$ and for $j \geq N+1$,

$$\bar{\theta}_j = \bar{\theta}_{j-1} + \rho_j h(\theta_{j-1}). \tag{2.9}$$

By construction, for $j \geq N+1$, $\bar{\theta}_j - \theta_j = \sum_{i=N+1}^{j} \rho_i \xi_i$ which implies that $\sup_{j\geq N} |\bar{\theta}_j - \theta_j| \leq \delta'$. On the other hand, for $j \geq N+1$,

$$\bar{\theta}_j = \bar{\theta}_{j-1} + \rho_j h(\bar{\theta}_{j-1}) + \rho_j \left(h(\theta_{j-1}) - h(\bar{\theta}_{j-1}) \right), \tag{2.10}$$

and since $|\bar{\theta}_{j-1} - \theta_{j-1}| \leq \delta' \leq \delta$, Eq. (2.8) shows that $|h(\theta_{j-1}) - h(\bar{\theta}_{j-1})| \leq \beta$. Thus, Eq. (2.7) implies that, whenever $\bar{\theta}_{j-1} \in \mathcal{K}_\delta \setminus \mathcal{N}$, $w(\bar{\theta}_j) \leq w(\bar{\theta}_{j-1}) - \rho_j \varepsilon$. Now Eq. (2.8) implies that $|w(\bar{\theta}_j) - w(\bar{\theta}_{j-1})| \leq \eta$ for any $\bar{\theta}_{j-1} \in \mathcal{K}_\delta$ and $|w(\theta_j) - w(\bar{\theta}_j)| \leq \eta$ for any $\theta_j \in \mathcal{K}$, which concludes the proof. □

LEMMA 2.5. *Let ε be real constants, n be an integer and let $-\infty < a_1 < b_1 < \cdots < a_n < b_n < \infty$ be real numbers. Let $\{u_j\}$ be a bounded real sequence such that, for any $\eta > 0$ there exists an integer J such that for all $j \geq J$,*

$$u_j \leq u_{j-1} - \rho_j \varepsilon + (\eta + \rho_j \varepsilon) \mathbb{1}_A(u_{j-1}) \quad A = \bigcup_{i=1}^{n} [a_i, b_i]. \tag{2.11}$$

Then, the limit points of the sequence $\{u_j\}$ are included in A.

Proof. As $\{u_j\}$ is bounded, it has at least one limit point from the Bolzano-Weierstrass theorem. Let us denote $\check{a}$ one of these limit points; since $\{u_j\}$ is bounded and satisfies Eq. (2.11), $\check{a} \geq a_1$. Now let us proceed by contradiction and assume that there exists $l \in \{1, 2, \ldots, n\}$ such that $\check{a} \in (b_l, a_{l+1})$, with the convention that $a_{n+1} = \infty$. For any $\epsilon > 0$ sufficiently small $[\check{a} - \epsilon, \check{a} + \epsilon] \subset A^c$. Now, for any integer j and any set $B \subset \mathbb{R}$, we define:

$$\tau_B(j) = \inf\{k \geq j : u_k \in B\},$$

with the convention $\inf \emptyset = \infty$. Since $\sum_{k=1}^{\infty} \rho_k = \infty$ and $\{u_k\}_{k\geq 0}$ is bounded, Eq. (2.11) implies that for any $\eta > 0$ and $j \geq J$, $\sigma(j) := \tau_A(j) < \infty$. Note also that for $k = j, \ldots, \sigma(j)$, $u_k \leq u_j$. Since $\check{a} \in (b_l, a_{l+1})$ is a limit point, for any integer j, $\kappa(j) := \tau_{(b_l,\infty)}(j) < \infty$. Let $\eta > 0$ be such that, for any $j \geq J$, $0 < \eta < (\check{a} - \epsilon - b_l)/2$. Then for $j \geq J$, $u_{\kappa[\sigma(j)]} < (\check{a} - \epsilon + b_l)/2$ and for $k = \kappa[\sigma(j)], \ldots, \kappa(\sigma(\kappa[\sigma(j)])) - 1$, $u_k \leq u_{\kappa[\sigma(j)]}$, which implies that for any $i \geq \kappa(\sigma(J))$, $u_i \leq (\check{a} - \epsilon + b_l)/2$ and contradicts the fact that $\check{a}$ is a limit point. Now using the same type of argument, one can show that if an accumulation point $\check{a} \in [a_k, b_k]$ for some $k \in 1, \ldots, n-1$, then there cannot be any accumulation point in $[a_l, b_l]$ for $n \geq l > k$. As a consequence there cannot be an accumulation point in an interval other than $[a_k, b_k]$. □

Proof. [Theorem 2.3] We first prove that $\lim_{j\to\infty} w(\theta_j)$ exists. For any $\alpha > 0$, define the set $[w(\mathcal{L} \cap \mathcal{K})]_\alpha := \{x \in \mathbb{R} : d(x, w(\mathcal{L} \cap \mathcal{K})) \leq \alpha\}$. Since $\|w\|_{\mathcal{K}} < \infty$, $[w(\mathcal{L} \cap \mathcal{K})]_\alpha$ is a finite union of disjoint intervals of length at least equal to 2α. By Lemma 2.4, there exist positive constants δ, ε, λ, such that for any $\delta' \in (0, \delta]$, $\lambda' \in (0, \lambda]$ and $\eta > 0$, one may find an integer N and a sequence $\{\bar{\theta}_j\}_{j\geq N}$ such that,

$$\sup_{j\geq N} |\theta_j - \bar{\theta}_j| \leq \delta' \quad \text{and} \quad \sup_{j\geq N} |w(\theta_j) - w(\bar{\theta}_j)| \leq \eta$$

and

$$w(\bar{\theta}_j) \leq w(\bar{\theta}_{j-1}) - \rho_j \varepsilon + (\eta + \rho_j \varepsilon) \mathbb{1}_{[w(\mathcal{L}\cap\mathcal{K})]_\alpha}(w(\bar{\theta}_{j-1})) \quad \text{for any} \quad j \geq N+1,$$

where we have made the choice $\mathcal{N} = w^{-1}(\mathrm{int}([w(\mathcal{L} \cap \mathcal{K})]_\alpha))$ and used that $\mathbb{1}_{\mathcal{N}}(\theta) \leq \mathbb{1}_{[w(\mathcal{L}\cap\mathcal{K})]_\alpha}(w(\theta))$. By Lemma 2.5, the limit points of the sequence $\{w(\bar{\theta}_j)\}$ are in $[w(\mathcal{L} \cap \mathcal{K})]_\alpha$ and since $\sup_{j\geq N} |\theta_j - \bar{\theta}_j| \leq \delta'$

the limit points of the sequence $\{w(\theta_j)\}_{j\geq 0}$ are in $[w(\mathcal{L}\cap\mathcal{K})]_{\alpha'}$ for $\alpha' = \alpha+\eta$. Since α and η can be chosen arbitrarily small, this implies that the limit points of the sequence $\{w(\theta_j)\}_{j\geq 0}$ are included in $\bigcap_{\alpha>0}[w(\mathcal{L}\cap\mathcal{K})]_\alpha$. Because $\mathcal{L}\cap\mathcal{K}$ is a compact subset of $\mathbb{R}^{n_\theta}$ and w is continuous, $w(\mathcal{L}\cap\mathcal{K})$ is a compact subset of $\mathbb{R}$ which implies that: $w(\mathcal{L}\cap\mathcal{K}) = \bigcap_{\alpha>0}[w(\mathcal{L}\cap\mathcal{K})]_\alpha$. Thus, the limit points of $\{w(\theta_j)\}$ belong to the set $w(\mathcal{L}\cap\mathcal{K})$.

On the other hand, $\limsup_{j\to\infty}|w(\theta_j)-w(\theta_{j-1})| = 0$, which implies that the set of limit points of $\{w(\theta_j)\}$ is an interval. Because $w(\mathcal{L})$ has an empty interior, the only intervals included in $w(\mathcal{L}\cap\mathcal{K})$ are isolated points, which shows that the limit $\lim_{j\to\infty} w(\theta_j)$ exists.

We now proceed to proving that $\limsup_{j\to\infty} d(\theta_j,\mathcal{L}\cap\mathcal{K}) = 0$. Let $\mathcal{N}\subset\mathcal{K}$ be an arbitrary neighbourhood of $\mathcal{L}\cap\mathcal{K}$. From Lemma 2.4 there exist constants $\delta>0$, $\varepsilon>0$, $\lambda>0$ such that for any $\delta'\in(0,\delta]$, $\lambda'\in(0,\lambda]$ and $\eta>0$ one may find an integer N and a sequence $\{\bar\theta_j\}_{j\geq N}$ such that

$$\sup_{j\geq N}|\theta_j-\bar\theta_j|\leq\delta',\quad \sup_{j\geq N}\rho_j\leq\lambda'\quad\text{and}\quad\sup_{j\geq N}|w(\theta_j)-w(\bar\theta_j)|\leq\eta$$

and

$$w(\bar\theta_j)\leq w(\bar\theta_{j-1})-\rho_j\varepsilon+(\eta+\rho_j\varepsilon)\mathbb{1}_{\mathcal{N}}(\bar\theta_{j-1})\quad\text{for any}\quad j\geq N+1.$$

For $j\geq N$, define $\tau(j) := \inf\left\{k\geq 0, \bar\theta_{k+j}\in\mathcal{N}\right\}$. For any integer p, define $\tau^p(j) := \tau(j)\wedge p$, where $a\wedge b = \min(a,b)$.

$$w(\bar\theta_{j+\tau^p(j)})-w(\bar\theta_j) = \sum_{i=j+1}^{j+\tau^p(j)}\left\{w(\bar\theta_i)-w(\bar\theta_{i-1})\right\}\leq -\varepsilon\sum_{i=j+1}^{j+\tau^p(j)}\rho_i, \tag{2.12}$$

with the convention that, for any sequence $\{a_i\}$ and any integer l, $\sum_{i=l+1}^{l} a_i = 0$. Therefore,

$$\begin{aligned} w(\theta_{j+\tau^p(j)})-w(\theta_j) &= w(\theta_{j+\tau^p(j)})-w(\bar\theta_{j+\tau^p(j)})+w(\bar\theta_{j+\tau^p(j)})-w(\bar\theta_j)+w(\bar\theta_j)-w(\theta_j)\\ &\leq 2\eta-\varepsilon\sum_{i=j+1}^{j+\tau^p(j)}\rho_i.\end{aligned}$$

Since $\{w(\theta_j)\}$ converges, for any $\varepsilon'>0$ there exists $N'>N$ such that, for all $j\geq N'$,

$$-\varepsilon' < w(\theta_{j+\tau^p(j)})-w(\theta_j)\leq 2\eta-\varepsilon\sum_{i=j+1}^{j+\tau^p(j)}\rho_i \tag{2.13}$$

This implies that, for all $j\geq N'$ and all integer $p\geq 0$,

$$\sum_{i=j+1}^{j+\tau^p(j)}\rho_i\leq C(\varepsilon',\eta) := \varepsilon^{-1}\left(\varepsilon'+2\eta\right). \tag{2.14}$$

Since $\sum_{i=j+1}^{j+\tau(j)}\rho_i = \lim_{p\to\infty}\sum_{i=j+1}^{j+\tau^p(j)}\rho_i$ and $\sum_{i=1}^{\infty}\rho_i = \infty$, the previous relation implies that, for all $j\geq N'$, $\tau(j)<\infty$ and $\sum_{i=j+1}^{j+\tau(j)}\rho_i\leq C(\varepsilon',\eta)$. For any integer p, $\theta_{j+p}-\theta_j = \sum_{i=j+1}^{j+p}\rho_i h(\theta_{i-1})+\sum_{i=j+1}^{j+p}\rho_i\xi_i$, which implies that

$$|\theta_{j+p}-\theta_j|\leq\|h\|_{\mathcal{K}}\sum_{i=j+1}^{j+p}\rho_i+\left|\sum_{i=j+1}^{j+p}\rho_i\xi_i\right|.$$

Applying this inequality for $j \geq N'$ and $p = \tau(j)$ and using that, by definition, $\bar{\theta}_{j+\tau(j)} \in \mathcal{N}$,

$$d(\theta_j, \mathcal{N}) \leq |\bar{\theta}_{j+\tau(j)} - \theta_{j+\tau(j)}| + |\theta_{j+\tau(j)} - \theta_j| \leq \delta' + \|h\|_{\mathcal{K}}\, C(\varepsilon', \eta) + \left| \sum_{i=j+1}^{j+\tau(j)} \rho_i \xi_i \right| .$$

Since η, δ' and ε' can be chosen arbitrarily small, and $\limsup_{k\to\infty} \sup_{l \geq k} \left| \sum_{i=k}^{l} \rho_i \xi_i \right| = 0$, the latter inequality shows that $\lim_{j\to\infty} d(\theta_j, \mathcal{N}) = 0$. Since $\mathcal{N}$ is arbitrary, we thus have $\lim_{j\to\infty} d(\theta_j, \mathcal{L} \cap \mathcal{K}) = 0$. □

Note that the boundedness is here one of the required assumption. It is therefore natural to try to apply Theorem 2.2. This is what motivates the next section, where we describe a modification of the stochastic approximation algorithm which ensures that the conditions of Theorem 2.2 are satisfied. We consider here the Markov state dependent noise as it covers many applications of interest, encompasses the exogeneous scenario and as we shall see leads to general and verifiable conditions.

3. Markov state-dependent noise. In this section, we describe our stochastic approximation procedure with adaptive truncation sets and introduce the relevant notation required in the Markovian state dependent noise scenario (see [24, Section 6.6, p 159] for a detailed description and numerous examples). We first introduce a version without truncations of the algorithm in this setting (Subsection 3.2), and describe our adaptive procedure in terms of this plain algorithm in Subsection 3.1. This will prove extremely useful when proving that our procedure is stable in Section 4 and in particular Section 5.

It is assumed hereafter that the state-space X and the parameter space Θ are equipped with a countably generated σ-field, $\mathcal{B}(\mathsf{X})$ and $\mathcal{B}(\Theta)$ (and measurability will always be defined w.r.t to these σ-fields).

3.1. Non-homogeneous chain. Let $\boldsymbol{\rho} = \{\rho_n\}$ be a monotone non-increasing sequence with $\rho_0 \leq 1$, define the product space $\bar{\mathsf{X}} := \mathsf{X} \cup \{x_c\} \times \bar{\Theta} := \Theta \cup \{\theta_c\}$, where $\theta_c \notin \Theta$ and $x_c \notin \mathsf{X}$ are two arbitrary cemetery points, and define the *non-homogeneous* Markov chain $\{Y_n^{\boldsymbol{\rho}} := (X_n, \theta_n)\}$ on $\bar{\mathsf{X}} \times \bar{\Theta}$ as follows. Set $\theta_0 = \theta \in \Theta$, $X_0 = x \in \mathsf{X}$, and for $n \geq 0$,

$$\theta_{n+1} = \begin{cases} \theta_n + \rho_{n+1} H(\theta_n, X_{n+1}) & \text{and} \quad X_{n+1} \sim P_{\theta_n}(X_n, \cdot) \quad \text{if} \quad \theta_n \in \Theta, \\ \theta_c & \text{and} \quad X_{n+1} = x_c \quad \text{if} \quad \theta_n \notin \Theta, \end{cases} \tag{3.1}$$

where it is assumed that the family of Markov transition probabilities $\{P_\theta, \theta \in \Theta\}$ and the field H satisfy the following conditions :

(**A2**) For any $\theta \in \Theta$, the Markov kernel P_θ has a single stationary distribution π_θ, $\pi_\theta P_\theta = \pi_\theta$. In addition $H : \Theta \times \mathsf{X} \to \Theta$ is measurable, for all $\theta \in \Theta$, $\int_{\mathsf{X}} |H(\theta, x)| \pi_\theta(dx) < \infty$.

