CONVERGENCE OF MARKOVIAN STOCHASTIC APPROXIMATION WITH DISCONTINUOUS DYNAMICS

G. FORT *†, E. MOULINES *, A. SCHRECK *, AND M. VIHOLA ‡

Abstract. This paper is devoted to the convergence analysis of stochastic approximation algorithms of the form $\theta_{n+1} = \theta_n + \gamma_{n+1} H_{\theta_n}(X_{n+1})$ where $\{\theta_n, n \in \mathbb{N}\}$ is a $\mathbb{R}^d$-valued sequence, $\{\gamma_n, n \in \mathbb{N}\}$ is a deterministic step-size sequence and $\{X_n, n \in \mathbb{N}\}$ is a controlled Markov chain. We study the convergence under weak assumptions on smoothness-in-θ of the function $\theta \mapsto H_\theta(x)$. It is usually assumed that this function is continuous for any x; in this work, we relax this condition. Our results are illustrated by considering stochastic approximation algorithms for (adaptive) quantile estimation and a penalized version of the vector quantization.

Key words. Stochastic approximation, discontinuous dynamics, state-dependent noise, controlled Markov chain.

AMS subject classifications. 62L20, secondary: 90C15, 65C40

1. Introduction.

Stochastic Approximation (SA) methods have been introduced by [35] as algorithms to find the roots of $h : \Theta \to \mathbb{R}^d$ where Θ is an open subset of $\mathbb{R}^d$ (equipped with its Borel σ-field $\mathcal{B}(\Theta)$) when only noisy measurements of h are available. More precisely, let X be a space equipped with a countably generated σ-field $\mathcal{X}$, $\{P_\theta, \theta \in \Theta\}$ be a family of transition kernels on $(\mathsf{X}, \mathcal{X})$ and $H : \mathsf{X} \times \Theta \to \mathbb{R}^d$, $(x, \theta) \mapsto H_\theta(x)$ be a measurable function. We consider

$$\theta_{n+1} = \theta_n + \gamma_{n+1} H_{\theta_n}(X_{n+1}) \tag{1.1}$$

where $\{\gamma_n, n \in \mathbb{N}\}$ is a sequence of deterministic nonnegative step sizes and $\{X_n, n \in \mathbb{N}\}$ is a controlled Markov chain, *i.e.*, for any non-negative measurable function f,

$$\mathbb{E}[f(X_{n+1})|\mathcal{F}_n] = P_{\theta_n} f(X_n), \quad \mathbb{P}-\text{a.s.}\ , \quad \mathcal{F}_n = \sigma((X_\ell, \theta_\ell), \ell \leq n).$$

It is assumed that for each $\theta \in \Theta$, P_θ admits a unique stationary distribution π_θ and that $h(\theta) = \int_{\mathsf{X}} H_\theta(x) \pi_\theta(\mathrm{d}x) = \pi_\theta(H_\theta)$ (assuming that $\pi_\theta(|H_\theta|) < \infty$). This setting encompasses the cases $\{X_n, n \in \mathbb{N}\}$ is a (non-controlled) Markov chain by choosing $P_\theta = P$ for any θ; the Robbins-Monro case by choosing $P_\theta(x, \cdot) = \pi_\theta(\cdot)$ where π_θ is a distribution on X; the case when $\{X_n, n \in \mathbb{N}\}$ is an i.i.d. sequence with distribution π by choosing $P_\theta(x, \cdot) = \pi(\cdot)$ for any x, θ.

The goal of this paper is to provide almost sure convergence results of the sequence $\{\theta_n, n \in \mathbb{N}\}$ under conditions on the regularity of the $\theta \mapsto H_\theta(x)$ which does not include continuity, which is usually assumed in the literature. When $\{X_n, n \in \mathbb{N}\}$ is a controlled Markov chain, [9] and [37, Theorem 4.1] establish a.s. convergence under the assumption that for any x, $\theta \mapsto H_\theta(x)$ is Hölder-continuous. This assumption traces back to [22, Eq. (4.2)] and the same assumption is imposed in [3, assumption (DRI2) and Proposition 6.1.].

In order to prove convergence, a preliminary step is to establish that the sequence $\{\theta_n, n \in \mathbb{N}\}$ is $\mathbb{P}$-a.s. in a compact set of Θ, a property referred to as *stability* in [24]. It is common in applications that stability fails to hold or it cannot be theoretically guaranteed. When the 'unconstrained' process as stated above can be shown to be

†corresponding author. mail: gersende.fort@telecom-paristech.fr
*LTCI ; Télécom ParisTech & CNRS
‡University of Jyväskylä ; Department of Mathematics and Statistics

stable, the proof often requires problem specific arguments; see for instance [36, 4]. Different algorithmic modifications for ensuring stability have been suggested in the literature. It is sometimes possible to modify H without modifying the stationary points in order to ensure stability as suggested in [28] (see also [7] and [18] for applications of this approach). An alternative is to adapt the step sizes that control the growth of the iterates ([19]). Another idea is to replace the single draw in (1.1) by a Monte Carlo sum over many realizations of X_{n+1} ([39]). Such modifications usually require quite precise understanding of the properties of the system in order to be implemented efficiently. The values of θ_n may simply be constrained to take values in a compact set K [23, 24]. The choice of the constraint set K requires prior information about the stationary points of h, as ill-chosen constraint set K may even lead to spurious convergence on the boundary of K. It is possible to modify the projection approach by constraining θ_n to take values in compact sets K_n, which eventually cover the whole parameter space $\cup_n K_n = \Theta$ [1]. In the controlled Markov chain setup, this approach requires relatively good control on the ergodic properties of the related Markov kernels near the 'boundary' of Θ [5].

We focus on the self-stabilized stochastic approximation algorithm with controlled Markov chains [3], which is based on truncations on random varying sets as suggested in [13]. The main difference to the expanding projections approach of [1, 5] is the occasional 'restart' of the process; see Section 2. The main advantages of this algorithm include that it does not introduce spurious convergence as projections to fixed set, it provides automatic calibration of the step size, but it does not require precise control of the behavior of the system near the boundary of Θ like the expanding projections and the averaging approaches. The convergence properties of the algorithm under the controlled Markov chain setup is studied in [3] and in a different setup in [27]. This algorithm has been used in various applications, including adaptive Monte Carlo [25] and adaptive Markov chain Monte Carlo [2].

Theorem 2.1 provides sufficient conditions implying that the number of truncations is finite almost surely. Therefore, this stabilized algorithm follows the equation (1.1) after some random (but almost surely finite) number of iterations. We then prove the almost sure convergence of $\{\theta_n, n \in \mathbb{N}\}$ to a connected component of a limiting set which contains the roots of h. We also provide a new set of sufficient conditions for the almost sure convergence of a SA sequence which weakens the conditions used in earlier contributions (see Proposition 4.6). We illustrate our results for (adaptive) quantile and multidimensional median approximation. We also analyze a penalized version of the 0-neighbors Kohonen algorithm.

The paper is organized as follows: the stabilized stochastic approximation algorithm, the main assumptions and the convergence result are given in Section 2. Section 3 is devoted to applications. The proofs are postponed to Sections 4 and 5.

2. Main results. Let $\{\mathcal{K}_q, q \in \mathbb{N}\}$ be a sequence of compact subsets of Θ such that

$$\bigcup_{i\geq 0} \mathcal{K}_i = \Theta, \quad \text{and} \quad \mathcal{K}_i \subset \operatorname{int}(\mathcal{K}_{i+1}), \quad i \geq 0, \tag{2.1}$$

where $\operatorname{int}(A)$ denotes the interior of the set A. The stable algorithm, described in Algorithm 2, proceeds as follows. We first run $\mathtt{SA}(\boldsymbol{\gamma}, \mathcal{K}_0, x, \theta)$ (see Algorithm 1) until the first time instant for which $\theta_n \notin \mathcal{K}_0$. When it occurs, *(i)* the active set is replaced with a larger one $\mathcal{K}_1$, *(ii)* the stepsize sequence $\boldsymbol{\gamma}$ is shifted and replaced with the sequence $\boldsymbol{\gamma}^{\leftarrow 1} \stackrel{\text{def}}{=} \{\gamma_{1+k}, k \in \mathbb{N}\}$ and *(iii)* $\mathtt{SA}(\boldsymbol{\gamma}^{\leftarrow 1}, \mathcal{K}_1, x, \theta)$ is run. The above procedure is repeated until convergence; see e.g. [3]).