The existence and uniqueness of the invariant distribution can be guaranteed under classical irreducibility and recurrence conditions (see *e.g.* [26, Chapter 9,10]). We denote $h(\theta) := \int_{\mathsf{X}} H(\theta, x) \pi_\theta(dx)$ the mean-field associated to this stochastic approximation procedure and define the noise sequence $\{\xi_n = H(\theta_{n-1}, X_n) - h(\theta_{n-1})\}$. Following [5], we will often write $H_\theta(x)$ as an equivalent expression for $H(\theta, x)$, h_θ for $h(\theta)$, etc...

We denote $\mathcal{F} = \{\mathcal{F}_n, n \geq 0\}$ the natural filtration of this Markov chain, with $\mathcal{F}_n := \sigma((X_l, \theta_l), l \in \{0, \ldots, n\})$ and $\mathbb{P}^{\boldsymbol{\rho}}_{x,\theta}$ the probability measure on the canonical space $\left((\mathsf{X} \times \Theta)^{\mathbb{N}}, (\mathcal{B}(\mathsf{X}) \otimes \mathcal{B}(\Theta))^{\otimes \mathbb{N}}\right)$ generated by the non-homogeneous Markov chain $\{Y_n^{\boldsymbol{\rho}}\}$ started from the initial conditions $(X_0, \theta_0) = (x, \theta) \in \mathsf{X} \times \Theta$ and using the sequence $\boldsymbol{\rho}$. Finally it will be useful in the sequel to introduce $\{Q_{\rho_n}\}$ the sequence of transition probabilities that generates the inhomogeneous Markov chain $\{Y_n^\rho\}$, where for $\rho \geq 0$, Q_ρ is defined for any $(x, \theta) \in \mathsf{X} \times \Theta$ as,

$$Q_\rho(x, \theta; A \times B) = \int_A P_\theta(x, dy) \delta_{\theta + \rho H(\theta, y)}(B), \quad A \in \mathcal{B}(\mathsf{X}), B \in \mathcal{B}(\Theta).$$

3.2. Homogeneous chain. Let $\{\mathcal{K}_q, q \geq 0\}$ be a sequence of compact subsets of Θ such that

$$\bigcup_{q \geq 0} \mathcal{K}_q = \Theta, \quad \text{and} \quad \mathcal{K}_q \subset \operatorname{int}(\mathcal{K}_{q+1}), \quad q \geq 0, \tag{3.2}$$

where $\operatorname{int}(A)$ denotes the interior of set A. Let $\boldsymbol{\gamma} = \{\gamma_k\}$ and $\boldsymbol{\epsilon} = \{\epsilon_k\}$ be two monotone non-increasing sequences of positive numbers and let K be a subset of X. Let $\Phi : \mathsf{X} \times \Theta \to \mathsf{K} \times \mathcal{K}_0$ be a measurable function and $\phi : \mathbb{Z}^+ \to \mathbb{Z}$ be a function such that $\phi(k) > -k$ for any k. Our stochastic approximation algorithm with adaptive truncation sets is defined as an *homogeneous* Markov chain on $\mathsf{Z} := \mathsf{X} \times \Theta \times \mathbb{N} \times \mathbb{N} \times \mathbb{N}$

$$\{Z_n := (X_n, \theta_n, \kappa_n, \varsigma_n, \nu_n)\} \in \mathsf{Z}^{\mathbb{N}}, \tag{3.3}$$

with the following transition at iteration $n+1$,

- If $\nu_n = 0$, then draw $(X_{n+1}, \theta_{n+1}) \sim Q_{\gamma_{\varsigma_n}}(\Phi(X_n, \theta_n); \cdot)$; otherwise draw $(X_{n+1}, \theta_{n+1}) \sim Q_{\gamma_{\varsigma_n}}(X_n, \theta_n; \cdot)$.
- If $|\theta_{n+1} - \theta_n| \leq \epsilon_{\varsigma_n}$ and $\theta_{n+1} \in \mathcal{K}_{\kappa_n}$, then set: $\kappa_{n+1} = \kappa_n$, $\varsigma_{n+1} = \varsigma_n + 1$ and $\nu_{n+1} = \nu_n + 1$; otherwise, set $\nu_{n+1} = 0$, $\kappa_{n+1} = \kappa_n + 1$, $\varsigma_{n+1} = \varsigma_n + \phi(\nu_n)$.

In words, κ, ς and ν are counters: κ is the index of the current active truncation set; ν counts the number of iterations since the last reinitialization; ς is the current index in the sequences $\{\gamma_n\}$ and $\{\epsilon_n\}$, and therefore defines the current proposal kernel Q_γ. The event $\{\nu_n = 0\}$ means that a reinitialization occurs and the condition on ϕ ensures that the algorithm is reinitialized with a value for γ_{ς_n} smaller than that used the last time such an event occurred. This algorithm is reminiscent of the algorithm with adaptive truncation sets proposed in [11], [10]. When the current iterate wanders outside the active truncation set or when the difference between two successive values of the parameter is larger than a time-dependent threshold, then the algorithm is reinitialised with a smaller initial value of the stepsize and a larger truncation set. Various choices for the function ϕ can be considered. For example, the choice $\phi(k) = 1$ for all $k \in \mathbb{N}$ coincides with the procedure proposed in [10]: in this case $\varsigma_n = n$. Another sensible choice consists of setting $\phi(k) = 1 - k$ for all $k \in \mathbb{N}$, in which case the number of iterations between two successive reinitialisations is not taken into account.

The intuitive motivation for this modification of the original stochastic approximation recursion lies in Theorem 2.2. Indeed, in order to ensure the stability of the algorithm it is required that the sizesteps be not too large and that the average effect of the noise be small in order for the drift $h(\theta)$ to dominate, and confine the recursion to a compact set. The reprojections act as a -drastic- drift towards the center of Θ when $\{\theta_n\}$ grows too rapidly and allow one to reinitialize the algorithm with a smaller sizestep and weaker noise inside a "ring" of the type $\{\theta \in \Theta : w(\theta) \in (M_0, M_1]\}$ (M_0 and M_1 are defined in (A1)) where the drift is strictly positive. The fact that M_0 and M_1 are unknown *a priori* is the reason for the adaptive truncations, which ensure that one eventually selects $\mathcal{K}_q$ large enough in order to have $\mathcal{L} \cap \mathcal{K}_q \neq \varnothing$. As we shall see the limitation imposed on the increments of the sequence $\{\theta_n\}$ is required in order to ensure some type of homogeneity of the chain $\{\xi_n\}$, and therefore ergodicity properties of the noise sequence $\{\xi_n\}$.

In the light of this heuristic, one can naturally propose many variations on this theme. We suggest here two possible extensions. First one can suggest other strategies in order to adapt the magnitude of the sizesteps. Let $\{\gamma_{n,l}, n \geq 0, l \geq 0\}$ be an array of stepsizes. Then, when a reprojection occurs, instead of jumping forward in a unique sequence of stepsizes, it is possible to simply change the sequence of stepsizes, from say l to $l+1$. Another interesting variant of the proposed scheme consists of reinitialising the algorithm when $|\theta_n - \theta_{n-1}| > \epsilon_{\varsigma_{n-1}}$ without changing the truncation set. In either cases the proof of convergence follows using the same type of arguments as those presented in this paper.

We now introduce some further notation and briefly state our main result. For μ a probability on Z, we denote $\bar{\mathbb{P}}_\mu$ (resp. $\bar{\mathbb{E}}_\mu$) the probability (resp. the expectation) on the canonical space $(\mathsf{Z}^{\mathbb{N}}, \mathcal{B}(\mathsf{Z})^{\otimes\mathbb{N}})$ associated to the Markov chain $\{Z_n\}$ with initial distribution μ. For $z \in \mathsf{Z}$ we set $\bar{\mathbb{P}}_z := \bar{\mathbb{P}}_{\delta_z}$, $\bar{\mathbb{E}}_z := \bar{\mathbb{E}}_{\delta_z}$ and for $(x,\theta) \in \mathsf{X} \times \Theta$

$$\bar{\mathbb{P}}_{x,\theta} := \bar{\mathbb{P}}_{x,\theta,0,0,0} \quad \text{and} \quad \bar{\mathbb{E}}_{x,\theta} := \bar{\mathbb{E}}_{x,\theta,0,0,0}. \tag{3.4}$$

This probability measure depends upon the deterministic sequences $\boldsymbol{\gamma} = \{\gamma_n\}$ and $\boldsymbol{\epsilon} = \{\epsilon_n\}$; this will be implicit hereafter in order to alleviate notation. We define recursively $\{T_n, n \geq 0\}$ the sequence of successive reinitialisation times

$$T_{n+1} = \inf\{k \geq T_n + 1,\ \nu_k = 0\}, \quad \text{with} \quad T_0 = 0, \tag{3.5}$$

where by convention $\inf\{\emptyset\} = \infty$. In the following sections we prove that under (A1), some regularity conditions on the family of transition probabilities $\{P_\theta, \theta \in \Theta\}$ and the sequences $\boldsymbol{\gamma}$ and $\boldsymbol{\epsilon}$ then

$$\inf_{(x,\theta)\in\mathsf{K}\times\mathcal{K}_0} \bar{\mathbb{P}}_{x,\theta}\left(\sup_{n\geq 0} \kappa_n < \infty\right) = \inf_{(x,\theta)\in\mathsf{K}\times\mathcal{K}_0} \bar{\mathbb{P}}_{x,\theta}\left(\bigcup_{n=0}^{\infty}\{T_n = \infty\}\right) = 1,$$

i.e., the number of reinitializations of the procedure described above is finite $\bar{\mathbb{P}}_{x,\theta}$-a.e., for every $(x,\theta) \in \mathsf{K} \times \mathcal{K}_0$. Convergence will then follow using Theorem 2.3 for example.

4. Bound on $\bar{\mathbb{P}}_{x,\theta}(T_n < \infty)$.

In this section we establish in Proposition 4.2 a bound on $\bar{\mathbb{P}}_{x,\theta}(T_n < \infty)$ in terms of the fluctuations of the noise sequence of the algorithm between successive reinitializations. Let $\mathcal{K}$ be a compact subset of Θ and let $\boldsymbol{\epsilon} = \{\epsilon_n\}$ be a non-increasing sequence of positive numbers. We introduce

$$\sigma(\mathcal{K},\epsilon) = \sigma(\mathcal{K}) \wedge \nu(\boldsymbol{\epsilon}), \quad \sigma(\mathcal{K}) = \inf\{k \geq 1, \theta_k \notin \mathcal{K}\}, \quad \text{and} \quad \nu(\boldsymbol{\epsilon}) = \inf\{k \geq 1, |\theta_k - \theta_{k-1}| \geq \epsilon_k\},$$

and for a sequence $\mathbf{a} = \{a_k\}$ and an integer l, we define $\mathbf{a}^{\leftarrow l} = \{a_k^{\leftarrow l}\}$ as $a_k^{\leftarrow l} = a_{k+l}$. We now prove the following lemma, which relates the expectation of the homogeneous Markov chain $\{Z_n\}$ defined in Subsection 3.2 to the expectation of a non-homogeneous Markov chain $\{Y_n^{\boldsymbol{\rho}}\}$ defined in Subsection 3.1, for a particular $\boldsymbol{\rho}$.

Lemma 4.1. *For any $m \geq 1$, for any non-negative measurable function $\Psi_m : (\mathsf{X}\times\Theta)^m \to \mathbb{R}^+$, for any integers p and q, for any $x, \theta \in \mathsf{X}\times\Theta$,*

$$\bar{\mathbb{E}}_{x,\theta,p,q,0}\left[\Psi_m(X_1,\theta_1,\ldots,X_m,\theta_m)\mathbb{1}_{\{T_1\geq m\}}\right] = \mathbb{E}^{\gamma^{\leftarrow q}}_{\Phi(x,\theta)}\left[\Psi_m(X_1,\theta_1,\ldots,X_m,\theta_m)\mathbb{1}_{\{\sigma(\mathcal{K}_p,\epsilon^{\leftarrow q})\geq m\}}\right]. \tag{4.1}$$

Proof. We proceed by induction. Let $\Psi_1 : \mathsf{X}\times\Theta \to \mathbb{R}^+$. We notice that $\mathbb{1}_{\{T_1\geq 1\}} = 1$. This, combined with the definition of $\bar{\mathbb{E}}_{x,\theta,p,q,0}$, leads to

$$\bar{\mathbb{E}}_{x,\theta,p,q,0}[\Psi_1(X_1,\theta_1)\mathbb{1}_{\{T_1\geq 1\}}] = Q_{\gamma_q}(\Phi(x,\theta);\Psi_1), \tag{4.2}$$

and by definition

$$\mathbb{E}^{\gamma^{\leftarrow q}}_{\Phi(x,\theta)}[\Psi_1(X_1,\theta_1)\mathbb{1}_{\{\sigma(\mathcal{K}_p,\epsilon^{\leftarrow q})\geq 1\}}] = Q_{\gamma_q}(\Phi(x,\theta);\Psi_1). \tag{4.3}$$

Now assume that the property is true for some $m \geq 1$. It is sufficient to prove the induction for functions Ψ_{m+1} of the form

$$\Psi_{m+1}(x_1, \theta_1, \ldots, x_{m+1}, \theta_{m+1}) = \psi_{m+1}(x_{m+1}, \theta_{m+1}) \Psi_m(x_1, \theta_1, \ldots, x_m, \theta_m), \tag{4.4}$$

with $\psi_{m+1} : \mathsf{X} \times \Theta \to \mathbb{R}^+$. In order to alleviate notation we will often write Ψ_m (resp. ψ_m) for $\Psi_m(x_1, \theta_1, \ldots, x_m, \theta_m)$ (resp. $\psi_m(x_m, \theta_m)$) in what follows. Consider

$$\bar{\mathbb{E}}_{x,\theta,p,q,0}[\Psi_{m+1}(X_1, \theta_1, \ldots, X_{m+1}, \theta_{m+1}) \mathbb{1}_{\{T_1 \geq m+1\}}] = \\ \bar{\mathbb{E}}_{x,\theta,p,q,0}[\psi_{m+1} \Psi_m \mathbb{1}_{\{T_1 \geq m\}} \mathbb{1}_{\{\theta_m \in \mathcal{K}_{\kappa_m}\}} \mathbb{1}_{\{|\theta_m - \theta_{m-1}| < \epsilon_{\varsigma_m}\}}]. \tag{4.5}$$

Now, by definition of the stopping time T_1, we have

$$\mathbb{1}_{\{\theta_m \in \mathcal{K}_{\kappa_m}\}} \mathbb{1}_{\{|\theta_m - \theta_{m-1}| < \epsilon_{\varsigma_m}\}} \mathbb{1}_{\{T_1 \geq m\}} = \mathbb{1}_{\{\theta_m \in \mathcal{K}_{\kappa_0}\}} \mathbb{1}_{\{|\theta_m - \theta_{m-1}| < \epsilon_{\varsigma_0 + m}\}} \mathbb{1}_{\{T_1 \geq m\}} \\ = \mathbb{1}_{\{\theta_m \in \mathcal{K}_{\kappa_0}\}} \mathbb{1}_{\{|\theta_m - \theta_{m-1}| < \epsilon_{\varsigma_0 + m}\}} \mathbb{1}_{\{\sigma(\mathcal{K}_{\kappa_0}, \epsilon^{\leftarrow \varsigma_0}) \geq m\}},$$

from which we may deduce, using the induction assumption, that

$$\begin{aligned} &\bar{\mathbb{E}}_{x,\theta,p,q,0}[\Psi_{m+1} \mathbb{1}_{\{T_1 \geq m+1\}}] \\ &= \bar{\mathbb{E}}_{x,\theta,p,q,0}\left[\bar{\mathbb{E}}_{x,\theta,p,q,0}[\psi_{m+1} | X_m, \theta_m, \kappa_m, \varsigma_m, \nu_m] \, \mathbb{1}_{\{\theta_m \in \mathcal{K}_{\kappa_m}\}} \mathbb{1}_{\{|\theta_m - \theta_{m-1}| < \epsilon_{\varsigma_m}\}} \Psi_m \mathbb{1}_{\{T_1 \geq m\}}\right] \\ &= \bar{\mathbb{E}}_{x,\theta,p,q,0}\left[Q_{\gamma_{q+m+1}}(X_m, \theta_m; \psi_{m+1}) \mathbb{1}_{\{\theta_m \in \mathcal{K}_p\}} \mathbb{1}_{\{|\theta_m - \theta_{m-1}| < \epsilon_{q+m}\}} \Psi_m \mathbb{1}_{\{T_1 \geq m\}}\right] \\ &= \mathbb{E}^{\gamma^{\leftarrow q}}_{\Phi(x,\theta)}\left[Q_{\gamma_{q+m+1}}(X_m, \theta_m; \psi_{m+1}) \mathbb{1}_{\{\theta_m \in \mathcal{K}_p\}} \mathbb{1}_{\{|\theta_m - \theta_{m-1}| < \epsilon_{q+m}\}} \Psi_m \mathbb{1}_{\{\sigma(\mathcal{K}_p, \epsilon^{\leftarrow q}) \geq m\}}\right] \\ &= \mathbb{E}^{\gamma^{\leftarrow q}}_{\Phi(x,\theta)}\left[Q_{\gamma_{q+m+1}}(X_m, \theta_m; \psi_{m+1}) \Psi_m \mathbb{1}_{\{\sigma(\mathcal{K}_p, \epsilon^{\leftarrow q}) \geq m+1\}}\right] \\ &= \mathbb{E}^{\gamma^{\leftarrow q}}_{\Phi(x,\theta)}\left[\Psi_{m+1} \mathbb{1}_{\{\sigma(\mathcal{K}_p, \epsilon^{\leftarrow q}) \geq m+1\}}\right]. \end{aligned}$$

which concludes the proof. □

Define, for any compact set $\mathcal{K} \subset \Theta$, $\boldsymbol{\epsilon} = \{\epsilon_k\}$, $\boldsymbol{\rho} = \{\rho_k\}$ and $1 \leq l \leq n$ the partial sum

$$S_{l,n}(\boldsymbol{\epsilon}, \boldsymbol{\rho}, \mathcal{K}) := \mathbb{1}_{\{\sigma(\mathcal{K}, \boldsymbol{\epsilon}) \geq n\}} \sum_{k=l}^{n} \rho_k (H(\theta_{k-1}, X_k) - h(\theta_{k-1})), \tag{4.6}$$

and for any $\delta \geq 0$ and any $M \in (M_0, M_1]$,

$$A(\delta, \boldsymbol{\epsilon}, M, \boldsymbol{\rho}) := \sup_{\theta \in \mathcal{K}_0} \sup_{x \in \mathsf{K}} \left\{ \mathbb{P}^{\boldsymbol{\rho}}_{\Phi(x,\theta)} \left[\sup_{k \geq 1} |S_{1,k}(\boldsymbol{\epsilon}, \boldsymbol{\rho}, \mathcal{W}_M)| > \delta \right] + \mathbb{P}^{\boldsymbol{\rho}}_{\Phi(x,\theta)}[\nu(\boldsymbol{\epsilon}) < \sigma(\mathcal{W}_M)] \right\}, \tag{4.7}$$

where $\mathcal{K}_0$ is defined in Eq. (3.2), $\mathcal{W}_M$, M_0 and M_1 are defined in (A1).

Proposition 4.2. *Assume (A1) and that $\mathcal{K}_0 \subset \mathcal{W}_{M_0}$ (where M_0 is defined in (A1)). Then for any $M \in (M_0, M_1]$ there exist an integer n_0 and a constant $\delta_0 > 0$ such that, for any $n > n_0$, we have*

$$\sup_{(x,\theta) \in \mathsf{K} \times \mathcal{K}_0} \bar{\mathbb{P}}_{x,\theta}[T_n < \infty] \leq \prod_{l=n_0}^{n-1} \sup_{q \geq l} A(\delta_0, \boldsymbol{\epsilon}^{\leftarrow q}, M, \boldsymbol{\gamma}^{\leftarrow q}),$$

where T_n is defined in Eq. (3.5).

Proof. By Theorem 2.2, for any $M \in (M_0, M_1]$ there exist constants $\delta_0 > 0$ and $\lambda_0 > 0$ such that, for all $\theta_0 \in \mathcal{W}_{M_0}$ (where M_0 is defined in (A1)), all integer $m \geq 1$, all sequences $\{\rho_k\}$ of non-negative real numbers and all sequences $\{\xi_k\}$ of n_θ-dimensional vectors satisfying $\sup_{1\leq k\leq m} \rho_k \leq \lambda_0$ and $\sup_{1\leq k \leq m} \left|\sum_{j=1}^k \rho_j \xi_j\right| \leq \delta_0$, we have $\sup_{1\leq k\leq m} w(\theta_k) \leq M$, where $\theta_k = \theta_{k-1} + \rho_k h(\theta_k) + \rho_k \xi_k$.

Now, choose n_0 such that $\mathcal{W}_M \subset \mathcal{K}_{n_0}$ and $\gamma_{n_0} \leq \lambda_0$, where λ_0 is given in Theorem 2.2. The existence of such a n_0 follows from **(i)** for all $M \in (M_0, M_1]$, the level set $\mathcal{W}_M$ is compact and $\bigcup_{p=0}^\infty \mathcal{K}_p$ is an increasing covering of Θ **(ii)** $\gamma_p \downarrow 0$ as $p \to \infty$. We notice that for any $l \geq 0$

$$T_{l+1} = T_l + T_1 \circ \tau^{T_l}, \tag{4.8}$$

where τ denotes the shift operator on the canonical space associated to the chain $\{Z_n\}$. Consequently, by the strong Markov property

$$\bar{\mathbb{P}}_{x,\theta}[T_{l+1} < \infty] = \bar{\mathbb{E}}_{x,\theta}\left[\mathbb{1}_{\{T_l<\infty\}} \bar{\mathbb{P}}_{Z_{T_l}}\left(T_1 < \infty\right)\right]. \tag{4.9}$$

Using Lemma 4.1, we have

$$\bar{\mathbb{P}}_{Z_{T_l}}\{T_1 < \infty\}\mathbb{1}_{\{T_l<\infty\}} = C(X_{T_l}, \theta_{T_l}, l, \varsigma_{T_l})\mathbb{1}_{\{T_l<\infty\}},$$

where, for any $x, \theta \in \mathsf{X} \times \Theta$ and any integers p and q,

$$C(x, \theta, p, q) = \mathbb{P}^{\boldsymbol{\gamma}^{\leftarrow q}}_{\Phi(x,\theta)}\left(\sigma(\mathcal{K}_p, \boldsymbol{\epsilon}^{\leftarrow q}) < \infty\right).$$

Now, for $p \geq n_0$, we have $\mathcal{W}_M \subset \mathcal{K}_{n_0} \subset \mathcal{K}_p$, showing that for any $x, \theta \in \mathsf{X}\times\Theta$ and integers $p, q \geq n_0$,

$$\begin{aligned} C(x,\theta,p,q) &\leq \mathbb{P}^{\boldsymbol{\gamma}^{\leftarrow q}}_{\Phi(x,\theta)}\left[\sigma(\mathcal{W}_M) \wedge \nu(\boldsymbol{\epsilon}^{\leftarrow q}) < \infty\right] \\ &\leq \mathbb{P}^{\boldsymbol{\gamma}^{\leftarrow q}}_{\Phi(x,\theta)}\left[\sigma(\mathcal{W}_M) < \infty, \sigma(\mathcal{W}_M) \leq \nu(\boldsymbol{\epsilon}^{\leftarrow q})\right] + \mathbb{P}^{\boldsymbol{\gamma}^{\leftarrow q}}_{\Phi(x,\theta)}\left[\nu(\boldsymbol{\epsilon}^{\leftarrow q}) < \sigma(\mathcal{W}_M)\right]. \end{aligned}$$

By Theorem 2.2, for any integer $m \geq 0$ and any integer $q \geq n_0$, we have

$$\{\sigma(\mathcal{W}_M) = m, m \leq \nu(\boldsymbol{\epsilon}^{\leftarrow q})\} \subset \left\{\sup_{k\in\{1,\ldots,m\}} |S_{1,k}(\boldsymbol{\epsilon}^{\leftarrow q}, \boldsymbol{\gamma}^{\leftarrow q}, \mathcal{W}_M)| > \delta_0\right\},$$

which implies that for any $x, \theta \in \mathsf{X}\times\Theta$, any $l \geq n_0$ and any $q \geq n_0$

$$C(x,\theta,l,q) \leq \mathbb{P}^{\boldsymbol{\gamma}^{\leftarrow q}}_{\Phi(x,\theta)}\left(\sup_{k\geq 1}|S_{1,k}(\boldsymbol{\epsilon}^{\leftarrow q}, \boldsymbol{\gamma}^{\leftarrow q}, \mathcal{W}_M)| > \delta_0\right) + \mathbb{P}^{\boldsymbol{\gamma}^{\leftarrow q}}_{\Phi(x,\theta)}(\nu(\boldsymbol{\epsilon}^{\leftarrow q}) < \sigma(\mathcal{W}_M)) \leq A(\delta_0, \boldsymbol{\epsilon}^{\leftarrow q}, M, \boldsymbol{\gamma}^{\leftarrow q}).$$

Combining the results above, we have, noting that $\varsigma_{T_l} \geq l$,

$$\bar{\mathbb{P}}_{Z_{T_l}}\left[T_1 < \infty\right]\mathbb{1}_{\{T_l<\infty\}} \leq \quad A(\delta_0, \boldsymbol{\epsilon}^{\leftarrow \varsigma_{T_l}}, M, \boldsymbol{\gamma}^{\leftarrow \varsigma_{T_l}})\mathbb{1}_{\{T_l<\infty\}} \leq \quad \sup_{q\geq l} A(\delta_0, \boldsymbol{\epsilon}^{\leftarrow q}, M, \boldsymbol{\gamma}^{\leftarrow q})\mathbb{1}_{\{T_l<\infty\}};$$

the proof now follows from a straightforward backward induction using Eq. (4.9) for $l = n_0, \ldots, n-1$ and $n > n_0$. □

COROLLARY 4.3. *Assume (A1) and that $\mathcal{K}_0 \subset \mathcal{W}_{M_0}$ (where M_0 is defined in (A1)). Then for any $M \in (M_0, M_1]$ and $n \geq n_0$, there exists a constant $C < \infty$ such that for any $m \geq n$,*

$$\bar{\mathbb{P}}_{x,\theta}\left[\sup_{k\geq 1}\kappa_k \geq m\right] \leq C\left(\sup_{q\geq n} A(\delta_0, \boldsymbol{\epsilon}^{\leftarrow q}, M, \boldsymbol{\gamma}^{\leftarrow q})\right)^m,$$

where $\{\kappa_k\}$ is the counter corresponding to the number of reinitialisations defined in Eq. (3.3)*.*

Proof. We have

$$\left\{\sup_{k\geq 1} \kappa_k \geq m\right\} \subset \{T_m < \infty\}$$

and consequently

$$\bar{\mathbb{P}}_{x,\theta}\left(\sup_{k\geq 1} \kappa_k \geq m\right) \leq \bar{\mathbb{P}}_{x,\theta}\left(T_m < \infty\right) \leq \prod_{l=n_0}^{m-1} \sup_{q\geq l} A(\delta_0, \boldsymbol{\epsilon}^{\leftarrow q}, M, \boldsymbol{\gamma}^{\leftarrow q})$$
$$\leq \prod_{l=n_0}^{n-1} \sup_{q\geq n_0} A(\delta_0, \boldsymbol{\epsilon}^{\leftarrow q}, M, \boldsymbol{\gamma}^{\leftarrow q}) \prod_{l=n}^{m-1} \sup_{q\geq n} A(\delta_0, \boldsymbol{\epsilon}^{\leftarrow q}, M, \boldsymbol{\gamma}^{\leftarrow q}) \leq C \left(\sup_{q\geq n} A(\delta_0, \boldsymbol{\epsilon}^{\leftarrow q}, M, \boldsymbol{\gamma}^{\leftarrow q})\right)^m . \quad (4.10)$$

□

In the next section we derive conditions on the family of Markov kernels $\{P_\theta, \theta \in \Theta\}$ and on the sequences $\boldsymbol{\epsilon} = \{\epsilon_k\}$ and $\boldsymbol{\gamma} = \{\gamma_k\}$ which ensure that $\sup_{q\geq n} A(\delta_0, \boldsymbol{\epsilon}^{\leftarrow q}, M, \boldsymbol{\gamma}^{\leftarrow q}) < 1$ for n large enough. It should be emphasized here that this involves studying only the fluctuations of the canonical "interprojections" processes, *i.e.* $\{Y_n^{\boldsymbol{\rho}}\}$ for $\boldsymbol{\rho} = \boldsymbol{\gamma}^{\leftarrow \varsigma_{T_0}}, \boldsymbol{\gamma}^{\leftarrow \varsigma_{T_1}}, \boldsymbol{\gamma}^{\leftarrow \varsigma_{T_2}}, \ldots$