Input: A positive sequence $\boldsymbol{\rho} = \{\rho_n, n \in \mathbb{N}\}$;
A point $(x, \theta) \in \mathsf{X} \times \Theta$;
A subset $\mathcal{K} \subseteq \Theta$
Initialization: $(X_0, \theta_0) = (x, \theta)$, $n = 0$;
repeat
  Draw $X_{n+1} \sim P_{\theta_n}(X_n, \cdot)$;
  $\theta_{n+1} = \theta_n + \rho_{n+1} H_{\theta_n}(X_{n+1})$;
  $n \leftarrow n + 1$
until $\theta_n \notin \mathcal{K}$;

Algorithm 1: Stochastic Approximation algorithm $\mathtt{SA}(\boldsymbol{\rho}, \mathcal{K}, x, \theta)$

Input : A positive sequence $\boldsymbol{\gamma} = \{\gamma_n, n \in \mathbb{N}\}$ and a $(x, \theta) \in \mathsf{X} \times \Theta$;
Output: The sequence $\{(X_n, \theta_n, I_n), n \in \mathbb{N}\}$
$(X_0, \theta_0, I_0) = (x, \theta, 0)$;
$n = 0$; `/*` n`: number of iterations */`
$\zeta_0 = 0$; `/*` ζ`: number of iterations in the current active set */`
repeat
  if $\zeta_n = 0$ **then**
    $X_{n+1/2} = x$, $\theta_{n+1/2} = \theta$;
  else
    $X_{n+1/2} = X_n$, $\theta_{n+1/2} = \theta_n$;
  end
  Draw $X_{n+1} \sim P_{\theta_{n+1/2}}(X_{n+1/2}, \cdot)$;
  $\theta_{n+1} = \theta_{n+1/2} + \gamma_{I_n + \zeta_n + 1} H_{\theta_{n+1/2}}(X_{n+1})$;
  if $\theta_{n+1} \in \mathcal{K}_{I_n}$ **then**
    $I_{n+1} = I_n$, $\zeta_{n+1} = \zeta_n + 1$
  else
    $I_{n+1} = I_n + 1$, $\zeta_{n+1} = 0$
  end
  $n \leftarrow n + 1$
until *convergence of the sequence* $\{\theta_n, n \in \mathbb{N}\}$;

Algorithm 2: A Stable Stochastic Approximation algorithm

Consider the following assumptions:

H1 *The function* $(x, \theta) \mapsto H_\theta(x)$ *from* $\mathsf{X} \times \Theta$ *to* $\mathbb{R}^d$ *is measurable. There exists a measurable function* $W : \mathsf{X} \to [1, \infty)$ *such that for any compact set* $\mathcal{K} \subset \Theta$, $\sup_{\theta \in \mathcal{K}} |H_\theta|_W < \infty$ *where* $|f|_W \stackrel{\text{def}}{=} \sup_{\mathsf{X}} |f|/W$.

H2 *(a) For any* θ *in* Θ, *the kernel* P_θ *has a unique invariant distribution* π_θ.

(b) For any compact $\mathcal{K} \subseteq \Theta$, *there exist constants* $C > 0$ *and* $\lambda \in (0, 1)$ *such that for any* $x \in \mathsf{X}$, $l \geq 0$, $\sup_{\theta \in \mathcal{K}} \|P_\theta^l(x, .) - \pi_\theta\|_W \leq C \lambda^l W(x)$ *and* $\sup_{\theta \in \mathcal{K}} \pi_\theta(W) < \infty$, *where* $\|\mu\|_W \stackrel{\text{def}}{=} \sup_{\{f, |f|_W \leq 1\}} |\mu(f)|$.

(c) There exists $p > 1$ *and for any compact set* $\mathcal{K} \subset \Theta$, *there exist constants* $\varrho \in (0, 1)$ *and* $b < \infty$ *such that* $\sup_{\theta \in \mathcal{K}} P_\theta W^p(x) \leq \varrho W^p(x) + b$.

When $P_\theta = P$ for any θ, sufficient conditions for H2-(b) are given in [29, Chapters

10 and 15]: they are mainly implied by the drift condition H2-(c) assuming the level sets $\{W \leq M\}$ are petite [16, Lemma 2.3].

We also introduce an assumption on the smoothness-in-θ of the transition kernels $\{P_\theta, \theta \in \Theta\}$. Let $\langle \cdot, \cdot \rangle$ denote the usual scalar product in $\mathbb{R}^d$ and $|\cdot|$ denote the associated norm. Denote

$$D_W(\theta, \theta') \stackrel{\mathrm{def}}{=} \sup_{x \in \mathsf{X}} \frac{\|P_\theta(x, \cdot) - P_{\theta'}(x, \cdot)\|_W}{W(x)} . \tag{2.2}$$

H3 *There exists $\upsilon \in (0, 1]$ such that for any compact $\mathcal{K} \subset \Theta$,*

$$\sup_{\theta, \theta' \in \mathcal{K}} \frac{D_W(\theta, \theta')}{|\theta - \theta'|^\upsilon} < \infty .$$

In [3, Section 6] it is assumed that there exists $\alpha \in (0, 1]$ such that for any compact set $\mathcal{K}$, $\sup_{\theta_1, \theta_2 \in \mathcal{K}} |\theta_1 - \theta_2|^{-\alpha} \; |H_{\theta_1} - H_{\theta_2}|_W < \infty$. We consider here a weaker condition.
H4 *Let $\alpha \in (0, 1]$. For any compact set $\mathcal{K} \subseteq \Theta$, there exists a constant $C > 0$ such that for all $\delta > 0$,*

$$\sup_{\theta \in \mathcal{K}} \int \pi_\theta(\mathrm{d}x) \sup_{\{\theta' \in \mathcal{K}, |\theta' - \theta| \leq \delta\}} |H_{\theta'}(x) - H_\theta(x)| \leq C \, \delta^\alpha .$$

Denote by h the mean field

$$h(\theta) \stackrel{\mathrm{def}}{=} \int \pi_\theta(\mathrm{d}x) \; H_\theta(x) . \tag{2.3}$$

The following assumption is classical in stochastic approximation theory (see for example [9, Part II, Section 1.6], or [10, Section 3.3], [26]).
H5 *There exists a continuously differentiable function $w : \Theta \to [0, \infty)$ such that*

(a) *For any $M > 0$, the level set $\{\theta \in \Theta, w(\theta) \leq M\}$ is a compact set of Θ.*
(b) *The set $\mathcal{L}$ of stationary points, defined by*

$$\mathcal{L} \stackrel{\mathrm{def}}{=} \{\theta \in \Theta, \langle \nabla w(\theta), h(\theta) \rangle = 0\} , \tag{2.4}$$

is compact.
(c) *For any $\theta \in \Theta \setminus \mathcal{L}$, $\langle \nabla w(\theta), h(\theta) \rangle < 0$.*

Note that under H2, H3 and H4, h is Hölder-continuous on Θ (see Lemma 4.9 below). We finally provide conditions on the stepsize sequence $\boldsymbol{\gamma} = \{\gamma_n, n \in \mathbb{N}\}$.
H6 $\boldsymbol{\gamma} = \{\gamma_0/(n+1)^\beta, n \in \mathbb{N}\}$ *with $\gamma_0 > 0$ and*

$$\beta \in \left(\frac{1}{p} \vee \frac{1 + (\alpha \wedge \upsilon)/p}{1 + (\alpha \wedge \upsilon)} ; 1 \right] ,$$

where υ and α are respectively defined in H3 and H4.
We denote by $\overline{\mathbb{P}}_{x,\theta,i}$ (resp. $\overline{\mathbb{E}}_{x,\theta,i}$) the canonical probability (resp. the canonical expectation) associated to the process $\{(X_n, \theta_n, I_n), n \in \mathbb{N}\}$ defined by Algorithm 2 when $(X_0, \theta_0, I_0) = (x, \theta, i)$. The main results of this contribution is summarized in the following theorem which shows that

(i) the number of updates of the active set is finite almost surely;
(ii) the process converges to the set of stationary points.

Theorem 2.1 *Let $\{\mathcal{K}_n, n \in \mathbb{N}\}$ be a compact sequence satisfying (2.1) and $(x_\star, \theta_\star) \in \mathsf{X} \times \mathcal{K}_0$. Assume H1 to H6. The sequence $\{(X_n, \theta_n), n \in \mathbb{N}\}$ given by Algorithm 2 started from $(x_\star, \theta_\star)$ is stable:*

$$\overline{\mathbb{P}}_{x_\star,\theta_\star,0}\left(\bigcup_{i\geq 0}\bigcap_{k\geq 0}\{\theta_k \in \ \mathcal{K}_i\}\right) = 1. \tag{2.5}$$

If in addition, one of the following assumptions holds
(i) $w(\mathcal{L})$ has an empty interior,
(ii) ∇w is locally Lipschitz on Θ, H2-(c) is satisfied with $p \geq 2$, and H6 is strengthened with the condition $\beta > 1/2$,
then the sequence $\{\theta_k, k \geq 0\}$ converges to a connected component of $\mathcal{L}$:

$$\overline{\mathbb{P}}_{x_\star,\theta_\star,0}\left(\lim_{k\to\infty} \mathrm{d}(\theta_k, \mathcal{L}) = 0, \lim_k |\theta_{k+1} - \theta_k| = 0\right) = 1, \tag{2.6}$$

where $\mathrm{d}(x, A)$ denotes the distance from x to the set A.

Proof. The proof is postponed to Section 4. □

3. Examples. For any $x \in \mathbb{R}^d$ and any $r > 0$, we define $\mathcal{B}(x, r) \stackrel{\text{def}}{=} \{y \in \mathbb{R}^d, |y - x| \leq r\}$.