5. Control of the fluctuations. Our aim is now to find a bound for $A(\delta, \boldsymbol{\epsilon}, M, \boldsymbol{\rho})$ defined in Eq. (4.7), which requires the following conditions to hold. Define, for $V : \mathsf{X} \to [1, \infty)$ and $g : \mathsf{X} \to \mathbb{R}^{n_\theta}$ the norm

$$\|g\|_V = \sup_{x\in\mathsf{X}} \frac{|g(x)|}{V(x)}. \quad (5.1)$$

Consider the following assumptions

(**A3**) For any $\theta \in \Theta$, the Poisson equation $g - P_\theta g = H_\theta - \pi_\theta(H_\theta)$ has a solution g_θ. There exist a function $W : \mathsf{X} \to [1, \infty]$ such that $\{x \in \mathsf{X}, W(x) < \infty\} \neq \emptyset$, constants $\alpha \in (0, 1]$, $p \geq 2$ such that for any compact subset $\mathcal{K} \subset \Theta$,

(i)

$$\sup_{\theta\in\mathcal{K}} \|H_\theta\|_W < \infty, \quad (5.2)$$
$$\sup_{\theta\in\mathcal{K}} (\|g_\theta\|_W + \|P_\theta g_\theta\|_W) < \infty, \quad (5.3)$$
$$\sup_{(\theta,\theta')\in\mathcal{K}} |\theta - \theta'|^{-\alpha} \{\|g_\theta - g_{\theta'}\|_W + \|P_\theta g_\theta - P_{\theta'} g_{\theta'}\|_W\} < \infty. \quad (5.4)$$

(ii) there exist constants $\{C_k, k \geq 0\}$ such that, for any $k \in \mathbb{N}$, for any sequence $\boldsymbol{\rho} = \{\rho_k\}$ and for any $x \in \mathsf{X}$,

$$\sup_{\theta\in\mathcal{K}} \mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}[W^p(X_k)\mathbb{1}_{\{\sigma(\mathcal{K})\geq k\}}] \leq C_k W^p(x), \quad (5.5)$$

(iii) there exist $\epsilon > 0$ and a constant C such that for any sequence $\boldsymbol{\rho} = \{\rho_k\}$ and for any $x \in \mathsf{X}$,

$$\sup_{\theta\in\mathcal{K}} \mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}[W^p(X_k)\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu_\epsilon\geq k\}}] \leq C W^p(x). \quad (5.6)$$

where

$$\nu_\epsilon = \inf\{k \geq 1, |\theta_k - \theta_{k-1}| > \epsilon\}. \quad (5.7)$$

Assumption (A3) states the existence and the regularity of the solutions of Poisson's equation for the family of transition kernels $\{P_\theta, \theta \in \Theta\}$. The conditions stated above are non primitive; a set of more tractable conditions implying (A3) are given in Section 6. Poisson's equation has proven to be fundamental to the analysis of additive functionals of Markov chains, in particular for establishing limit theorems such as the (functional) central limit theorem (see *e.g.* [5], [27], [26, chapter 17], [19], [16]); The existence of solutions to Poisson's equation is well established for geometrically ergodic Markov chains (see [27], [26, chapter 17]); it has been more recently proven under assumptions weaker than geometric ergodicity (see [19, Theorem 2.3]); the regularity of the solution of Poisson's equation has been studied, under various ergodicity and regularity conditions on the mapping $\theta \mapsto P_\theta$ by [5], [4]. We stress here on the fact that the function W is global but that the bounds in Eqs. (5.2), (5.3), (5.4), (5.5) and (5.6) depend on the particular compact $\mathcal{K}$ under consideration. We have

LEMMA 5.1. *Assume (A3). Let $\mathcal{K}$ be a compact subset of Θ and $s \in \mathbb{N}$. There exists a constant C such that for any sequence $\boldsymbol{\epsilon} = \{\epsilon_k\}$ satisfying $0 < \epsilon_k \leq \epsilon$ for all $k \geq s$ (where ϵ is defined in (A3-iii)), for any sequence $\boldsymbol{\rho} = \{\rho_k\}$ and for any $x \in \mathsf{X}$,*

$$\sup_{\theta \in \mathcal{K}} \sup_{k \geq 0} \mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[W^p(X_k) \mathbb{1}_{\{\sigma(\mathcal{K}, \boldsymbol{\epsilon}) \geq k\}}\right] \leq C W^p(x).$$

Proof. Under (A3), there exists a constant C such that, for any sequence $\boldsymbol{\rho} = \{\rho_k\}$ and any $x \in \mathsf{X}$ we have

$$\sup_{\theta \in \mathcal{K}} \mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}[W^p(X_k) \mathbb{1}_{\{\sigma(\mathcal{K}) \wedge \nu_\epsilon \geq k\}}] \leq C W^p(x).$$

where ν_ϵ is defined in (5.7). For any sequence $\boldsymbol{\epsilon} = \{\epsilon_k\}$ such that $\epsilon_k \leq \epsilon$ for any $k \geq s$,

$$\begin{aligned}
&\mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}[W^p(X_{k+s}) \mathbb{1}_{\{\sigma(\mathcal{K}) \wedge \nu(\boldsymbol{\epsilon}) \geq k+s\}}] \\
&\quad = \mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[\mathbb{E}^{\boldsymbol{\rho}^{\leftarrow s}}_{X_s, \theta_s}[W^p(X_k) \mathbb{1}_{\{\sigma(\mathcal{K}) \wedge \nu(\boldsymbol{\epsilon}^{\leftarrow s}) \geq k\}}] \mathbb{1}_{\{\sigma(\mathcal{K}) \wedge \nu(\boldsymbol{\epsilon}) \geq s\}}\right] \\
&\quad \leq \mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[\sup_{\theta \in \mathcal{K}} \mathbb{E}^{\boldsymbol{\rho}^{\leftarrow s}}_{X_s, \theta}[W^p(X_k) \mathbb{1}_{\{\sigma(\mathcal{K}) \wedge \nu_\epsilon \geq k\}}] \mathbb{1}_{\{\sigma(\mathcal{K}) \geq s\}}\right] \\
&\quad \leq C \mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}[W^p(X_s) \mathbb{1}_{\{\sigma(\mathcal{K}) \geq s\}}],
\end{aligned}$$

and the proof is concluded by (A3). □

PROPOSITION 5.2. *Assume (A3). Let $\mathcal{K}$ be a compact subset of Θ and let $\boldsymbol{\rho} = \{\rho_k\}$ and $\boldsymbol{\epsilon} = \{\epsilon_k\}$ be two non-increasing sequences of positive numbers such that $\lim_{k \to \infty} \epsilon_k = 0$. Then, for p as defined in (A3),*

1. *There exists a constant C such that, for any $(x, \theta) \in \mathsf{X} \times \mathcal{K}$ and any integer l, any $\delta > 0$*

$$\mathbb{P}^{\boldsymbol{\rho}}_{x,\theta}\left(\sup_{n \geq l} |S_{l,n}(\boldsymbol{\epsilon}, \boldsymbol{\rho}, \mathcal{K})| \geq \delta\right) \leq C \delta^{-p} \left\{\left(\sum_{k=l}^{\infty} \rho_k^2\right)^{p/2} + \left(\sum_{k=l}^{\infty} \rho_k \epsilon_k^\alpha\right)^p\right\} W^p(x). \tag{5.8}$$

2. *There exists a constant C such that, for any $(x, \theta) \in \mathsf{X} \times \mathcal{K}$,*

$$\mathbb{P}^{\boldsymbol{\rho}}_{x,\theta}(\nu(\boldsymbol{\epsilon}) < \sigma(\mathcal{K})) \leq C \left\{\sum_{k=1}^{\infty} (\epsilon_k^{-1} \rho_k)^p\right\} W^p(x). \tag{5.9}$$

The proof is in Appendix A. We finally need a condition on the stepsize sequences which will ensure that $A(\delta, \boldsymbol{\epsilon}^{\leftarrow q}, M, \boldsymbol{\rho}^{\leftarrow q}) \to 0$ when $q \to \infty$.

(**A4**) The sequences $\boldsymbol{\gamma} = \{\gamma_k\}$ and $\boldsymbol{\epsilon} = \{\epsilon_k\}$ are non-increasing, positive, and satisfy, $\sum_{k=0}^{\infty} \gamma_k = \infty$, $\lim_{k\to\infty} \epsilon_k = 0$ and

$$\sum_{k=1}^{\infty} \left\{ \gamma_k^2 + \gamma_k \epsilon_k^{\alpha} + (\epsilon_k^{-1}\gamma_k)^p \right\} < \infty,$$

where p and α are defined in (A3).

For instance, we may assume that $\sum_{k\geq 0} \gamma_k = \infty$ and $\sum_{k=0}^{\infty} \gamma_k^{\delta} < \infty$ for some $1 < \delta \leq p(1+\alpha)/(p+\alpha)$. Then, (A4) is verified by setting $\epsilon_k = C\gamma_k^{\eta}$ for some constant C and some η such that

$$\frac{\delta - 1}{\alpha} \leq \eta \leq \frac{p-\delta}{p}.$$

It is now straightforward to establish the following results:

Proposition 5.3. *Assume (A3) and (A4). Then, for any subset* $\mathsf{K} \subset \mathsf{X}$ *such that* $\sup_{x\in\mathsf{K}} W(x) < \infty$ *any* $M \in (M_0, M_1]$ *and any* $\delta > 0$ *we have* $\lim_{k\to\infty} A(\delta, \boldsymbol{\epsilon}^{\leftarrow k}, M, \boldsymbol{\gamma}^{\leftarrow k}) = 0$*, where* $A(\delta, \boldsymbol{\epsilon}, M, \boldsymbol{\rho})$ *is given by Eq. (4.7).*

We may now summarize the discussion above to obtain the following stability result.

Theorem 5.4. *Assume (A1) to (A4). Then, for any subset* $\mathsf{K} \subset \mathsf{X}$ *such that* $\sup_{x\in\mathsf{K}} W(x) < \infty$, $\mathcal{K}_0 \subset \mathcal{W}_{M_0}$ *(where* M_0 *is defined in (A1)) and any* $\rho \in (0,1)$*, there exists a constant* $C < \infty$ *such that, for all* $(x,\theta) \in \mathsf{X} \times \Theta$*,*

$$\bar{\mathbb{P}}_{x,\theta}\left[\sup_{n\geq 1} \kappa_n \geq k\right] \leq C\rho^k.$$

Hence, under the stated conditions, the tail probability of the number of reinitialization decreases faster than any exponential and $\sup_{n\geq 1} \kappa_n$ is finite $\bar{\mathbb{P}}_{x,\theta}$-a.s. Combining this result with Theorem 2.3, it is possible to obtain the following global convergence result.

Theorem 5.5. *Assume (A1) to (A4). Let* $\mathsf{K} \subset \mathsf{X}$ *be such that* $\sup_{x\in\mathsf{K}} W(x) < \infty$ *and that* $\mathcal{K}_0 \subset \mathcal{W}_{M_0}$ *(where* M_0 *is defined in (A1)), and let* $\{Z_n\}$ *be as defined by Eq. (3.3). Then, for all* $(x,\theta) \in \mathsf{X} \times \Theta$*, we have* $\lim_{k\to\infty} d(\theta_k, \mathcal{L}) = 0$*,* $\bar{\mathbb{P}}_{x,\theta}$*-a.s.*

Proof. Define, for $k \geq 1$,

$$B_k = \limsup_{l\to\infty} \sup_{T_{k-1}+l\leq n} \left| \mathbb{1}_{\{n<T_k\}} \sum_{j=T_{k-1}+l}^{n} \gamma_{\varsigma_j} (H(\theta_{j-1}, X_j) - h(\theta_{j-1})) \right| \mathbb{1}_{\{T_{k-1}<\infty\}},$$

where ς_j and T_k are defined in Section 3. We first show that, for any k and any $\delta > 0$, $\bar{\mathbb{P}}_{x,\theta}(|B_k| \geq \delta) = 0$ for all $(x,\theta) \in \mathsf{X} \times \Theta$. We have, by the strong Markov property and the definition Eq. (4.6) that for $l \geq 1$

$$\bar{\mathbb{P}}_{x,\theta}\left(\sup_{T_{k-1}+l\leq n} \left| \mathbb{1}_{\{n<T_k\}} \sum_{j=T_{k-1}+l}^{n} \gamma_{\varsigma_j} (H(\theta_{j-1}, X_j) - h(\theta_{j-1})) \right| \mathbb{1}_{\{T_{k-1}<\infty\}} \geq \delta \right) \\ \leq \bar{\mathbb{E}}_{x,\theta}\left\{ C_l(X_{T_{k-1}}, \theta_{T_{k-1}}, \delta, \mathcal{K}_{k-1}, \varsigma_{T_{k-1}}) \mathbb{1}_{\{T_{k-1}<\infty\}} \right\},$$

where for any $x, \theta \in \mathsf{X} \times \Theta$, any $\delta > 0$, any set $\mathcal{K} \subset \Theta$ and any integer q,

$$C_l(x, \theta, \delta, \mathcal{K}, q) = \mathbb{P}^{\boldsymbol{\gamma}^{\leftarrow q}}_{\Phi(x,\theta)} \left(\sup_{n \geq l} |S_{l,n}(\boldsymbol{\epsilon}^{\leftarrow q}, \boldsymbol{\gamma}^{\leftarrow q}, \mathcal{K})| \geq \delta \right).$$

By Proposition 5.2, for any compact subset $\mathcal{K}$, there exists a constant C such that, for all $q \geq 0$,

$$\sup_{(x,\theta) \in \mathsf{X} \times \Theta} C_l(x, \theta, \delta, \mathcal{K}, q) \leq C \delta^{-p} \left\{ \left(\sum_{j=l}^{\infty} \gamma_j^2 \right)^{p/2} + \left(\sum_{j=l}^{\infty} \gamma_j \epsilon_j^{\alpha} \right)^{p} \right\},$$

which implies that, for all $k \geq 0$, $\bar{\mathbb{P}}_{x,\theta}(|B_k| \geq \delta) = 0$. Corollary 4.3 and Proposition 5.3 show that, for all $(x, \theta) \in \mathsf{K} \times \mathcal{K}_0$, $\kappa = \sup_k \kappa_k < \infty$ $\bar{\mathbb{P}}_{x,\theta}$-a.e. Set, for $k \geq 0$, $\bar{\theta}_k = \theta_{k+T_{\kappa-1}}$, $\bar{\gamma}_k = \gamma_{k+\varsigma_{T_{\kappa-1}}}$ and

$$\xi_k = H(\bar{\theta}_{k-1}, X_{k+T_{\kappa-1}}) - h(\bar{\theta}_{k-1}), \quad k \geq 1.$$

Then, $\bar{\theta}_k = \bar{\theta}_{k-1} + \bar{\gamma}_k h(\bar{\theta}_{k-1}) + \bar{\gamma}_k \xi_k$ and, since $T_\kappa = \infty$, for all $(x, \theta) \in \mathsf{K} \times \mathcal{K}_0$,

$$\limsup_{l \to \infty} \sup_{n \geq l} \left| \sum_{k=l}^{n} \bar{\gamma}_k \xi_k \right| = B_\kappa = 0, \quad \bar{\mathbb{P}}_{x,\theta} - \text{a.e.}.$$

The proof follows from Theorem 2.3. □

6. Drift conditions. In this section, we give conditions which imply (A3) in terms of a minorisation of the Markov kernel on a small set and a drift condition toward this small set (see [26] for the definitions and main results). Denote, for $V : \mathsf{X} \to [1, \infty)$, $\mathcal{L}_V := \{g : \mathsf{X} \to \mathbb{R}^{n_\theta}, \sup_{x \in \mathsf{X}} \|g\|_V < \infty\}$ where $\|\cdot\|_V$ is defined in Eq. (5.1).