3.1. Quantile estimation. Let P be a Markov kernel on $\mathsf{X} \subseteq \mathbb{R}^d$ having a stationary distribution π. Let $\phi : \mathsf{X} \to \mathbb{R}$ be a measurable function. We want to compute the quantile $q \in (0, 1)$ under π of the random variable $\phi(X)$. Quantile estimation has been considered in [14, Chapter 1]; more refined algorithms can also be found in [7, 15]. We consider the stochastic approximation procedure $\theta_{n+1} = \theta_n + \gamma_{n+1} H_{\theta_n}(X_{n+1})$ where

$$H_\theta(x) \stackrel{\text{def}}{=} q - \mathbb{1}_{\{\phi(x)\leq\theta\}}, \tag{3.1}$$

and $\{X_k, k \geq 0\}$ is a Markov chain with Markov kernel P. In this example, the Markov kernel is kept fixed *i.e.* $P_\theta = P$ and $\pi_\theta = \pi$ for all $\theta \in \mathbb{R}$

Proposition 3.1 *Assume that the push-forward measure of π by ϕ has a density w.r.t. the Lebesgue measure on $\mathbb{R}$, bounded on $\mathbb{R}$ and $\int |\phi(y)|\, \pi(y)\mathrm{d}y < \infty$. Assume also that H2 is satisfied with $P_\theta = P$ and $\pi_\theta = \pi$ for any θ. Then H1, H4 and H5 are satisfied with $\alpha = 1$ and w, $\mathcal{L}$ given by*

$$w(\theta) = \frac{1}{2}\int |\theta - \phi(y)|\, \pi(y)\, \mathrm{d}y + \left(\frac{1}{2} - q\right)\theta, \tag{3.2}$$

$$\mathcal{L} = \{\theta \in \mathbb{R} : \mathbb{P}(\phi(y) \leq \theta) = q\}. \tag{3.3}$$

Furthermore, $w(\mathcal{L})$ has an empty interior.

Proof. The proof is postponed to Section 5.1. □

Therefore, by Theorem 2.1, Algorithm 2 applied with a sequence $\{\gamma_n, n \geq 0\}$ satisfying H6, provides a sequence $\{\theta_n, n \geq 0\}$ converging almost-surely to the quantile of order q of $\phi(X)$ when $X \sim \pi$.

3.2. Stochastic Approximation Cross-Entropy (SACE) algorithm. Let $q \in (0, 1)$, $\mathsf{X} \subseteq \mathbb{R}^d$ and p be a density on X w.r.t. the Lebesgue measure. The goal is to find the q-th quantile θ of $\phi(X)$, *i.e.* θ such that $\int \mathbb{1}_{\{\phi(x)\geq\theta\}} p(x)\mathrm{d}x = 1 - q$.

We are particularly interested in extreme quantiles, *i.e.* $q \approx 1$ for which plain Monte Carlo methods are not efficient. We consider an approach combining MCMC and the cross-entropy method (see e.g. [21, Chapter 13]). Let $\mathcal{P} = \{g_\nu, \nu \in \mathcal{V} \subseteq \mathbb{R}^v\}$ be a parametric family of distributions w.r.t. the Lebesgue measure on $\mathsf{X} \subseteq \mathbb{R}^d$. The importance sampling estimator amounts to compute, for a given value of θ,

$$n^{-1} \sum_{i=1}^{n} \mathbb{1}_{\{\phi(Z_i) \geq \theta\}} w_\nu(Z_i), \tag{3.4}$$

where $\{Z_i, i \in \mathbb{N}\}$ is an i.i.d. sequence distributed under the instrumental distribution g_ν and $w_\nu(z) = p(z)/g_\nu(z)$ is the *importance weight function*. The choice of the parameter ν is of course critical to reduce the variance of the estimator. The optimal importance sampling distribution, also called the zero-variance importance distribution, is proportional to $\mathbb{1}_{\{\phi(z) \geq \theta\}} p(z)$. Note that the optimal sampling distribution is known up to a normalizing constant, which is the tail probability of interest.

The cross-entropy method amounts to choose the parameter ν by minimizing the Kullback-Leibler divergence of g_ν from the optimal importance distribution, or equivalently choose $\nu = \hat{\nu}$ with

$$\hat{\nu} = \mathrm{argmax}_{\nu \in \mathcal{V}} \int \log g_\nu(y) \mathbb{1}_{\{\phi(y) \geq \theta\}} p(y) \mathrm{d}y. \tag{3.5}$$

This integral is not directly available but can be approximated by Markov Chain Monte Carlo,

$$\hat{\nu} = \mathrm{argmax}_{\nu \in \mathcal{V}} \frac{1}{m} \sum_{i=1}^{m} \log g_\nu(Y_i) \tag{3.6}$$

where $\{Y_i, i \in \mathbb{N}\}$ is a Markov chain with transition kernel Q_θ where Q_θ has stationary density $p_\theta(z) \propto \mathbb{1}_{\{\phi(z) \geq \theta\}} p(z)$.

In the sequel, it is assumed that $\mathcal{P}$ is a canonical exponential family, *i.e.* there exist measurable functions $S : \mathsf{X} \to \mathbb{R}^v$, $A : \mathbb{R}^d \to \mathbb{R}^+$, $B : \mathbb{R}^v \to \mathbb{R}$ such that, for all $\nu \in \mathcal{V}$ and for all $x \in \mathsf{X}$,

$$g_\nu(x) = A(x) \exp\left(B(\nu) + \langle \nu, S(x) \rangle\right).$$

In such a case, solving the optimization problem (3.6) amounts to estimate the sufficient statistics $\bar{S}_m = m^{-1} \sum_{i=1}^{m} S(Y_i)$ and then to compute the maximum of $\nu \mapsto B(\nu) + \langle \nu, \bar{S}_m \rangle$. We assume that for any $s \in \mathbb{R}^v$, the function $\nu \mapsto B(\nu) + \langle \nu, s \rangle$ admits a unique maximum on $\mathcal{V}$ denoted by $\hat{\nu}(s)$. When estimating the quantile, the value of θ is not known a priori, and the above process should be used several times for different values of θ, which may be cumbersome.

In the Stochastic Approximation version of the Cross Entropy (SACE algorithm), we replace the Monte Carlo approximations (3.4) and (3.6) by stochastic approximations. Given a sequence of step-sizes $\{\gamma_n, n \geq 0\}$ and a family of MCMC kernels $\{Q_\theta, \theta \in \mathbb{R}\}$ such that Q_θ admits $p(x)\mathbb{1}_{\{\phi(x) \geq \theta\}}$ as unique invariant distribution, the SACE algorithm proceeds as follows

This algorithm can be casted into the stochastic approximation form $\vartheta_{n+1} = \vartheta_n + \gamma_{n+1} H_{\vartheta_n}(X_{n+1})$, by setting

$$\vartheta_n = \begin{bmatrix} \theta_n \\ \sigma_n \end{bmatrix} \qquad X_n = \begin{bmatrix} Y_n \\ Z_n \end{bmatrix} \qquad H_{(\theta,\sigma)}(y, z) = \begin{bmatrix} q - \mathbb{1}_{\{\phi(z) < \theta\}} \; p(z)/g_{\hat{\nu}(\sigma)}(z) \\ S(y) - \sigma \end{bmatrix}.$$

Input : Initial values: $x \in \mathsf{X}$, $\nu_0 \in \mathcal{V}$, $\sigma_0 \in S(\mathsf{X})$
Output: The sequence $\{\theta_n, n \in \mathbb{N}\}$
$n = 0$; `/* n: number of iterations */`
$\nu_0 = \hat{\nu}(\sigma_0)$;
$X_0 = x$;
repeat
  Conditionally to the past, draw independently $Y_{n+1} \sim Q_{\theta_n}(Y_n, \cdot)$ and $Z_{n+1} \sim g_{\nu_n}$;
  $\theta_{n+1} = \theta_n + \gamma_{n+1} \left(q - \mathbb{1}_{\{\phi(Z_{n+1}) < \theta_n\}} p(Z_{n+1})/g_{\nu_n}(Z_{n+1})\right)$;
  $\sigma_{n+1} = (1 - \gamma_{n+1})\sigma_n + \gamma_{n+1} S(Y_{n+1})$;
  $\nu_{n+1} = \hat{\nu}(\sigma_{n+1})$;
  $n \leftarrow n + 1$
until *convergence of* $\{\theta_n, n \in \mathbb{N}\}$;

Algorithm 3: SACE algorithm

It is easily seen from Algorithm 3 that $\{X_n, n \in \mathbb{N}\}$ is a controlled Markov chain: the conditional distribution of X_{n+1} given the past is $P_{\vartheta_n}(X_n, \cdot)$ where

$$P_{(\theta,\sigma)}((y,z), \mathrm{d}(y',z')) = Q_\theta(y, \mathrm{d}y')\, g_{\hat{\nu}(\sigma)}(z')\mathrm{d}z';$$

this kernel possesses a unique invariant distribution with density

$$\pi_{(\theta,\sigma)}(y,z) = p_\theta(y)\, g_{\hat{\nu}(\sigma)}(z)\,.$$

Therefore, the mean field function h is given by (up to a transpose)

$$(\theta, \sigma) \mapsto \left(\int \mathbb{1}_{\{\phi(z) \geq \theta\}} p(z) \mathrm{d}z - q, \ \int S(y) p_\theta(y) \mathrm{d}y - \sigma \right).$$