(**DRI**) For any $\theta \in \Theta$, P_θ is ψ-irreducible and aperiodic [1]. In addition there exist a function $V : \mathsf{X} \to [1, \infty)$, constants $p \geq 2$ and $\beta \in [0, 1]$ such that for any compact subset $\mathcal{K} \subset \Theta$,

(DRI1) there exist an integer m, constants $0 < \lambda < 1$, b, κ, $\delta > 0$ and a probability measure ν such that

$$\sup_{\theta \in \mathcal{K}} P_\theta^m V^p(x) \leq \lambda V^p(x) + b \mathbb{1}_{\mathsf{C}}(x), \tag{6.1}$$

$$\sup_{\theta \in \mathcal{K}} P_\theta V^p(x) \leq \kappa V^p(x) \qquad \forall x \in \mathsf{X}, \tag{6.2}$$

$$\inf_{\theta \in \mathcal{K}} P_\theta^m(x, A) \geq \delta \nu(A) \qquad \forall x \in \mathsf{C}, \quad \forall A \in \mathcal{B}(\mathsf{X}). \tag{6.3}$$

(DRI2) there exists C such that, for all $x \in \mathsf{X}$,

$$\sup_{\theta \in \mathcal{K}} |H_\theta(x)| \leq C V(x),$$

$$\sup_{(\theta,\theta') \in \mathcal{K}} |\theta - \theta'|^{-\beta} |H_\theta(x) - H_{\theta'}(x)| \leq C V(x).$$

(DRI3) there exists C such that, for all $(\theta, \theta') \in \mathcal{K} \times \mathcal{K}$,

$$\|P_\theta g - P_{\theta'} g\|_V \leq C \|g\|_V |\theta - \theta'|^\beta \qquad \forall g \in \mathcal{L}_V, \tag{6.4}$$

$$\|P_\theta g - P_{\theta'} g\|_{V^p} \leq C \|g\|_{V^p} |\theta - \theta'|^\beta, \qquad \forall g \in \mathcal{L}_{V^p}. \tag{6.5}$$

[1] We use in this article the standard terminology and the notations introduced in [26, Chapter 4,5]

Assumption (DRI1) is classical in the Markov chain literature; it implies the existence of a stationary distribution π_θ for all $\theta \in \Theta$ and V^p-uniform ergodicity, *i.e.* for each $\theta \in \Theta$ there exist constants $C_\theta < \infty$ and $\rho_\theta \in [0,1)$, such that for any function $f \in \mathcal{L}_{V^p}$ and any integer $k > 0$

$$\|P_\theta^k f - \pi_\theta(f)\|_{V^p} \leq C_\theta \rho_\theta^k \|f\|_{V^p} .$$

Note that the constants C_θ and ρ_θ may be bounded over the compact sets of Θ, *i.e.* for each $\mathcal{K} \subset \Theta$, there exists $\bar{C} < \infty$ and $\bar{\rho} \in [0,1)$, such that $\sup_{\theta\in\mathcal{K}} C_\theta \leq \bar{C}$ and $\sup_{\theta\in\mathcal{K}} \rho_\theta \leq \bar{\rho}$. The regularity of the kernels $\theta \to P_\theta$ expressed in V and V^p norm is naturally less classical. The main result of this section is:

Proposition 6.1. *Assume (DRI). Then (A2) and (A3) are satisfied and for any* $0 < \alpha < \beta$,

$$\sup_{(\theta,\theta')\in\mathcal{K}\times\mathcal{K}} |\theta - \theta'|^{-\alpha} \, |h(\theta) - h(\theta')| < \infty. \tag{6.6}$$

The proof is in Appendix B.

7. Controlled MCMC algorithm. Markov chain Monte Carlo (MCMC), introduced by [25], is a popular computational method for generating samples from virtually any distribution π defined on a space $\mathsf{X} \subset \mathbb{R}^{n_x}$ (for some integer n_x). The method consists of simulating an ergodic Markov chain $\{X_n, n \geq 0\}$ on X with transition probability P such that π is a *stationary* distribution for this chain, *i.e* $\pi P = \pi$. The generated samples can then be used in order to estimate integrals of the type

$$\pi(\psi) := \int_{\mathsf{X}} \psi(x)\, \pi(dx),$$

for some π-integrable function $\psi : \mathsf{X} \to \mathbb{R}^{n_\psi}$, for some integer n_ψ, using estimators of the type

$$S_n(\psi) = \frac{1}{n} \sum_{k=1}^{n} \psi(X_k). \tag{7.1}$$

In general the transition probability P of the Markov chain depends on some tuning parameter, say θ, defined on some space $\Theta \subset \mathbb{R}^{n_\theta}$ for some integer n_θ, and the convergence properties of the Monte Carlo averages in Eq. (7.1) might highly depend on a proper choice of this parameter.

We illustrate this here with the classical Metropolis-Hastings (MH) update, but it should be stressed at this point that the results presented in this paper apply to much more general settings. The MH algorithm requires the choice of a *proposal distribution* q. In order to simplify the discussion, we will here assume that π and q admit densities with respect to the Lebesgue measure λ^{Leb}, denoted with an abuse of notation π and q hereafter. The rôle of the distribution q consists of proposing potential transitions y for the Markov chain $\{X_n\}$. Given that the chain is currently at x, a candidate y is accepted with probability $\alpha(x,y)$ defined as

$$\alpha(x,y) = \begin{cases} 1 \wedge \frac{\pi(y)}{\pi(x)} \frac{q(y,x)}{q(x,y)} & \text{if } \pi(x) q(x,y) > 0 \\ 1 & \text{otherwise,} \end{cases}$$

where $a \wedge b := \min(a,b)$. Otherwise it is rejected and the Markov chain stays at its current location x. The transition kernel P of this Markov chain takes the form for $x, A \in \mathsf{X} \times \mathcal{B}(\mathsf{X})$

$$P(x,A) = \int_A \alpha(x,y) q(x,y) \lambda^{\mathrm{Leb}}(dy) + \mathbb{1}_A(x) \int_{\mathsf{X}} (1 - \alpha(x,y)) q(x,y) \lambda^{\mathrm{Leb}}(dy). \tag{7.2}$$

The Markov chain P is reversible with respect to π, and therefore admits π as invariant distribution. Conditions on the proposal distribution q that guarantee irreducibility and positive recurrence are mild and many satisfactory choices are possible.

7.1. Symmetric random walk MH. We focus here on the symmetric increments random-walk MH algorithm (hereafter SRWM), which corresponds to the case where $q(x,y) = q(x-y)$ for some symmetric probability density q. Other examples are considered in [2]. The transition kernel of the SRWM algorithm is then given for $x, A \in \mathsf{X} \times \mathcal{B}(\mathsf{X})$ by

$$P_q^{\text{SRW}}(x,A) = \int_{A-x} \left(1 \wedge \frac{\pi(x+z)}{\pi(x)}\right) q(z)\,\lambda^{\text{Leb}}(dz) + \mathbb{1}_A(x) \int_{\mathsf{X}-x} \left(1 - \left(1 \wedge \frac{\pi(x+z)}{\pi(x)}\right)\right) q(z)\,\lambda^{\text{Leb}}(dz), \quad x \in \mathsf{X}, A \in \mathcal{B}(\mathsf{X}), \tag{7.3}$$

where $A - x := \{z \in \mathsf{X}, x+z \in A\}$. A classical choice for the proposal distribution is $q = \phi_{0,\Gamma}$, where $\phi_{\mu,\Gamma}$ is the density of a multivariate normal distribution with mean μ and covariance matrix Γ. We will later on refer to this algorithm as the N-SRWM. It is well known that either too small or too large a covariance matrix will result in highly positively correlated Markov chains, and therefore estimators $S_n(\psi)$ with large variance. In practice this covariance matrix Γ is determined by trial and error, using several realisations of the Markov chain. This hand-tuning requires some expertise and can be time-consuming.

In order to circumvent this problem, in the context of the N-SRWM update described above, [20] have proposed to "learn Γ on the fly". Their algorithm can be summarized as follows (see [20])

$$\begin{aligned} \mu_{n+1} &= \mu_n + \gamma_{n+1}(X_{n+1} - \mu_n) \qquad\qquad n \geq 0 \\ \Gamma_{n+1} &= \Gamma_n + \gamma_{n+1}((X_{n+1} - \mu_n)(X_{n+1} - \mu_n)^{\mathsf{T}} - \Gamma_n), \end{aligned} \tag{7.4}$$

where

- X_{n+1} is drawn from $P_{\theta_n}(X_n, \cdot)$, where for $\theta = (\mu, \Gamma)$, $P_\theta := P^{\text{SRW}}_{\phi_{0,\lambda\Gamma}}$ with $\lambda > 0$ a constant scaling factor depending only on the dimension of the state-space n_x and kept constant across the iterations,
- $\gamma = \{\gamma_n\}$ is a non-increasing sequence of positive stepsizes such that $\sum_{n=1}^{\infty} \gamma_n = \infty$ and $\sum_{n=1}^{\infty} \gamma_n^{1+\delta} < \infty$ for some $\delta > 0$ ([20] have suggested the choice $\gamma_n = 1/n$).

It was realised in [3] that such a scheme is a particular case of a more general framework akin to stochastic control, combined with the use of the Robbins-Monro procedure. More precisely, let $\theta = (\mu, \Gamma) \in \Theta$, where $\Theta := \mathbb{R}^{n_x} \times \mathcal{C}_+^{n_x}$ and $\mathcal{C}_+^{n_x}$ is the cone of positive $n_x \times n_x$ matrices, then

$$H(x;\theta) = (x - \mu, (x-\mu)(x-\mu)^{\mathsf{T}} - \Gamma)^{\mathsf{T}}. \tag{7.5}$$

With this notation, the recursion in Eq. (7.4) may be written in the standard Robbins-Monro form as

$$\theta_{n+1} = \theta_n + \gamma_{n+1} H(X_{n+1}, \theta_n), \quad n \geq 0, \tag{7.6}$$

with $X_{n+1} \sim P_{\theta_n}(X_n, \cdot)$. For the present example, assuming that $\int_{\mathsf{X}} |x|^2 \pi(dx) < \infty$, one can easily check that

$$h(\theta) = \int_{\mathsf{X}} H(x,\theta)\pi(dx) = (\mu_\pi - \mu, (\mu_\pi - \mu)(\mu_\pi - \mu)^{\mathsf{T}} + \Gamma_\pi - \Gamma)^{\mathsf{T}}, \tag{7.7}$$

with μ_π and Γ_π the mean and covariance of the target distribution. It is assumed in the sequel that Γ_π is positive definite. We now analyze the corresponding homogeneous Markov chain $\{Z_n, n \geq 0\}$ as defined in Section 3, *i.e.* prove under mild conditions on π that (A1)-(A3) are satisfied.

7.2. Condition (A1). In the algorithm described above the parameter estimates μ_n and Γ_n take the form of maximum likelihood estimates under the i.i.d. multivariate normal model. It therefore comes as no surprise if the appropriate Lyapunov function is

$$w(\mu, \Gamma) = -\int_{\mathsf{X}} \log\left(\frac{\pi(x)}{\phi_{0,\Gamma}(x)}\right)\pi(dx), \tag{7.8}$$

the Kullback-Leibler divergence between the target density π and a normal density $\phi_{0,\Gamma}$.

PROPOSITION 7.1. *Let h be as defined in Eq.* (7.7) *where π satisfies (M). Then (A1) is satisfied with w as in Eq. 7.8. Furthermore $\mathcal{L}$ is reduced to a single point, $\theta_\pi := (\mu_\pi, \Gamma_\pi)$.*

Proof. h is naturally continuous (and as we shall see later is in fact Lipschitz continuous under (DRI) from Proposition 6.1). Now w is equal, up to multiplicative and additive constants, to

$$\log\det\Gamma + (\mu - \mu_\pi)^{\mathsf{T}}\Gamma^{-1}(\mu - \mu_\pi) + \mathrm{Tr}(\Gamma^{-1}\Gamma_\pi). \tag{7.9}$$

Using straightforward algebra, one can show that there exists a constant $C > 0$ such that

$$C\Big\langle \nabla w(\mu,\Gamma), h(\mu,\Gamma)\Big\rangle = -2(\mu-\mu_\pi)^{\mathsf{T}}\Gamma^{-1}(\mu-\mu_\pi) \\ - \mathrm{Tr}(\Gamma^{-1}(\Gamma-\Gamma_\pi)\Gamma^{-1}(\Gamma-\Gamma_\pi)) - \big((\mu-\mu_\pi)^{\mathsf{T}}\Gamma^{-1}(\mu-\mu_\pi)\big)^2, \tag{7.10}$$

that is $\Big\langle \nabla w(\theta), h(\theta)\Big\rangle \leq 0$ for any $\theta = (\mu, \Gamma) \in \Theta$, with equality if and only if $\Gamma = \Gamma_\pi$ and $\mu = \mu_\pi$. As $w(\Theta) = [w(\mu_\pi, \Gamma_\pi), \infty)$ and w is continuous, any $w(\mu_\pi, \Gamma_\pi) < M_0 < M_1 < \infty$ satisfy (A1-i) and (A1-ii), and (A1-iii) is automatically satisfied. Now as the set of stationary points $\mathcal{L}$ is reduced to a single point, (A1-iv) is also satisfied. □

7.3. Condition (A3). In order to check (A3) in this case, we check (DRI). The geometric ergodicity of the RWMH kernel has been studied by [29] and refined in [22]; the regularity of the RWMH has, to the best of our knowledge, not been considered in the literature. The geometric ergodicity of the RWMH kernel mainly depends on the tail properties of the target distribution π. We will therefore restrict our discussion to target distributions that satisfy the following set of conditions. These are not minimal but easy to check in practice (see [22] for details).

(**M**) The probability density π has the following properties:
- (i) It is bounded, bounded away from zero on every compact set and continuously differentiable.
- (ii) It is super-exponential, *i.e.*
$$\lim_{|x|\to+\infty} \left\langle \frac{x}{|x|}, \nabla \log \pi(x) \right\rangle = -\infty.$$
- (iii) The contours $\partial\mathsf{A}(x) = \{y : \pi(y) = \pi(x)\}$ are asymptotically regular, *i.e.*
$$\lim_{|x|\to+\infty} \sup \left\langle \frac{x}{|x|}, \frac{\nabla\pi(x)}{|\nabla\pi(x)|} \right\rangle < 0.$$

Note that this condition implicitely implies the existence and finiteness of μ_π and Γ_π. We now establish uniform minorisation and drift conditions for P_q^{SRW} defined in Eq. (7.3). Let $\mathcal{M}(\mathsf{X})$ denote the set of probability densities w.r.t. the Lebesgue measure λ^{Leb}. Let $\varepsilon > 0$ and $\delta > 0$ and define the subset $\mathcal{K}_{\delta,\varepsilon} \subset \mathcal{M}(\mathsf{X})$,

$$\mathcal{K}_{\delta,\varepsilon} = \{q \in \mathcal{M}(\mathsf{X}), q(z) = q(-z) \quad \text{and} \quad |z| \leq \varepsilon \Rightarrow q(z) \geq \delta\}. \tag{7.11}$$

PROPOSITION 7.2. *Assume (M). For any $\eta \in (0,1)$, set $W = \pi^{-\eta}/(\inf_{\mathsf{X}} \pi^{-\eta})$. Then,*

1. *Any non-empty compact set* $\mathsf{C} \subset \mathsf{X}$ *is a (1,δ)-small set for some* $\delta > 0$ *and some measure* ν*,*

$$\forall (x, A) \in \mathsf{C} \times \mathcal{B}(\mathsf{X}) \qquad \inf_{q \in \mathcal{K}_{\delta,\varepsilon}} P_q^{\text{SRW}}(x, A) \geq \delta \nu(A). \tag{7.12}$$

2. *Furthermore, for any* $\delta > 0$ *and* $\varepsilon > 0$*,*

$$\sup_{q \in \mathcal{K}_{\delta,\varepsilon}} \limsup_{|x| \to +\infty} \frac{P_q^{\text{SRW}} W(x)}{W(x)} < 1, \tag{7.13}$$

$$\sup_{(x,q) \in \mathsf{X} \times \mathcal{K}_{\delta,\varepsilon}} \frac{P_q^{\text{SRW}} W(x)}{W(x)} < +\infty. \tag{7.14}$$

3. *Let* $q, q' \in \mathcal{M}(\mathsf{X})$ *be two symmetric probability distributions. Then, for any* $r \in [0,1]$ *and any* $g \in \mathcal{L}_{W^r}$ *we have*

$$\left\| P_q^{\text{SRW}} g - P_{q'}^{\text{SRW}} g \right\|_{W^r} \leq 2 \left\| g \right\|_{W^r} \int_{\mathsf{X}} |q(z) - q'(z)| \lambda^{\text{Leb}}(dz). \tag{7.15}$$

Proof. For any $x \in \mathsf{X}$, define the acceptance region $\mathsf{A}(x) = \{z \in \mathsf{X} - x; \pi(x+z) \geq \pi(x)\}$ and the rejection region $\mathsf{R}(x) = \{z \in \mathsf{X} - x; \pi(x+z) < \pi(x)\}$. From the definition (7.11) of $\mathcal{K}_{\delta,\varepsilon}$ [29, Theorem 2.2] applies for any $q \in \mathcal{K}_{\delta,\varepsilon}$ and we can conclude that (7.12) is satisfied. Noting that the two sets $\mathsf{A}(x)$ and $\mathsf{R}(x)$ do not depend on the proposal distribution q and using the conclusion of the proof of Theorem 4.3 of [22] we have

$$\inf_{q \in \mathcal{K}_{\delta,\varepsilon}} \liminf_{|x| \to +\infty} \int_{\mathsf{A}(x)} q(z) \lambda^{\text{Leb}}(dz) > 0,$$

so that from the conclusion of the proof of Theorem 4.1 of [22],

$$\sup_{q \in \mathcal{K}_{\delta,\varepsilon}} \limsup_{|x| \to +\infty} \frac{P_q^{\text{SRW}} W(x)}{W(x)} = 1 - \inf_{q \in \mathcal{K}_{\delta,\varepsilon}} \liminf_{|x| \to +\infty} \int_{\mathsf{A}(x)} q(z) \lambda^{\text{Leb}}(dz) < 1,$$

which proves (7.13). Finally, for any $q \in \mathcal{K}_{\delta,\varepsilon}$,

$$\frac{P_q^{\text{SRW}} W(x)}{W(x)} = \int_{\mathsf{A}(x)} \frac{\pi(x+z)^{-\eta}}{\pi(x)^{-\eta}} q(z) \lambda^{\text{Leb}}(dz) + \int_{\mathsf{R}(x)} \left(1 - \frac{\pi(x+z)}{\pi(x)} + \frac{\pi(x+z)^{1-\eta}}{\pi(x)^{1-\eta}} \right) q(z) \lambda^{\text{Leb}}(dz)$$
$$\leq \sup_{0 \leq u \leq 1} (1 - u + u^{1-\eta}),$$

which proves (7.14). Now notice that

$$P_q^{\text{SRW}} g(x) - P_{q'}^{\text{SRW}} g(x) = \int_{\mathsf{X}} \alpha(x, x+z)(q(z) - q'(z)) g(x+z) \lambda^{\text{Leb}}(dz) +$$
$$g(x) \int_{\mathsf{X}} \alpha(x, x+z)(q'(z) - q(z)) \lambda^{\text{Leb}}(dz).$$

We therefore focus, for $r \in [0,1]$ and $g \in \mathcal{L}_{W^r}$, on the term

$$\frac{\left| \int_{\mathsf{X}} \alpha(x, x+z)(q(z) - q'(z)) g(x+z) \lambda^{\text{Leb}}(dz) \right|}{\|g\|_{W^r} W^r(x)} \leq \frac{\int_{\mathsf{X}} \alpha(x, x+z) |q(z) - q'(z)| W^r(x+z) \lambda^{\text{Leb}}(dz)}{W^r(x)} =$$
$$= \int_{\mathsf{A}(x)} \frac{\pi(x+z)^{-r\eta}}{\pi(x)^{-r\eta}} |q(z) - q'(z)| \lambda^{\text{Leb}}(dz) + \int_{\mathsf{R}(x)} \frac{\pi(x+z)^{1-r\eta}}{\pi(x)^{1-r\eta}} |q(z) - q'(z)| \lambda^{\text{Leb}}(dz)$$
$$\leq \int_{\mathsf{X}} |q(z) - q'(z)| \lambda^{\text{Leb}}(dz).$$

We now conclude that for any $x \in \mathsf{X}$ and any $g \in \mathcal{L}_{W^r}$,

$$\frac{|P_q^{\text{SRW}} g(x) - P_{q'}^{\text{SRW}} g(x)|}{W^r(x)} \leq 2 \|g\|_{W^r} \int_{\mathsf{X}} |q(z) - q'(z)| \lambda^{\text{Leb}}(dz).$$

□

One can specialise the regularity property (7.15) to the N-SRWM, where the proposal distribution q_θ is a zero-mean normal distribution with covariance matrix Γ, and for simplicity we set $q_\Gamma := \phi_{0,\Gamma}$.

LEMMA 7.3. *Let $\mathcal{K}$ be a convex compact subset of $\mathcal{C}_+^{n_x}$ and set $W = \pi^{-\eta}/(\inf_{\mathsf{X}} \pi)^{-\eta}$ for some $\eta \in (0,1)$. For any $r \in [0,1]$, any $\Gamma, \Gamma' \in \mathcal{K} \times \mathcal{K}$, $g \in \mathcal{L}_{W^r}$, we have*

$$\left\| P_{q_\Gamma}^{\text{SRW}} g - P_{q_{\Gamma'}}^{\text{SRW}} g \right\|_{W^r} \leq \frac{2n_x}{\lambda_{\min}(\mathcal{K})} \|g\|_{W^r} |\Gamma - \Gamma'|, \tag{7.16}$$

where $\lambda_{\min}(\mathcal{K})$ is the minimum possible eigenvalue for matrices in $\mathcal{K}$.

Proof. We have

$$\int_{\mathsf{X}} |q_\Gamma(z) - q_{\Gamma'}(z)| dz = \int_{\mathsf{X}} \left| \int_0^1 \frac{d}{dv} q_{\Gamma + v(\Gamma' - \Gamma)}(z) dv \right| dz$$

and let $\Gamma_v = \Gamma + v(\Gamma' - \Gamma)$, so that

$$\frac{d}{dv} \log q_{\Gamma + v(\Gamma' - \Gamma)}(z) = -\frac{1}{2} \operatorname{Tr} \left[\Gamma_v^{-1} (\Gamma' - \Gamma) + \Gamma_v^{-1} z z^{\mathsf{T}} \Gamma_v^{-1} (\Gamma' - \Gamma) \right]$$

and consequently

$$\int_{\mathsf{X}} \left| \int_0^1 \frac{d}{dv} q_{\Gamma + v(\Gamma' - \Gamma)}(z) dv \right| dz \leq |\Gamma' - \Gamma| \int_0^1 |\Gamma_v^{-1}| dv \leq \frac{n_x}{\lambda_{\min}(\mathcal{K})} |\Gamma' - \Gamma|,$$

where we have used the following inequality,

$$|\operatorname{Tr}[\Gamma_v^{-1} z z^{\mathsf{T}} \Gamma_v^{-1} (\Gamma' - \Gamma)]| \leq |\Gamma' - \Gamma| \operatorname{Tr}[\Gamma_v^{-1} \Gamma_v^{-1} z z^{\mathsf{T}}].$$

□

COROLLARY 7.4. *For any compact subset $\mathcal{K}$ of $\mathcal{C}_+^{n_x}$, there exists $C < \infty$ such that*

$$\left\| P_{q_\Gamma}^{\text{SRW}} g - P_{q_{\Gamma'}}^{\text{SRW}} g \right\|_{W^r} \leq C \|g\|_{W^r} |\Gamma - \Gamma'|. \tag{7.17}$$

7.4. Convergence of the adaptive MCMC algorithm. The main result of this section is:

THEOREM 7.5. *Let $\pi \in \mathcal{M}(\mathsf{X})$ satisfying (M). Let $\{Z_n\}$ be the homogenous Markov chain be defined as in Section 3, with H as in Eq. (7.5), $P_\theta := P_{\phi_{0,\lambda\Gamma}}^{\text{SRW}}$ for some $\lambda > 0$ with $\theta = (\mu, \Gamma) \in \Theta = \mathbb{R}^{n_x} \times \mathcal{C}_+^{n_x}$, K a compact set, and $\boldsymbol{\gamma} = \{\gamma_n\}$ and $\boldsymbol{\epsilon} = \{\epsilon_n\}$ satisfying (A4). Then, (A1) to (A3) are satisfied for any $\mathcal{K}_0$ and $\theta_n \to \theta_\pi$ w.p. 1, where $\theta_\pi := (\mu_\pi, \Gamma_\pi)$ is the unique stationary point of $\{\theta_n\}$.*

Proof. (A1) is implied by Proposition 7.1. (A2) is satisfied by construction of P_θ, from (M) and the definition of H in Eq. (7.5). Now we prove that (DRI) is satisfied. Choose $V^p = W = \pi^{-\eta}/\inf_{\mathsf{X}} \pi^{-\eta}$ for

$p \geq 2$ in Proposition 7.2: then (DRI1) and (DRI3) are satisfied. Now (DRI2) is satisfied since from Eq. (7.5)

$$|H_\theta(x) - H_{\theta'}(x)| \leq |\mu - \mu'|\{1 + |\mu + \mu'| + 2|x|\} + |\Gamma - \Gamma'|, \tag{7.18}$$

and $\|x\|_W + \||x|^2\| < \infty$ from (M). Theorem 5.5 now applies.
□

This result is an important step for the study of the asymptotic properties of $\{S_n\}$ in [2], in particular the proof that $\{S_n\}$ satisfies a central limit theorem.

8. Acknowledgements. The authors would like to thank Stas Volkov and Sumeetpal Singh for their careful reading of parts of the paper and their helpful comments.

Appendix A. Proof of Proposition 5.2. Denote

$$D(\boldsymbol{\epsilon}, \boldsymbol{\rho}, \mathcal{K}, x) = \sup_{k\geq 1} \sup_{\theta \in \mathcal{K}} \mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}[W^p(X_k) \mathbb{1}_{\{\sigma(\mathcal{K}) \wedge \nu(\boldsymbol{\epsilon}) \geq k\}}].$$

We first consider the case $l = 1$. Denote

$$T_n = \sum_{k=1}^{n} \rho_k \left(g_{\theta_{k-1}}(X_k) - P_{\theta_{k-1}} g_{\theta_{k-1}}(X_k) \right) \mathbb{1}_{\{\sigma(\mathcal{K}) \wedge \nu(\boldsymbol{\epsilon}) \geq k\}},$$

Using $\mathbb{1}_{\{\sigma(\mathcal{K}) \wedge \nu(\boldsymbol{\epsilon}) \geq k\}} = \mathbb{1}_{\{\sigma(\mathcal{K}) \wedge \nu(\boldsymbol{\epsilon}) \geq k+1\}} + \mathbb{1}_{\{\sigma(\mathcal{K}) \wedge \nu(\boldsymbol{\epsilon}) = k\}}$, we may write $T_n = \sum_{i=1}^5 T_n^{(i)}$ where

$$T_n^{(1)} = \sum_{k=1}^{n} \rho_k \left(g_{\theta_{k-1}}(X_k) - P_{\theta_{k-1}} g_{\theta_{k-1}}(X_{k-1}) \right) \mathbb{1}_{\{\sigma(\mathcal{K}) \wedge \nu(\boldsymbol{\epsilon}) \geq k\}}, \tag{A.1}$$

$$T_n^{(2)} = \sum_{k=1}^{n-1} \rho_{k+1} \left(P_{\theta_k} g_{\theta_k}(X_k) - P_{\theta_{k-1}} g_{\theta_{k-1}}(X_k) \right) \mathbb{1}_{\{\sigma(\mathcal{K}) \wedge \nu(\boldsymbol{\epsilon}) \geq k+1\}}, \tag{A.2}$$

$$T_n^{(3)} = \sum_{k=1}^{n-1} (\rho_{k+1} - \rho_k) P_{\theta_{k-1}} g_{\theta_{k-1}}(X_k) \mathbb{1}_{\{\sigma(\mathcal{K}) \wedge \nu(\boldsymbol{\epsilon}) \geq k+1\}}, \tag{A.3}$$

$$T_n^{(4)} = \rho_1 P_{\theta_0} g_{\theta_0}(X_0) \mathbb{1}_{\{\sigma(\mathcal{K}) \wedge \nu(\boldsymbol{\epsilon}) \geq 1\}} - \rho_n P_{\theta_{n-1}} g_{\theta_{n-1}}(X_n) \mathbb{1}_{\{\sigma(\mathcal{K}) \wedge \nu(\boldsymbol{\epsilon}) \geq n\}}, \tag{A.4}$$

$$T_n^{(5)} = -\sum_{k=1}^{n-1} \rho_k P_{\theta_{k-1}} g_{\theta_{k-1}}(X_k) \mathbb{1}_{\{\sigma(\mathcal{K}) \wedge \nu(\boldsymbol{\epsilon}) = k\}}. \tag{A.5}$$

We now evaluate bounds for $T_n^{(i)}$, $i = 1, \ldots, 4$. In the sequel C denotes a constant which depends only upon the compact set $\mathcal{K}$ through the quantities defined in the assumptions and whose value may change upon each appearance. We have

$$\sup_{\theta\in\mathcal{K}} \mathbb{E}^{\boldsymbol{\rho}}_{\theta,x}\left[\sup_{n\geq 0}\left|T^{(1)}_n\right|^p\right] \leq C\left(\sum_{k=0}^{\infty}\rho_k^2\right)^{p/2} \sup_{\theta\in\mathcal{K}}\sup_k \mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[W^p(X_k)\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu(\boldsymbol{\epsilon})\geq k\}}\right], \tag{A.6}$$

$$\sup_{\theta\in\mathcal{K}} \mathbb{E}^{\boldsymbol{\rho}}_{\theta,x}\left[\sup_{n\geq 0}\left|T^{(2)}_n\right|^p\right] \leq C\left(\sum_{k=1}^{\infty}\rho_k\epsilon_k^{\alpha}\right)^{p} \sup_{\theta\in\mathcal{K}}\sup_k \mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[W^p(X_k)\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu(\boldsymbol{\epsilon})\geq k\}}\right], \tag{A.7}$$

$$\sup_{\theta\in\mathcal{K}} \mathbb{E}^{\boldsymbol{\rho}}_{\theta,x}\left[\sup_{n\geq 0}\left|T^{(3)}_n\right|^p\right] \leq C\rho_1^p \sup_{\theta\in\mathcal{K}}\sup_k \mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[W^p(X_k)\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu(\boldsymbol{\epsilon})\geq k\}}\right], \tag{A.8}$$

$$\sup_{\theta\in\mathcal{K}} \mathbb{E}^{\boldsymbol{\rho}}_{\theta,x}\left[\sup_{n\geq 0}\left|T^{(4)}_n\right|^p\right] \leq C\left(\sum_{k=1}^{\infty}\rho_k^2\right)^{p/2} \sup_{\theta\in\mathcal{K}}\sup_k \mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[W^p(X_k)\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu(\boldsymbol{\epsilon})\geq k\}}\right]. \tag{A.9}$$

The proof of these inequalities can be adapted from [5, Part II, Section 3.2] - see also [4], Chapter 6, Lemma 6.2-6.4.