We establish that SACE satisfies H4 in the case $\mathcal{P}$ and π satisfy the following assumptions

E1
1. *There exists $\alpha \in (0,1]$ and for any compact set $\mathcal{K}$ of $S(\mathsf{X})$, there exists a constant C such that $|B(\hat{\nu}(\sigma)) - B(\hat{\nu}(\sigma'))| + |\hat{\nu}(\sigma) - \hat{\nu}(\sigma')| \leq C\,|\sigma - \sigma'|^\alpha$ for any $\sigma, \sigma' \in \mathcal{K}$.*
2. *The push-forward distribution of p by ϕ possesses a bounded density w.r.t. the Lebesgue measure on $\mathbb{R}$. In addition, there exists $\delta > 0$ such that*

$$\int (1 + |S(x)|) \exp\left(1 + \delta\,|S(x)|\right)\, p(x) \mathrm{d}x < \infty.$$

Proposition 3.2 *Assume E1. Then H4 holds with α given by E1.*

Proof. The proof is postponed to Section 5.2. □

Consider a bridge network: the network is composed with nodes and d edges $\{e_1, \cdots, e_d\}$ with length $\{U_\ell, \ell \leq d\}$. Fix two nodes N_1, N_2 in the graph; we are interested in the length of the shortest path from N_1 to N_2 defined by

$$\phi(U_1, \cdots, U_d) = \min_{\mathcal{C}} \sum_{\ell\, \mathrm{s.t.}\, e_\ell \in \mathcal{C}} \mathsf{a}_\ell\, U_\ell$$

where $\mathcal{C}$ denotes a path from N_1 to N_2 ($\mathcal{C}$ is a set of edges) and $\mathsf{a}_\ell > 0$ (see Figure 1). It is assumed that the lengths $\{U_\ell, \ell \leq d\}$ are i.i.d. and uniformly distributed on $[0,1]$.

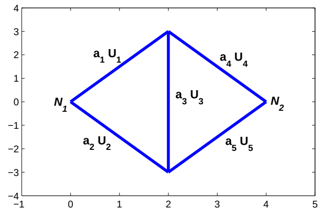

FIG. 1. *A bridge network, in the case* $d = 5$

We are interested in computing a threshold $\theta_\star$ such that the probability $\phi(U_1, \cdots, U_d)$ exceeds $\theta_\star$ is $1-q$ in the case q is close to one. In this example,

$$\mathsf{X} = [0,1]^d, \qquad p(u_1, \cdots, u_d) = \prod_{\ell=1}^{d} \mathbb{1}_{\{[0,1]\}}(u_\ell).$$

The importance sampling distribution g_ν is a product of Beta$(\nu_\ell, 1)$ distributions:

$$\mathcal{V} = \left(\mathbb{R}^+ \setminus \{0\}\right)^d, \qquad g_\nu(u_1, \cdots, u_d) = \prod_{\ell=1}^{d} \nu_\ell \, (u_\ell)^{\nu_\ell - 1} \, \mathbb{1}_{\{[0,1]\}}(u_\ell).$$

g_ν is from a canonical exponential family with $S(u_1, \cdots, u_d) = (\ln u_\ell)_{1 \leq \ell \leq d}$ and $B(\nu_1, \cdots, \nu_d) = \sum_{\ell=1}^{d} \ln(\nu_\ell)$. Furthermore, for any $1 \leq \ell \leq d$, $(\hat{\nu}(s_1, \cdots, s_d))_\ell = (-s_\ell)^{-1}$.

In this example, $\Theta = \mathbb{R} \times \prod_{\ell=1}^{d} (-\infty, 0)$. The assumptions E1 are easily verified with $\alpha = 1$ (details are omitted). For the MCMC samplers $\{Q_\theta, \theta \in \mathbb{R}\}$, we use a Gibbs sampler: note that for any $\ell \in \{1, \cdots, d\}$ and any $\{u_j, j \neq \ell\}$, $u_\ell \mapsto \phi(u_1, \cdots, u_d)$ is increasing. Hence, the conditional distribution of the ℓ-th variable conditionally to the others when the joint distribution is proportional to $p(x)\mathbb{1}_{\{\phi(x) \geq \theta\}}$ is a uniform distribution on $[(\theta - \phi(u^*))_+, 1]$ where $u^* = (u_1, \cdots, u_{\ell-1}, 0, u_{\ell+1}, \cdots, u_d)$.

We illustrate the convergence of SACE for the bridge network displayed on Figure 1 in the case $[\mathsf{a}_1, \cdots, \mathsf{a}_d] = [1, 2, 3, 1, 2]$. On Figure 2[left], we show a path of $\{\theta_n, n \geq 0\}$ for different runs of the algorithm corresponding to the different quantiles q. On Figure 2[right], we show the path of the d components of $\hat{\nu}(\sigma_n)$ in the case $q = 0.001$. Not surprisingly, the largest values of the parameters $\hat{\nu}(\sigma_n)_\ell$ at convergence are reached with $\ell = 1, 4$; they correspond to the shortest range of path length ($a_1 = a_4 = 1$).

3.3. Median in multi-dimensional spaces. In a multivariate setting different extensions of the median have been proposed in the literature (see for instance [11] and [12] and the references therein). We focus here on the spatial median, also named geometric median which is probably the most frequently used. The median θ of a random vector X taking values in $\mathbb{R}^d$ with $d \geq 2$ is

$$\theta = \arg\min_{m \in \mathbb{R}^d} \mathbb{E}[|X - m|]$$

The median θ is uniquely defined unless the support of the distribution of X is concentrated on a one dimensional subspace of $\mathbb{R}^d$. Note also that it is translation invariant.

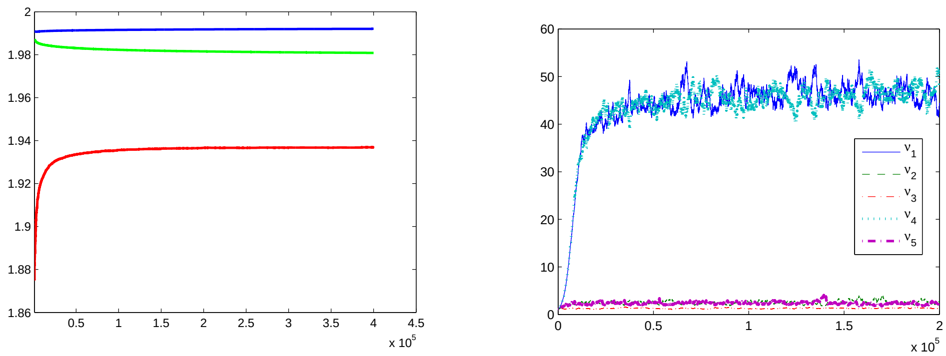

FIG. 2. *[left] Paths of θ_n for different values of q ($q = 10^{-3}, 10^{-4}$ and 10^{-5}) - the first* 1000 *iterations are discarded. [Right] Path of the d components of $\hat{\nu}(\sigma_n)$, when $q = 0.001$*

Following [11], we consider the stochastic approximation procedure with

$$H_\theta(x) \stackrel{\text{def}}{=} \frac{x-\theta}{|x-\theta|} \mathbb{1}_{\{x \neq \theta\}}, \tag{3.7}$$

and the following assumptions

E2 *the condition H2 is satisfied with $\pi_\theta = \pi$ for any θ. The density π is bounded on $\mathbb{R}^d$ and $\int |x|\, \pi(x)\mathrm{d}x < \infty$.*

Proposition 3.3 *Assume E2. Let $\theta_\star$ be the unique solution of $\int H_\theta(x)\pi(x)\mathrm{d}x = 0$. Then H1, H4 and H5 are satisfied with w and $\mathcal{L}$ given by*

$$w(\theta) = \int |x-\theta|\, \pi(x)\mathrm{d}x, \qquad \mathcal{L} = \{\theta_\star\},$$

and with $\alpha = d/(1+d)$.

Proof. The proof is postponed to Section 5.3. □

Here again, $w(\mathcal{L})$ has an empty interior; and Theorem 2.1 provides sufficient conditions on the kernels $\{P_\theta, \theta \in \Theta\}$ and on the sequence $\{\gamma_n, n \geq 0\}$ implying that $\{\theta_n, n \geq 0\}$ converges almost-surely to $\theta_\star$.

3.4. Vector quantization. Vector quantization consists of approximating a random vector X in $\mathbb{R}^d$ by a random vector taking at most N values in $\mathbb{R}^d$. In this section, we assume that

E3 *the distribution of X is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$ with density π having a bounded support: $\pi(x) = 0$ for any $|x| > \Delta$ for some $\Delta > 0$.*

Vector quantization plays a crucial role in source coding [38], numerical integration [31, 33] and nonlinear filtering [32, 34]. Several stochastic approximations procedure have been proposed to approximate the optimal quantizer; see [20, 8]. For $\theta = (\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(N)}) \in (\mathbb{R}^d)^N$, and for any $1 \leq i \leq N$, define the Voronoi cells associated to the dictionary θ by

$$C^{(i)}(\theta) \stackrel{\text{def}}{=} \left\{ u \in \mathbb{R}^d, \left|u - \theta^{(i)}\right| = \min_{1 \leq j \leq N} \left|u - \theta^{(j)}\right| \right\}.$$

These cells allow to approximate a random vector X by $\theta^{(i)}$ in the cell $C^{(i)}(\theta)$. Denote by $\widetilde{w}$ the mean squared quantization error (or distortion) given by

$$\widetilde{w}(\theta) \stackrel{\mathrm{def}}{=} \sum_{i=1}^{N} \mathbb{E}_\pi \left[\left| X - \theta^{(i)} \right|^2 \mathbb{1}_{\{C^{(i)}(\theta)\}}(X) \right].$$

The Kohonen algorithm (with 0 neighbors) is a stochastic approximation algorithm with field $\widetilde{H} : (\mathbb{R}^d)^N \times \mathbb{R}^d \to (\mathbb{R}^d)^N$ given by

$$\widetilde{H_\theta}(u) \stackrel{\mathrm{def}}{=} -2 \left((\theta^{(i)} - u) \mathbb{1}_{\{C^{(i)}(\theta)\}}(u) \right)_{1 \le i \le N}.$$

The convergence of the Kohonen algorithm has been established in dimension $d = 1$ for i.i.d. observations $\{X_n, n \in \mathbb{N}\}$ [31]. The case $d \ge 2$ is still an open question: one of the main difficulty arises from the non-coercivity of the distortion $\widetilde{w}$ and the non-smoothness property of the field $\widetilde{H_\theta}(x)$. The goal here is to go a step further in the study of the multidimensional case, including a generalization to the case the observations $\{X_k, k \ge 1\}$ are Markovian (see e.g. [33, 34]):

E4 *$\{P_\theta, \theta \in \Theta\}$ is a family of kernels satisfying H2 for some $p \ge 2$, H3 and such that $\pi_\theta = \pi$ for any $\theta \in \Theta$.*

We define

$$\Theta = \left\{ \theta = (\theta^{(1)}, \dots \theta^{(N)}) \in (\mathbb{R}^d)^N \cap (\mathcal{B}(0, \Delta))^N, \theta^{(i)} \neq \theta^{(j)} \text{ for all } i \neq j \right\};$$

and run Algorithm 2 with $H_\theta \in \mathbb{R}^{dN}$ defined by

$$H_\theta(x) = \widetilde{H_\theta}(x) - \lambda \left(\sum_{j \neq i} \frac{\theta^{(i)} - \theta^{(j)}}{\left| \theta^{(i)} - \theta^{(j)} \right|^4} \right)_{1 \le i \le N} \tag{3.8}$$

for some $\lambda > 0$ with the sequence of compact sets $\mathcal{K}_q = \{\theta \in \Theta : \min_{i \neq j} \left| \theta^{(i)} - \theta^{(j)} \right| \ge 1/q\}$. Set

$$w(\theta) = \widetilde{w}(\theta) + \frac{\lambda}{4} \sum_{i \neq j} \left| \theta^{(i)} - \theta^{(j)} \right|^{-2}. \tag{3.9}$$

Under E3, from [31, Proposition 9 and Lemma 29], w is continuously differentiable on Θ and $\nabla w(\theta) = -\int H_\theta(x) \pi(x)\, \mathrm{d}x$. Furthermore, for any compact set $\mathcal{K}_q$, there exists a constant C such that for any $\theta \in \mathcal{K}_q$, $\bar{\theta} \in \mathcal{K}_q$, $\left| \nabla w(\theta) - \nabla w(\bar{\theta}) \right| \le C \left| \theta - \bar{\theta} \right|$. Hence, ∇w is locally Lipschitz on Θ.

Lemma 3.4 *Assume E3 and E4. Then H1, H4 and H5 are satisfied with w defined by (3.9) and $\mathcal{L} \stackrel{\mathrm{def}}{=} \{\theta \in \Theta : \nabla w(\theta) = 0\}$.*

Proof. The proof is postponed in Section 5.4. □

Theorem 2.1 shows that under E3, E4, the penalized 0-neighbors Kohonen algorithm converges a.s. to a connected component of $\mathcal{L}$.

4. Proofs. In Section 4.1, we start with preliminary results on the stability and the convergence of stochastic approximation schemes. We provide a new set of sufficient conditions for the convergence of stable SA algorithms (see Proposition 4.6). In Section 4.2, some properties of the Poisson equations associated to the transition kernels $\{P_\theta, \theta \in \Theta\}$ are discussed. In Section 4.3, we first state a control of the perturbations $H_{\theta_n}(X_{n+1}) - h(\theta_n)$ (see Proposition 4.11) which is the key ingredient for the proof of Theorem 2.1; we then conclude Section 4.3 by giving the proof of Theorem 2.1. The proof of Proposition 4.11 is given in Section 4.4.

4.1. Stability and Convergence of Stochastic Approximation algorithms. Let $\{\vartheta_n, n \geq 0\}$ be defined, for all $\vartheta_0 \in \Theta$ and $n \geq 1$ by:

$$\vartheta_n = \vartheta_{n-1} + \rho_n h(\vartheta_{n-1}) + \rho_n \xi_n, \tag{4.1}$$

where $\{\rho_n, n \in \mathbb{N}\}$ is a sequence of positive numbers and $\{\xi_n, n \in \mathbb{N}\}$ is a sequence of $\mathbb{R}^d$-vectors. For any $L \geq 0$, the sequence $\{\tilde{\vartheta}_{L,k}, k \geq 0\}$ is defined by $\tilde{\vartheta}_{L,0} = \vartheta_L$, and for $k \geq 1$,

$$\tilde{\vartheta}_{L,k} = \tilde{\vartheta}_{L,k-1} + \rho_{L+k} h(\vartheta_{L+k-1}). \tag{4.2}$$

Lemma 4.1 *Assume H5. Let M_0 be such that $\{\vartheta_0\} \cup \mathcal{L} \subset \{\vartheta : w(\vartheta) \leq M_0\}$. There exist $\delta_\star > 0$ and $\lambda_\star > 0$ such that for any non-increasing sequence $\{\rho_n, n \in \mathbb{N}\}$ of positive numbers and any $\mathbb{R}^d$-valued sequence $\{\xi_n, n \in \mathbb{N}\}$*

$$\left(\rho_0 \leq \lambda_\star \text{ and } \sup_{k \geq 1} \left| \sum_{j=1}^k \rho_j \xi_j \right| \leq \delta_\star \right) \Longrightarrow \left(\sup_{k \geq 1} w(\vartheta_k) \leq M_0 + 1 \right).$$

Proof. See [3, Theorem 2.2]. □

Lemma 4.2 *Let $g : \Theta \to \mathbb{R}$ be a continuous function. For any compact set $\mathcal{K} \subset \Theta$ and $\delta > 0$, there exists $\eta > 0$ such that for all $\vartheta \in \mathcal{K}$ and $\vartheta' \in \Theta$ satisfying $|\vartheta - \vartheta'| \leq \eta$, $|g(\vartheta) - g(\vartheta')| \leq \delta$.*

Lemma 4.3 *Assume H5-(c) and h is continuous. For any compact set $\mathcal{K}$ of Θ such that $\mathcal{K} \cap \mathcal{L} = \emptyset$ and any $\delta \in (0, \inf_{\mathcal{K}} |\langle \nabla w, h \rangle|)$, there exist $\lambda > 0, \beta > 0$ such that for all $\vartheta \in \mathcal{K}$, $\rho \leq \lambda$ and $|\xi| \leq \beta$, $w(\vartheta + \rho h(\vartheta) + \rho \xi) \leq w(\vartheta) - \rho \delta$.*

Proof. See [3, Lemma 2.1(i)]. □

Lemma 4.4 *Assume H5, h is continuous, $\lim_n \rho_n = 0$, $\lim_k \sum_{j=1}^k \rho_j \xi_j$ exists and there exists $M > 0$ such that for any $n \geq 0$, $\vartheta_n \in \mathcal{K} \stackrel{\text{def}}{=} \{\theta \in \Theta : w(\theta) \leq M\}$. Then*

(i) for any $L \geq 0$ and $k \geq 0$, $\vartheta_{L+k+1} - \tilde{\vartheta}_{L,k+1} = \sum_{j=L}^{L+k} \rho_{j+1} \xi_{j+1}$.

(ii) $\limsup_{n \to \infty} |\vartheta_{n+1} - \vartheta_n| = 0$.

(iii) for any $\widetilde{M} > M$, there exists $\tilde{L}$ such that for any $L \geq \tilde{L}$, $\{\tilde{\vartheta}_{L,k}, k \geq 0\} \subset \widetilde{\mathcal{K}} \stackrel{\text{def}}{=} \{\theta \in \Theta : w(\theta) \leq \widetilde{M}\}$.

(iv) $\lim_{L \to \infty} \sup_{l \geq 0} |w(\tilde{\vartheta}_{L,l}) - w(\vartheta_{L+l})| = 0$.

Proof.

(i) (4.2) implies that $\tilde{\vartheta}_{L,k+1} - \vartheta_{L+k+1} = \tilde{\vartheta}_{L,k} - \vartheta_{L+k} - \rho_{L+k+1} \xi_{L+k+1}$ from which the proof follows.