Proof of (A.6). Under (A3),

$$\sup_{\theta\in\mathcal{K}} \mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[(|g_{\theta_k}(X_{k+1})|^p + |P_{\theta_k}g_{\theta_k}(X_{k+1})|^p)\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu(\boldsymbol{\epsilon})\geq k+1\}}\right] \leq CD(\boldsymbol{\epsilon},\boldsymbol{\rho},\mathcal{K},x).$$

Since

$$\mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[(g_{\theta_k}(X_{k+1}) - P_{\theta_k}g_{\theta_k}(X_k))\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu(\boldsymbol{\epsilon})\geq k+1\}}\middle|\, \mathcal{F}_k\right] = \\ (P_{\theta_k}g_{\theta_k}(X_k) - P_{\theta_k}g_{\theta_k}(X_k))\,\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu(\boldsymbol{\epsilon})\geq (k+1)\}} = 0,$$

$T^{(1)}_n$ is a ($\mathbb{R}^d$-valued) martingale. Using the Burkholder inequality ([21], Theorem 2.10), we have

$$\mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[\left|T^{(1)}_n\right|^p\right] \leq C_p\mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left(\sum_{k=0}^{n-1}\rho_{k+1}^2|g_{\theta_k}(X_{k+1}) - P_{\theta_k}g_{\theta_k}(X_k)|^2\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu(\boldsymbol{\epsilon})\geq k+1\}}\right)^{p/2}, \tag{A.10}$$

where C_p is a universal constant. Using Minkowski's inequality and $\sup_{\theta\in\mathcal{K}}(\|g_\theta\|_V + \|P_\theta g_\theta\|_V) < \infty$, (A3), we have

$$\mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[\left|T^{(1)}_n\right|^p\right] \leq C\left(\sum_{k=1}^{\infty}\rho_k^2\right)^{p/2} D(\boldsymbol{\epsilon},\boldsymbol{\rho},\mathcal{K},x).$$

Since $T^{(1)}_n$ is a martingale in $\mathcal{L}^p$, then $|T^{(1)}_n|$ is a non-negative submartingale in $\mathcal{L}^p$ and Doob's $\mathcal{L}^p$ inequality implies that

$$\mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[\sup_{n\geq 1}\left|T^{(1)}_n\right|^p\right] \leq C\left(\sum_{k=1}^{\infty}\rho_k^2\right)^{p/2} D(\boldsymbol{\epsilon},\boldsymbol{\rho},\mathcal{K},x),$$

which concludes the proof of (A.6).

Proof of (A.7). Under (A3), we have

$$\begin{aligned}
&\sup_{n\geq 1}\left|\sum_{k=1}^{n-1}\rho_{k+1}\left(P_{\theta_k}g_{\theta_k}(X_k)-P_{\theta_{k-1}}g_{\theta_{k-1}}(X_k)\right)\ \mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu(\boldsymbol{\epsilon})\geq k+1\}}\right|\\
&\leq C\sum_{k=0}^{\infty}\rho_{k+1}W(X_k)\ |\theta_k-\theta_{k-1}|^{\alpha}\ \mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu(\boldsymbol{\epsilon})\geq k+1\}},\\
&\leq C\sum_{k=0}^{\infty}\rho_{k+1}\epsilon_k^{\alpha}\ W(X_k)\ \mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu(\boldsymbol{\epsilon})\geq k+1\}}.
\end{aligned}$$

We conclude the proof by applying Minkowski's inequality.

Proof of (A.8). Under (A3),

$$\sup_{n\geq 1}\left|\sum_{k=1}^{n-1}(\rho_{k+1}-\rho_k)P_{\theta_{k-1}}g_{\theta_{k-1}}(X_k)\ \mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu(\boldsymbol{\epsilon})\geq k+1\}}\right|\leq C\sum_{k=1}^{\infty}(\rho_k-\rho_{k+1})W(X_k)\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu(\boldsymbol{\epsilon})\geq k+1\}},$$

and the proof follows from Minkowski's inequality.

Proof of (A.9). Under (A3),

$$\begin{aligned}
&\sup_{n\geq 1}\left|\rho_1P_{\theta_0}g_{\theta_0}(X_0)\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu(\boldsymbol{\epsilon})\geq 1\}}-\rho_nP_{\theta_{n-1}}g_{\theta_{n-1}}(X_n)\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu(\boldsymbol{\epsilon})\geq n\}}\right|^p\leq\\
&C\left(\rho_1^pW^p(X_0)\ \mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu(\boldsymbol{\epsilon})\geq 1\}}+\sup_{n\geq 1}\rho_n^pW^p(X_n)\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu(\boldsymbol{\epsilon})\geq n\}}\right)\leq C\sum_{k=1}^{\infty}\rho_k^pW^p(X_k)\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu(\boldsymbol{\epsilon})\geq k\}}.
\end{aligned}$$

The proof follows from (A3), the inequality $\sum_{k=1}^n\rho_k^p\leq\left(\sum_{k=1}^n\rho_k^2\right)^{p/2}$ for $p\geq 2$.

Since $T_n^{(5)}\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu(\boldsymbol{\epsilon})\geq n\}}=0$, we have

$$S_{1,n}(\boldsymbol{\epsilon},\boldsymbol{\rho},\mathcal{K})=T_n\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu(\boldsymbol{\epsilon})\geq n\}}=\sum_{i=1}^{4}T_n^{(i)}\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu(\boldsymbol{\epsilon})\geq n\}}.$$

The Markov inequality and Lemma 5.1 imply that

$$\mathbb{P}_{x,\theta}^{\boldsymbol{\rho}}\left(\sup_{n\geq 1}|S_{1,n}(\boldsymbol{\epsilon},\boldsymbol{\rho},\mathcal{K})|\geq\delta\right)\leq C\delta^{-p}\left\{\left(\sum_{k=1}^{\infty}\rho_k^2\right)^{p/2}+\left(\sum_{k=1}^{\infty}\rho_k\epsilon_k^{\alpha}\right)^p\right\}W^p(x).\qquad\text{(A.11)}$$

The proof for all l then follows from the Markov property: for all $(x,\theta)\in\mathsf{X}\times\mathcal{K}$,

$$\begin{aligned}
\mathbb{P}_{x,\theta}^{\boldsymbol{\rho}}\left(\sup_{n\geq 1}|S_{l+1,n}(\boldsymbol{\epsilon},\boldsymbol{\rho},\mathcal{K})|\geq\delta\right)&=\mathbb{E}_{x,\theta}^{\boldsymbol{\rho}}\left(\mathbb{P}_{X_l,\theta_l}^{\boldsymbol{\rho}^{\leftarrow l}}\left(\sup_{n\geq 1}|S_{1,n}(\boldsymbol{\epsilon}^{\leftarrow l},\boldsymbol{\rho}^{\leftarrow l},\mathcal{K})|\geq\delta\right)\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu(\boldsymbol{\epsilon})\geq l\}}\right)\\
&\leq\mathbb{E}_{x,\theta}^{\boldsymbol{\rho}}\left(\sup_{\theta\in\mathcal{K}}\mathbb{P}_{X_l,\theta}^{\boldsymbol{\rho}^{\leftarrow l}}\left(\sup_{n\geq 1}|S_{1,n}(\boldsymbol{\epsilon}^{\leftarrow l},\boldsymbol{\rho}^{\leftarrow l},\mathcal{K})|\geq\delta\right)\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu(\boldsymbol{\epsilon})\geq l\}}\right).
\end{aligned}$$

Since the sequence $\boldsymbol{\epsilon}$ is non-increasing, there exists an integer s such that for all l and all $k\geq s$, $\epsilon_k^{\leftarrow l}\leq\epsilon$, for all $k\geq s$ (where ϵ is defined in (A3)) and Lemma 5.1 shows that there exists a constant C such that for any l, for any $x\in\mathsf{X}$ and any monotone non-increasing sequence $\boldsymbol{\rho}$,

$$\sup_{\theta\in\mathcal{K}}\sup_{k\geq 0}\mathbb{E}_{x,\theta}^{\boldsymbol{\rho}^{\leftarrow l}}[W^p(X_k)\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu(\boldsymbol{\epsilon}^{\leftarrow l})\geq k\}}]\leq CW^p(x).$$

The proof follows from (A.11).

It remains to bound $\mathbb{P}^{\boldsymbol{\rho}}_{x,\theta}(\nu(\boldsymbol{\epsilon}) < \sigma(\mathcal{K})) \leq \mathbb{P}^{\boldsymbol{\rho}}_{x,\theta}(\nu(\boldsymbol{\epsilon}) \leq \sigma(\mathcal{K}))$.

$$\begin{aligned}
\mathbb{P}^{\boldsymbol{\rho}}_{x,\theta}(\nu(\boldsymbol{\epsilon}) \leq \sigma(\mathcal{K})) &= \sum_{k=1}^{\infty} \mathbb{P}^{\boldsymbol{\rho}}_{x,\theta}(\nu(\boldsymbol{\epsilon}) = k, \sigma(\mathcal{K}) \geq k) \\
&= \sum_{k=1}^{\infty} \mathbb{P}^{\boldsymbol{\rho}}_{x,\theta}\left(|H(\theta_{k-1}, X_k)| \geq \epsilon_k \rho_k^{-1}, \sigma(\mathcal{K}) \geq k, \nu(\boldsymbol{\epsilon}) = k\right) \\
&\leq C \sum_{k=1}^{\infty} (\epsilon_k^{-1} \rho_k)^p \sup_{k \geq 0} \sup_{\theta \in \mathcal{K}} \mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}[W^p(X_k) \mathbb{1}_{\{\sigma(\mathcal{K}) \wedge \nu(\boldsymbol{\epsilon}) \geq k\}}].
\end{aligned}$$

The proof follows from Lemma 5.1.

Appendix B. Proof of Proposition 6.1. The following proposition is a (partial) restatement of [15, Theorem 1] (see also [Theorem 2.3][26]).

PROPOSITION B.1. *Suppose that P is irreducible and aperiodic and that $P^m(x, \cdot) \geq \mathbb{1}_C(x) \delta \nu(\cdot)$ for a set $C \in \mathcal{B}(\mathsf{X})$, some integer m and $\delta > 0$ and that there is a drift to C in the sense that, for some $\lambda < 1$, b and a function $V : \mathsf{X} \to [1, \infty)$,*

$$PV(x) \leq \lambda V(x) \quad \forall x \notin C \quad \text{and} \quad \sup_{x \in C}(V(x) + PV(x)) \leq b. \tag{B.1}$$

Then, there exist constants K and $\rho < 1$, depending only upon m, δ, λ, b, such that, for all $x \in \mathsf{X}$, and all $g \in \mathcal{L}_V$

$$\|P^k g - \pi(g)\|_V \leq K \rho^n \, \|g\|_V. \tag{B.2}$$

In addition, $u = \sum_{n \geq 0}(P^k g - \pi(g))$ is a solution of the Poisson equation $u - Pu = g - \pi(g)$.

[26, Theorem 2.3] is stated in the strongly aperiodic case, *i.e.* where C is a $(1, \delta)$ small set. Explicit but intricate expressions for K and ρ (in terms of the constants m, δ, λ, b) are given in this reference. Partial extensions to the general aperiodic case is considered in [26, Theorem 2.4], based on splitting and regeneration techniques. Sharper and simpler bounds have been recently obtained in [15] using coupling technique. This result extends to V-norm results obtained earlier for the total variation distance by [31] (see also [30]). These results have been derived in the strongly aperiodic case; extensions to the general aperiodic case can be considered in the same framework.

PROPOSITION B.2. *Assume (DRI1)-(DRI3). Then, there exist a constant C and $\rho < 1$ such that, for all $g \in \mathcal{L}_{V^q}$, with $q = 1$ or $q = p$ and any $k \geq 0$*

$$\sup_{\theta \in \mathcal{K}} \|P^k_\theta g - \pi_\theta(g)\|_{V^q} \leq C \rho^k \|g\|_{V^q}, \tag{B.3}$$

$$\sup_{(\theta, \theta') \in \mathcal{K} \times \mathcal{K}} |\theta - \theta'|^{-\beta} \|P^k_\theta g - P^k_{\theta'} g\|_{V^q} \leq C \|g\|_{V^q}. \tag{B.4}$$

Proof. Eq. (B.3) follows from Proposition B.1. To prove (B.4) write, for all $(\theta, \theta') \in \Theta \times \Theta$, all $n \in \mathbb{N}$, and all $g \in \mathcal{L}_{V^q}$

$$P^n_\theta g(x) - P^n_{\theta'} g(x) = \sum_{j=0}^{n-1} P^j_\theta (P_\theta - P_{\theta'}) P^{n-j-1}_{\theta'} g(x) = \sum_{j=0}^{n-1} P^j_\theta (P_\theta - P_{\theta'}) (P^{n-j-1}_{\theta'} g(x) - \pi_{\theta'}(g)).$$

Eq. (B.3) shows that there exists a constant C such that, for any $l \geq 0$,

$$\sup_{\theta \in \mathcal{K}} \|P_\theta^l g - \pi_\theta(g)\|_{V^q} \leq C \|g\|_{V^q}\ \rho^l \quad \text{and} \quad \sup_{j \geq 0} \sup_{\theta \in \mathcal{K}} \|P_\theta^j V^q\|_{V^q} < \infty .$$

Under assumption (DRI3) we thus have, for any $l \geq 0$,

$$\|(P_\theta - P'_\theta)(P_{\theta'}^l g(x) - \pi_{\theta'}(g))\|_{V^q} \leq C|\theta - \theta'|^\beta \|(P_{\theta'}^l g(x) - \pi_{\theta'}(g))\|_{V^q} \leq C|\theta - \theta'|^\beta \|g\|_{V^q} \rho^l ,$$

which concludes the proof. □

Proof of Proposition 6.1. Under (DRI1), P_θ is positive recurrent and admits a single stationary measure π_θ, which verifies $\sup_{\theta \in \mathcal{K}} \pi_\theta(V^p) < \infty$ which implies that $\sup_{\theta \in \mathcal{K}} |h(\theta)| < \infty$.