(ii) Under H5-(a) and the continuity of h on Θ, $\sup_{\mathcal{K}} |h| < +\infty$. Since $\vartheta_k \in \mathcal{K}$ for any k, we get $|\tilde{\vartheta}_{L,k+1} - \tilde{\vartheta}_{L,k}| \leq \rho_{L+k+1} \sup_{\mathcal{K}} |h|$. Let $\epsilon > 0$. Under the stated assumptions, we may choose K_ϵ such that for any $k \geq K_\epsilon$ and $L \geq 0$, $|\tilde{\vartheta}_{L,k} - \tilde{\vartheta}_{L,k-1}| \leq \epsilon$ and L_ϵ such that for any $L \geq L_\epsilon$, $\sup_{l \geq 1} |\sum_{j=L}^{L+l} \rho_{j+1} \xi_{j+1} \mathbb{1}_{\vartheta_j \in \mathcal{K}}| \leq \epsilon$. By (i), for any $k \geq K_\epsilon$ and $L = L_\epsilon$,

$$|\vartheta_{L+k+1} - \vartheta_{L+k}| \leq \left| \tilde{\vartheta}_{L,k+1} - \tilde{\vartheta}_{L,k} \right| + \left| \vartheta_{L+k+1} - \tilde{\vartheta}_{L,k+1} \right| + \left| \tilde{\vartheta}_{L,k} - \vartheta_{L+k} \right| \leq 3\epsilon.$$

(iii) $\widetilde{\mathcal{K}}$ is compact by H5-(a). By Lemma 4.2, there exists $\eta > 0$ such that for all $\vartheta \in \mathcal{K}$, $\vartheta' \in \Theta$, $|\vartheta - \vartheta'| \leq \eta$, $|w(\vartheta) - w(\vartheta')| \leq \widetilde{M} - M$. There exists $\tilde{L}$ such that for any $L \geq \tilde{L}$, $\sup_{l \geq 1} |\sum_{j=L}^{L+l} \rho_{j+1} \xi_{j+1} \mathbb{1}_{\vartheta_j \in \mathcal{K}}| \leq \eta$. By (i), $\sup_{L \geq \tilde{L}} \sup_{k \geq 0} |\tilde{\vartheta}_{L,k} - \vartheta_{L+k}| \leq \eta$. Since $\vartheta_j \in \mathcal{K}$ for any $j \geq 0$, this implies that for any $L \geq \tilde{L}$ and $k \geq 0$, $w(\tilde{\vartheta}_{L,k}) \leq \widetilde{M}$ and $\tilde{\vartheta}_{L,k} \in \widetilde{\mathcal{K}}$.

(iv) The proof is on the same lines as the proof of (iii).

□

Lemma 4.5 *Let $\{v_n, n \geq 0\}$ and $\{\chi_n, n \geq 0\}$ be non-negative sequences and $\{\eta_n, n \geq 0\}$ be a sequence such that $\sum_n \eta_n$ exists. If for any $n \geq 0$, $v_{n+1} \leq v_n - \chi_n + \eta_n$ then $\sum_n \chi_n < \infty$ and $\lim_n v_n$ exists.*

Proof. Set $w_n = v_n + \sum_{k \geq n} \eta_k + M$ with $M \stackrel{\text{def}}{=} -\inf_n \sum_{k \geq n} \eta_k$ so that $\inf_n w_n \geq 0$. Then

$$0 \leq w_{n+1} \leq v_n - \chi_n + \eta_n + \sum_{k \geq n+1} \eta_k + M \leq w_n - \chi_n .$$

Hence, the sequence $\{w_n, n \geq 0\}$ is non-negative and non-increasing; therefore it converges. Furthermore, $0 \leq \sum_{k=0}^n \chi_k \leq w_0$ so that $\sum_n \chi_n < \infty$. Therefore, the convergence of $\{w_n, n \geq 0\}$ also implies the convergence of $\{v_n, n \geq 0\}$. This concludes the proof. □

Proposition 4.6 *Assume H 5. Let $\{\rho_n, n \in \mathbb{N}\}$ be a non-increasing sequence of positive numbers and $\{\xi_n, n \in \mathbb{N}\}$ be a sequence of $\mathbb{R}^d$-vectors. Assume*

(C-i) $h : \Theta \to \mathbb{R}^d$ *is continuous.*

(C-ii) $\{\vartheta_k, k \in \mathbb{N}\} \subset \mathcal{K} \stackrel{\text{def}}{=} \{\theta \in \Theta : w(\theta) \leq M\}$.

(C-iii) $\sum_k \rho_k = +\infty$ *and* $\lim_k \rho_k = 0$.

(C-iv) $\lim_{k \to \infty} \sum_{j=1}^k \rho_j \xi_j$ *exists.*

(C-v) one of the following conditions

- *(A)* $w(\mathcal{L})$ *has an empty interior*
- *(B)* ∇w *is locally Lipschitz on* Θ, *and the series* $\sum \rho_j \langle \nabla w(\vartheta_j), \xi_j \rangle$ *and* $\sum_j \rho_j^2 |\xi_j|^2$ *converge.*

Then $\{\vartheta_n, n \geq 0\}$ converges to a connected component of $\mathcal{L}$.

Proof. Assume first that $\lim_{n \to \infty} w(\vartheta_n)$ exists.

▶ `Step 1.` For $\alpha > 0$, let $\mathcal{L}_\alpha \stackrel{\text{def}}{=} \{\theta \in \Theta : \mathrm{d}(\theta, \mathcal{L}) < \alpha\}$ be the α-neighborhood of $\mathcal{L}$. Let $\epsilon > 0$. We prove that there exist L_ϵ and $\delta_1 > 0$ such that for any $L \geq L_\epsilon$,

(a) $\sup_{k \geq 0} \left| \tilde{\vartheta}_{L,k} - \vartheta_{L+k} \right| \leq \epsilon$ and $\sup_{k \geq 0} \left| w(\tilde{\vartheta}_{L,k}) - w(\vartheta_{L+k}) \right| \leq \epsilon$.

(b) for any $k \geq 0$, it holds: $\tilde{\vartheta}_{L,k} \notin \mathcal{L}_\alpha \Longrightarrow w(\tilde{\vartheta}_{L,k+1}) - w(\tilde{\vartheta}_{L,k}) \leq -\rho_{L+k+1}\delta_1$.

(c) the sequence $\{\tilde{\vartheta}_{L,k}, k \geq 0\}$ is infinitely often in $\mathcal{L}_\alpha$.

Let $\widetilde{M} > M$ and set $\widetilde{\mathcal{K}} \stackrel{\text{def}}{=} \{\theta \in \Theta : w(\theta) \leq \widetilde{M}\}$. Note that by H 5-(a), $\widetilde{\mathcal{K}} \subset \Theta$ is compact and since $\mathcal{L}_\alpha$ is open, $\widetilde{\mathcal{K}}^\alpha \stackrel{\text{def}}{=} \widetilde{\mathcal{K}} \setminus \mathcal{L}_\alpha$ is compact and $\widetilde{\mathcal{K}}^\alpha \cap \mathcal{L} = \emptyset$. By Lemma 4.3, there exist $\delta_1 > 0$, $\lambda_1 > 0$, $\beta_1 > 0$ such that

$$\left(\vartheta \in \widetilde{\mathcal{K}}^\alpha, \rho \leq \lambda_1, |\xi| \leq \beta_1 \right) \Longrightarrow w\left(\vartheta + \rho h(\vartheta) + \rho \xi\right) \leq w(\vartheta) - \rho \delta_1 . \tag{4.3}$$

By Lemma 4.2, there exists $\delta_2 \in (0, \epsilon)$ such that

$$(\vartheta \in \mathcal{K}, \vartheta' \in \Theta, |\vartheta - \vartheta'| \leq \delta_2) \Longrightarrow \left(|h(\vartheta) - h(\vartheta')| \leq \beta_1, \quad |w(\vartheta) - w(\vartheta')| \leq \epsilon \right). \tag{4.4}$$

By Lemma 4.4-(iii), (C-iii) and (C-iv), there exists $\widetilde{L}$ such that for any $L \geq \widetilde{L}$, $\{\tilde{\vartheta}_{L,k}, k \geq 0\} \subset \widetilde{\mathcal{K}}$ and

$$\sup_{k \geq 0} \rho_{L+k} \leq \lambda_1, \qquad \sup_{l \geq 1} \left| \sum_{j=L}^{L+l} \rho_{j+1} \xi_{j+1} \mathbb{1}_{\vartheta_j \in \mathcal{K}} \right| \leq \delta_2 . \tag{4.5}$$

(a) follows from Lemma 4.4-(i) and (4.5).

(b) Let $L \geq \tilde{L}$, $k \geq 0$ and $\tilde{\vartheta}_{L,k} \notin \mathcal{L}_\alpha$. The proof follows from (4.3) and

$$\tilde{\vartheta}_{L,k+1} = \tilde{\vartheta}_{L,k} + \rho_{L+k+1} h\left(\tilde{\vartheta}_{L,k}\right) + \rho_{L+k+1}\left(h\left(\vartheta_{L+k}\right) - h\left(\tilde{\vartheta}_{L,k}\right)\right),$$

using (4.4) and (4.5).

(c) The proof is by contradiction. Let $L \geq \widetilde{L}$, $k \geq 0$ and assume that for any $j \geq 0$, $\tilde{\vartheta}_{L,k+j} \in \widetilde{\mathcal{K}}^\alpha$. By (b), for any $j, k \geq 0$, $w\left(\tilde{\vartheta}_{L,k+j+1}\right) \leq w(\tilde{\vartheta}_{L,k+j}) - \rho_{L+k+j+1}\delta_1$ which implies under (C-iii) that $\lim_{j\to\infty} w(\tilde{\vartheta}_{L,k+j}) = -\infty$. Since w is continuous and nonnegative and $\mathcal{K}$ compact, $\inf_{\widetilde{\mathcal{K}}} w \geq 0$. This is a contradiction.

▶ `Step 2.` Let $\alpha > 0$ and $\epsilon > 0$. By Step 1 and Lemma 4.4-(iv), there exists L_ϵ such that for any $L \geq L_\epsilon$ and $k \geq 0$, $\sigma_{L,k} \stackrel{\text{def}}{=} \inf\{j \geq 0, \tilde{\vartheta}_{L,k+j} \in \mathcal{L}_\alpha\}$ is finite, $\sup_{k\geq 0}\left|\tilde{\vartheta}_{L,k} - \vartheta_{L+k}\right| \leq \epsilon$, $\sup_{k\geq 0}\left|w(\tilde{\vartheta}_{L,k}) - w(\vartheta_{L+k})\right| \leq \epsilon$ and

$$\delta_1 \sum_{\ell=1}^{\sigma_{L,k}} \rho_{L+k+\ell} \leq w(\tilde{\vartheta}_{L,k}) - w(\tilde{\vartheta}_{L,k+\sigma_{L,k}}).$$

Hence, for any $k \geq 0$, using $|\tilde{\vartheta}_{L,\ell} - \tilde{\vartheta}_{L,\ell-1}| \leq \rho_{L+\ell} \sup_{\mathcal{K}} |h|$,

$$\begin{aligned}
\mathrm{d}\left(\vartheta_{L+k}, \mathcal{L}_\alpha\right) &\leq \left|\vartheta_{L+k} - \tilde{\vartheta}_{L,k+\sigma_{L,k}}\right| \leq \left|\vartheta_{L+k} - \tilde{\vartheta}_{L,k}\right| + \left|\tilde{\vartheta}_{L,k} - \tilde{\vartheta}_{L,k+\sigma_{L,k}}\right| \\
&\leq \epsilon + \sum_{\ell=1}^{\sigma_{L,k}} \left|\tilde{\vartheta}_{L,k+\ell} - \tilde{\vartheta}_{L,k+\ell-1}\right| \leq \epsilon + \sup_{\mathcal{K}} |h| \sum_{\ell=1}^{\sigma_{L,k}} \rho_{L+k+\ell} \\
&\leq \epsilon + \delta_1^{-1} \sup_{\mathcal{K}} |h| \sup_{\ell\geq 0} \left|w(\tilde{\vartheta}_{L,k}) - w(\tilde{\vartheta}_{L,k+\ell})\right|, \\
&\leq \epsilon + 2\epsilon\delta_1^{-1} \sup_{\mathcal{K}} |h| + \delta_1^{-1} \sup_{\mathcal{K}} |h| \sup_{\ell\geq 0} \left|w(\vartheta_{L+k}) - w(\vartheta_{L+k+\ell})\right|.
\end{aligned}$$

This proves that $\lim_n \mathrm{d}(\vartheta_n, \mathcal{L}) = 0$ since $\lim_n w(\vartheta_n)$ exists. Finally, since by Lemma 4.4-(ii), $\lim_n |\vartheta_n - \vartheta_{n-1}| = 0$, the sequence $\{\vartheta_n, n \geq 0\}$ converges to a connected component of $\mathcal{L}$.

▶ `Step 3.` We now prove that $\lim_n w(\vartheta_n)$ exists. Under (C-v)-(A), the proof follows from [3, Theorem 2.3]. We prove that this limit exists under (C-v)-(B).

By H5-(a), (C-ii) and Lemma 4.4-(ii), there exist N and a compact set $\widetilde{\mathcal{K}}$ of Θ such that $\mathcal{K} \subseteq \widetilde{\mathcal{K}}$ and for any $n \geq N$ and $t \in [0,1]$, $\vartheta_n + t(\vartheta_{n+1} - \vartheta_n) \in \widetilde{\mathcal{K}}$. By (C-v)-(B), there exists a constant C such that for $\vartheta, \vartheta' \in \widetilde{\mathcal{K}}$, $|\nabla w(\vartheta) - \nabla w(\vartheta')| \leq C|\vartheta - \vartheta'|$ showing that, for $n \geq N$,

$$w(\vartheta_{n+1}) \leq w(\vartheta_n) + \langle \nabla w(\vartheta_n), \vartheta_{n+1} - \vartheta_n \rangle + C/2 \left|\vartheta_{n+1} - \vartheta_n\right|^2.$$

Using (4.1), we obtain

$$\begin{aligned}
&\langle \nabla w(\vartheta_n), \vartheta_{n+1} - \vartheta_n \rangle = \rho_{n+1} \langle \nabla w(\vartheta_n), h(\vartheta_n) \rangle + \rho_{n+1} \langle \nabla w(\vartheta_n), \xi_{n+1} \rangle \\
&|\vartheta_{n+1} - \vartheta_n|^2 \leq 2\rho_{n+1}^2 \left\{ \left(\sup_{\mathcal{K}} |h|\right)^2 + |\xi_{n+1}|^2 \right\}.
\end{aligned}$$

This yields for any $n \geq N$,

$$w(\vartheta_{n+1}) \leq w(\vartheta_n) - \rho_{n+1} \left|\langle \nabla w(\vartheta_n), h(\vartheta_n)\rangle\right| + \rho_{n+1} \langle \nabla w(\vartheta_n), \xi_{n+1}\rangle \\ + C\, \rho_{n+1}^2 \left(\left(\sup_{\mathcal{K}} |h| \right)^2 + |\xi_{n+1}|^2 \right).$$

Lemma 4.5 concludes the proof. □

4.2. Regularity in θ of the solution to the Poisson equation. Under the assumptions H1 and H2, for any $\theta \in \Theta$, there exists a function g solving the Poisson equation

$$g \mapsto H_\theta - \pi_\theta H_\theta = g - P_\theta g. \tag{4.6}$$

This solution, denoted by g_θ, is unique up to an additive constant and given by $g_\theta(x) \stackrel{\text{def}}{=} \sum_{n\geq 0} \{P_\theta^n H_\theta(x) - \pi_\theta H_\theta\}$. Finally, for any compact set $\mathcal{K}$ of Θ,

$$\sup_{\theta\in\mathcal{K}} |g_\theta|_W \leq \sup_{\theta\in\mathcal{K}} |H_\theta|_W \sum_{n\geq 0} \sup_{x\in\mathsf{X}} \frac{\sup_{\theta\in\mathcal{K}} \|P_\theta^n(x,\cdot) - \pi_\theta\|_W}{W(x)} < \infty. \tag{4.7}$$

Lemma 4.7 *Assume H2-(b). For any compact set $\mathcal{K} \subset \Theta$, there exists a constant C such that for any $\theta, \theta' \in \mathcal{K}$*

$$\sup_{n\geq 0} \sup_{x\in\mathsf{X}} \frac{\|P_\theta^n(x,.) - P_{\theta'}^n(x,.)\|_W}{W(x)} \leq C D_W(\theta,\theta'),$$

where D_W is defined by (2.2).

Proof. For any measurable function f such that $|f|_W \leq 1$, it holds

$$P_\theta^n f(x) - P_{\theta'}^n f(x) = \sum_{j=0}^{n-1} P_{\theta'}^j (P_\theta - P_{\theta'}) \left(P_\theta^{n-j-1} f(x) - \pi_\theta(f) \right).$$

For any $0 \leq j \leq n-1$,

$$\left| P_{\theta'}^j (P_\theta - P_{\theta'}) \left(P_\theta^{n-j-1} f(x) - \pi_\theta(f) \right) \right| \leq P_{\theta'}^j W(x) \left| (P_\theta - P_{\theta'}) \left(P_\theta^{n-j-1} f - \pi_\theta(f) \right) \right|_W \\ \leq D_W(\theta,\theta')\ P_{\theta'}^j W(x) \left| P_\theta^{n-j-1} f - \pi_\theta(f) \right|_W.$$

By H2-(b), there exist $C > 0$ and $\lambda \in (0,1)$ such that for any $\theta, \theta' \in \mathcal{K}$,

$$P_{\theta'}^j W(x) \left| P_\theta^{n-j-1} f - \pi_\theta(f) \right|_W \leq C \left(\lambda^j W(x) + \pi_{\theta'}(W) \right) \lambda^{n-j-1}.$$

This concludes the proof. □

Lemma 4.8 *Assume H2-(a),(b). For any compact set $\mathcal{K} \subset \Theta$, there exists $C > 0$ such that for any $\theta, \theta' \in \mathcal{K}$, $\|\pi_\theta - \pi_{\theta'}\|_W \leq C D_W(\theta,\theta')$.*