Proof. [Proof Eq. (6.6)] Let $x_0 \in \mathsf{X}$ and $k \in \mathbb{N}$. Write $h(\theta) - h(\theta') = A(\theta, \theta') + B(\theta, \theta') + C(\theta, \theta')$ where

$$A(\theta, \theta') = \left(h(\theta) - P_\theta^k H_\theta(x_0)\right) + \left(P_{\theta'}^k H_{\theta'}(x_0) - h(\theta')\right), \tag{B.5}$$

$$B(\theta, \theta') = P_\theta^k H_\theta(x_0) - P_{\theta'}^k H_\theta(x_0), \tag{B.6}$$

$$C(\theta, \theta') = P_{\theta'}^k H_\theta(x_0) - P_{\theta'}^k H_{\theta'}(x_0). \tag{B.7}$$

Propositions B.1 and B.2 show that there exist constants C and $\rho < 1$ such that, for all $(\theta, \theta') \in \mathcal{K} \times \mathcal{K}$,

$$|A(\theta, \theta')| \leq C\ \rho^k \sup_{\theta \in \mathcal{K}} \|H_\theta\|_V\ V(x_0),$$

$$|B(\theta, \theta')| \leq C \sup_{\theta \in \mathcal{K}} \|H_\theta\|_V\ |\theta - \theta'|^\beta\ V(x_0),$$

$$|C(\theta, \theta')| \leq \int_{\mathsf{X}} P_{\theta'}^k(x_0, dy) |H_\theta(y) - H_{\theta'}(y)| \leq C|\theta - \theta'|^\beta \int_{\mathsf{X}} P_{\theta'}^k(x_0, dy) V(y) \leq C|\theta - \theta'|^\beta\ V(x_0).$$

Hence, there exists a constant C such that, for all $(\theta, \theta') \in \mathcal{K} \times \mathcal{K}$,

$$|h(\theta) - h(\theta')| \leq C\ V(x_0)\ (\rho^k + |\theta - \theta'|^\beta). \tag{B.8}$$

The proof is concluded by setting $k = [\beta \log|\theta - \theta'| / \log(\rho)]$ (where $[x]$ is the integer part of x) if $|\theta - \theta'| \leq \delta < 1$ and $k = 1$ otherwise. □

Proof. [Proof of Eq. (5.4)] Using Eq.(6.6), Proposition B.1 and B.2, there exists a constant C such that, for all $(\theta, \theta') \in \mathcal{K}$ we have

$$|(P_\theta^k H_\theta(x) - h(\theta)) - (P_{\theta'}^k(x) H_{\theta'}(x) - h(\theta'))| \leq \\ |P_\theta^k H_\theta(x) - P_\theta^k H_{\theta'}(x)| + |P_\theta^k H_{\theta'}(x) - P_{\theta'}^k H_{\theta'}(x)| + |h(\theta) - h(\theta')| \leq C|\theta - \theta'|^\beta V(x).$$

On the other hand, by Proposition B.1, there exist constants $\rho < 1$ and C such that, for all $(\theta, \theta') \in \mathcal{K} \times \mathcal{K}$,

$$|(P_\theta^k H_\theta(x) - h(\theta)) - (P_{\theta'}^k H_{\theta'}(x) - h(\theta'))| \leq C \rho^k V(x).$$

Hence, for any s and $N \geq s$, we have

$$|P_\theta^s g_\theta(x) - P_{\theta'}^s g_{\theta'}(x)| \leq \sum_{k=s}^{\infty} |(P_\theta^k H_\theta(x) - h(\theta)) - (P_{\theta'}^k(x) - h(\theta'))| \leq CV(x)\ \left\{ N|\theta - \theta'|^\beta + \frac{\rho^{N+s}}{1 - \rho} \right\}.$$

The proof follows by setting $N = [\beta \log|\theta - \theta'|/\log\rho]$, for $|\theta - \theta'| \leq \delta < 1$, $\theta \neq \theta'$, $N = s$ otherwise, and using the fact that for any $0 < \alpha < \beta$, $|\theta - \theta'|^\beta \log|\theta - \theta'| = o(|\theta - \theta'|^\alpha)$. □

Proof. [Proof Eq. (5.6)] Let $\boldsymbol{\rho} = \{\rho_k, k \geq 0\}$ be a non-increasing sequence of positive numbers and let $\mathcal{K}$ be a compact subset of Θ. (DRI1), Eq. (6.2) shows that, for all $k \geq 0$, $l \geq 0$, all $x \in \mathsf{X}$,

$$\sup_{\theta \in \mathcal{K}} \mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[V^p(X_{k+l})\mathbb{1}_{\{\sigma(\mathcal{K}) \geq k+l\}} \mid \mathcal{F}_k\right] \leq \kappa^l V^p(X_k)\mathbb{1}_{\{\sigma(\mathcal{K}) \geq k\}}. \tag{B.9}$$

We will show that there exist constants $\epsilon > 0$, $0 < \rho < 1$ and C such that, for all k

$$\mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[V^p(X_{k+m})\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu_\epsilon \geq k+m\}} \mid \mathcal{F}_k\right] \leq \rho V^p(X_k)\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu_\epsilon \geq k\}} + C. \tag{B.10}$$

For $n \in \mathbb{N}$, write $n = um + v$, where $v \in \{0, \ldots, m-1\}$. (B.10) shows that

$$\mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[V^p(X_{um+v})\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu_\epsilon \geq um+v\}}\right] \leq \rho^u \mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}[V^p(X_v)\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu_\epsilon \geq v\}}] + \frac{C}{1-\rho}$$

and the proof follows from (B.9). It remains to prove (B.10). We repeatedly use the following lemma adapted from [5] (Lemma 3, p. 292.)

Lemma B.3. *Assume (DRI). Let $\psi : \Theta \times \mathsf{X} \to \mathbb{R}$ be a function verifying $\sup_{\theta\in\mathcal{K}} \|\psi_\theta\|_{V^p} < \infty$. Then, for any $\epsilon > 0$, for any $l \geq 1$ there exist a constant C such that, for all $k \geq 0$,*

$$\begin{aligned}\mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[\psi_{\theta_k}(X_{k+l})\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu_\epsilon \geq k+l\}} \mid \mathcal{F}_k\right] &\leq \mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[P_{\theta_k}\psi_{\theta_k}(X_{k+l-1})\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu_\epsilon \geq k+l-1\}} \mid \mathcal{F}_k\right] \\ &\quad + C\kappa^l\epsilon^\alpha \sup_{\theta\in\mathcal{K}} \|\psi_\theta\|_{V^p}\ V^p(X_k)\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu_\epsilon \geq k\}}.\end{aligned}$$

Proof.

$$\begin{aligned}&\mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[\psi(\theta_k, X_{k+l})\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu_\epsilon \geq k+l\}} \mid \mathcal{F}_k\right] = \mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[P_{\theta_{k+l-1}}\psi_{\theta_k}(X_{k+l-1})\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu_\epsilon \geq k+l\}} \mid \mathcal{F}_k\right] \\ &= \mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[P_{\theta_k}\psi_{\theta_k}(X_{k+l-1})\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu_\epsilon \geq k+l\}} \mid \mathcal{F}_k\right] + R_l\end{aligned}$$

where

$$R_l = \mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[(P_{\theta_{k+l-1}} - P_{\theta_k})\psi_{\theta_k}(X_{k+l-1})\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu_\epsilon \geq k+l\}} \mid \mathcal{F}_k\right].$$

Under (DRI3), there exists a constant C such that for all $x \in \mathsf{X}$

$$|(P_{\theta_{k+l-1}} - P_{\theta_k})\psi_{\theta_k}(x)\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu_\epsilon \geq k+l\}}| \leq C \sup_{\theta\in\mathcal{K}} \|\psi_\theta\|_{V^p}\ V^p(x)(l\epsilon)^\alpha \mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu_\epsilon \geq k+l\}}.$$

Finally, (DRI1) implies that

$$\mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[V^p(X_{k+l-1})\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu_\epsilon \geq k+l\}} \mid \mathcal{F}_k\right] \leq \kappa^l V^p(X_k)\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu_\epsilon \geq k\}},$$

which implies

$$|R_l| \leq C\kappa^l\ (l\epsilon)^\alpha \sup_{\theta\in\mathcal{K}} \|\psi_\theta\|_{V^p}\ V^p(X_k)\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu_\epsilon \geq k\}}.$$

□

Using repeatedly the lemma above, we may write

$$\begin{aligned}
&\mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[V^p(X_{k+m})\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu_\epsilon\geq k+m\}} \mid \mathcal{F}_k\right] \\
&\quad\leq \mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[P_{\theta_k}V^p(X_{k+m-1})\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu_\epsilon\geq k+m-1\}} \mid \mathcal{F}_k\right] + C_m\epsilon^\alpha V^p(X_k)\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu_\epsilon\geq k\}} \\
&\quad\leq \mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[P^2_{\theta_k}V^p(X_{k+m-2})\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu_\epsilon\geq k+m-2\}} \mid \mathcal{F}_k\right] + (C_m + C_{m-1}\kappa)\epsilon^\alpha V^p(X_k)\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu_\epsilon\geq k\}} \\
&\quad\vdots \\
&\quad\leq P^m_{\theta_k}V^p(X_k)\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu_\epsilon\geq k\}} + \left(\sum_{i=0}^{m-1} C_{m-i}\kappa^i\right)\epsilon^\alpha V^p(X_k)\mathbb{1}_{\{\sigma(\mathcal{K})\wedge\nu_\epsilon\geq k\}}.
\end{aligned}$$

The proof follows from (DRI) for ϵ sufficiently small. □

REFERENCES

[1] J. Abounadi, D. Bertsekas, and V. Borkar, *Stochastic approximation for nonexpansive maps: Application to q-learning algorithms*, SIAM J. on Control and Optimization, 41 (2002), pp. 1–22.

[2] C. Andrieu and E. Moulines, *On the ergodicity properties of some markov chain monte carlo algorithms.*, Technical report, U. of Bristol, Stats. Group, (submitted), 02-19 (2002).

[3] C. Andrieu and C. Robert, *Controlled MCMC for optimal sampling*, (2001).

[4] J. Bartusek, *Stochastic Approximation and Optimization of Markov Chains*, PhD thesis, 2000.

[5] A. Benveniste, M. Métivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*, Springer-Verlag, New York, 1990.

[6] V. Borkar, *Stability of annealing schemes and related processes*, Systems Control Lett., 41 (2000), pp. 325–331.

[7] V. Borkar and S. Meyn, *The o.d.e. method for convergence of stochastic approximation and reinforcement learning*, SIAM J. Control and Optimization, 38 (2000), pp. 447–469.

[8] R. Buche and H. Kushner, *Rate of convergence for constrained stochastic approximation algorithms*, SIAM J. on Control and Optimization, 40 (2001), pp. 1011–1041.

[9] H. Chen, *Stochastic approxmation with state-dependent noise*, Science in China Series E, 43 (2000), pp. 531–541.

[10] H. Chen, L. Guo, and A. Gao, *Convergence and robustness of the Robbins- Monro algorithm truncated at randomly varying bounds*, Stochastic Processes and their Applications, 27 (1988), pp. 217–231.

[11] H. Chen and Y.-M. Zhu, *Stochastic approximation procedures with randomly varying truncations*, Scientia Sinica 1, 29 (1986), pp. 914–926.

[12] H.-F. Chen, *Stochastic approximation and its applications*, vol. 64 of Nonconvex Optimization and its Applications, Kluwer Academic Publishers, Dordrecht, 2002.

[13] B. Delyon, *Stochastic approximation with decreasing gain: Convergence and asymptotic theory*, tech. report, Université de Rennes, 2000.

[14] B. Delyon, M. Lavielle, and E. Moulines, *On a stochastic approximation version of the EM algorithm*, Annals of Statistics, 27 (1999).

[15] R. Douc, E. Moulines, and J. Rosenthal, *Quantitative bounds for geometric convergence rates of Markov chains.* submitted to Annals of Applied Probability, 2002.

[16] M. Duflo, *Random Iterative Systems*, Applications of mathematics, Springer-Verlag, 1997.

[17] A. Gelman, G. Roberts, and W. Gilks, *Efficient Metropolis jumping rules*, in Bayesian Statistics, J. O. Berger, J. M. Bernardo, A. P. Dawid, and A. F. M. Smith, eds., vol. V, Oxford University Press, 1995.

[18] L. Gerencer, *Rate of convergence of recursive estimators*, SIAM J. on Control and Optimization, 30 (1992), pp. 1200–1227.

[19] P. W. Glynn and S. P. Meyn, *A Liapounov bound for solutions of the Poisson equation*, Annals of Probability, 24 (1996), pp. 916–931.

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

[21] P. Hall and C. Heyde, *Martingale Limit Theory and its Application*, Academic Press, New York, London, 1980.

[22] S. Jarner and E. Hansen, *Geometric ergodicity of Metropolis algorithms*, Stochastic Processes and Their Applications, 85 (2000), pp. 341–361.

[23] H. Kushner and D. Clark, *Stochastic Approximation for Constrained and Unconstrained Systems*, Springer-Verlag, Berlin, Heidelberg, 1978.

[24] H. Kushner and G. Yin, *Stochastic Approximation Algorithms and Applications*, Applications of Mathematics, Springer-Verlag, New-York, 1997.

[25] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and M. Teller, *Equations of state calculations by fast computing machines,*, Journal of Chemical Physics, 21 (1953), pp. 1087–1091.
[26] S. Meyn and R. Tweedie, *Markov Chains and Stochastic Stability.*, Communication and Control Engineering series, Springer-Verlag, London, 1993.
[27] E. Nummelin, *On the Poisson equation in the potential theory of a single kernel*, Math. Scand., 68 (1991), pp. 59–82.
[28] H. Robbins and S. Monro, *A stochastic approximation method*, Annals of mathematical statistics, 22 (1951), pp. 400–407.
[29] G. Roberts and R. Tweedie, *Geometric convergence and central limit theorem for multidimensional Hastings and Metropolis algorithms*, Biometrika, 83 (1996), pp. 95–110.
[30] ———, *Bounds on regeneration times and convergence rates for Markov chains.*, Stochastic Processes and Their Applications, 80 (1999), pp. 211–229.
[31] J. Rosenthal, *Minorization conditions and convergence rates for Markov chain Monte Carlo*, Journal American Statistical Association, 90 (1995), pp. 558–566.
[32] V. Tadic, *Stochastic gradient with random truncations*, European J. of Operational Research, 101 (1997), pp. 261–284.
[33] ———, *Stochastic approximations with random truncations, state dependent noise and discontinuous dynamics*, Stochastics and Stochastics reports, 64 (1998), pp. 283–326.