Proof. For any $x \in \mathsf{X}, n \in \mathbb{N}$,

$$\|\pi_\theta - \pi_{\theta'}\|_W \leq \left\|\pi_\theta - P_\theta^n(x,\cdot)\right\|_W + \left\|P_\theta^n(x,\cdot) - P_{\theta'}^n(x,\cdot)\right\|_W + \left\|P_{\theta'}^n(x,\cdot) - \pi_{\theta'}\right\|_W.$$

Let $\mathcal{K}$ be a compact subset of Θ. By H2-(b), there exist constants $C > 0$ and $\lambda \in (0,1)$ such that for any $n \in \mathbb{N}$ and $x \in \mathsf{X}$ $\sup_{\theta\in\mathcal{K}} \left\|\pi_\theta - P_\theta^n(x,\cdot)\right\|_W \leq C\lambda^n W(x)$. Moreover, using Lemma 4.7, there exists a constant $C' > 0$ such that for any $\theta, \theta' \in \mathcal{K}$ and any $x \in \mathsf{X}$, $\sup_{n\geq 0} \left\|P_\theta^n(x,\cdot) - P_{\theta'}^n(x,\cdot)\right\|_W \leq C' D_W(\theta,\theta') W(x)$. The proof follows, upon noting that x is fixed and arbitrarily chosen. □

Lemma 4.9 *Assume H1, H2-(a),(b) and H4. For any compact set $\mathcal{K} \subset \Theta$, there exists $C > 0$ such that for any $\theta, \theta' \in \mathcal{K}$,*

$$|h(\theta) - h(\theta')| \leq C \left(D_W(\theta, \theta') + |\theta - \theta'|^{\alpha} \right),$$

where D_W and α are given by (2.2) *and H4.*

Proof. Let $\mathcal{K}$ be a compact subset of Θ and θ, θ' in $\mathcal{K}$. By definition of h, it holds

$$|h(\theta) - h(\theta')| = |\pi_\theta H_\theta - \pi_{\theta'} H_{\theta'}| \leq \pi_\theta \left| H_\theta - H_{\theta'} \right| + |(\pi_\theta - \pi_{\theta'}) H_{\theta'}| .$$

Condition H4 implies that there exists a constant $C > 0$ such that for any $\theta, \theta' \in \mathcal{K}$, $\pi_\theta \left| H_\theta - H_{\theta'} \right| \leq C \left| \theta - \theta' \right|^{\alpha}$. By Lemma 4.8 and H1, there exist $C > 0$ such that for any $\theta, \theta' \in \mathcal{K}$, $|(\pi_\theta - \pi_{\theta'}) H_{\theta'}| \leq C D_W(\theta, \theta')$. The proof follows. □

For any $\vartheta \in \Theta$, $x \in \mathsf{X}$ and $L \geq 0$, set

$$\mathcal{H}_{\vartheta, L}(x) \stackrel{\text{def}}{=} \sup_{\{\theta \in \mathcal{K} : \|\theta - \vartheta\| \leq L\}} |H_\theta(x) - H_\vartheta(x)| . \tag{4.8}$$

Proposition 4.10 *Assume H1, H2-(a),(b), and H4. Let g_θ be the solution of (4.6). For any compact set $\mathcal{K} \subset \Theta$, there exist constants $C > 0$ and $\lambda \in (0,1)$ such that for any $\theta, \theta' \in \mathcal{K}$, $x \in \mathsf{X}$, $n \geq 1$, $L > 0$, and any $\vartheta \in \mathcal{K}$ such that $|\theta - \vartheta| \leq L$,*

$$\begin{aligned} |P_\theta g_\theta(x) - P_{\theta'} g_{\theta'}(x)| \leq C \Big\{ & \lambda^n W(x) + 2 \sum_{l=1}^{n} P_\vartheta^l \mathcal{H}_{\vartheta, L}(x) \\ & + n \left| \theta - \theta' \right|^{\alpha} + n \, D_W(\theta, \theta') + n \, D_W(\theta, \vartheta) W(x) + n \, D_W(\theta', \vartheta) W(x) \Big\}. \end{aligned}$$

Proof. For any $\theta, \theta' \in \Theta$ and any $n \geq 1$, we write

$$\begin{aligned} P_\theta g_\theta(x) - P_{\theta'} g_{\theta'}(x) = & \sum_{l > n} \{ P_\theta^l H_\theta(x) - h(\theta) \} - \sum_{l > n} \{ P_{\theta'}^l H_{\theta'}(x) - h(\theta') \} + \psi \left(h(\theta') - h(\theta) \right) \\ & + \sum_{l=1}^{n} \{ P_\theta^l H_\theta(x) - P_{\theta'}^l H_{\theta'}(x) \}. \end{aligned}$$

We first prove that

$$\begin{aligned} & \left| P_\theta g_\theta(x) - P_{\theta'} g_{\theta'}(x) - \sum_{l=1}^{n} \{ P_\theta^l H_\theta(x) - P_{\theta'}^l H_{\theta'}(x) \} \right| \\ & \qquad \leq C \left(\lambda^n W(x) + n \, D_W(\theta, \theta') + n \left| \theta - \theta' \right|^{\alpha} \right). \end{aligned} \tag{4.9}$$

By H1 and H2-(b), for any compact set $\mathcal{K}$, there exist constants $C > 0$ and $\lambda \in (0,1)$ such that for any x, n

$$\sup_{\theta \in \mathcal{K}} \left| \sum_{l > n} \{ P_\theta^l H_\theta(x) - h(\theta) \} \right| \leq C \lambda^{n+1} \sup_{\theta \in \mathcal{K}} |H_\theta|_W \, W(x).$$

(4.9) follows from this inequality and Lemma 4.9. We now establish an upper bound for $\left| \sum_{l=1}^{n} \{ P_\theta^l H_\theta(x) - P_{\theta'}^l H_{\theta'}(x) \} \right|$; we first write

$$\sum_{l=1}^{n} \left| P_\theta^l H_\theta(x) - P_{\theta'}^l H_{\theta'}(x) \right| \leq \sum_{l=1}^{n} \left| P_\theta^l H_\theta(x) - P_\vartheta^l H_\vartheta(x) \right| + \sum_{l=1}^{n} \left| P_\vartheta^l H_\vartheta(x) - P_{\theta'}^l H_{\theta'}(x) \right| .$$

For any $l \geq 1$ and $\vartheta \in \mathcal{K}$ such that $|\theta - \vartheta| \leq L$, we have

$$\begin{aligned} \left|P_\theta^l H_\theta(x) - P_\vartheta^l H_\vartheta(x)\right| &\leq P_\vartheta^l \left|H_\theta - H_\vartheta\right|(x) + \left|\left(P_\theta^l - P_\vartheta^l\right) H_\theta(x)\right| \\ &\leq P_\vartheta^l \mathcal{H}_{\vartheta,L}(x) + \left|\left(P_\theta^l - P_\vartheta^l\right) H_\theta(x)\right| . \end{aligned}$$

By Lemma 4.7, the second term is upper bounded by $C D_W(\theta,\vartheta) W(x)$ for a constant C depending upon $\mathcal{K}$ (and independent of L and l). This concludes the proof. □

4.3. Proof of Theorem 2.1. Define the shifted sequence

$$\boldsymbol{\gamma}^{\leftarrow q} = \{\gamma_{q+n}, n \in \mathbb{N}\}; \tag{4.10}$$

and for any measurable set $\mathcal{K}$ of Θ, define the exit-time from $\mathcal{K}$

$$\sigma(\mathcal{K}) = \inf\{n \geq 1, \theta_n \notin \mathcal{K}\}, \tag{4.11}$$

with the convention that $\inf \emptyset = +\infty$. If $I_N = I_{N-1} + 1 = i$ i.e. the i-th update of the active set occurs at iteration N then

$$X_{N+1} \sim P_{\theta_\star}(x_\star, \cdot), \qquad \theta_{N+1} = \theta_\star + \gamma_1^{\leftarrow i} H_{\theta_\star}(X_{N+1}), \tag{4.12}$$

and for any $\zeta \geq 1$, while $\theta_{N+\zeta} \in \mathcal{K}_{I_N}$,

$$X_{N+\zeta+1} \sim P_{\theta_{N+\zeta}}(X_{N+\zeta}, \cdot) \qquad \theta_{N+\zeta+1} = \theta_{N+\zeta} + \gamma_{\zeta+1}^{\leftarrow i} H_{\theta_{N+\zeta}}(X_{N+\zeta+1}).$$

This iterative scheme can be seen as a perturbation of the algorithm $\tau_{N+\zeta+1} = \tau_{N+\zeta} + \gamma_{\zeta+1}^{\leftarrow i} h(\tau_{N+\zeta})$ and we will show that the sequence $\{\theta_n, n \geq 0\}$ converges as soon as the perturbations $\{H_{\theta_k}(X_{k+1}) - h(\theta_k), k \in \mathbb{N}\}$ are small enough in some sense. We therefore preface the proof of Theorem 2.1 by preliminary results on the control of

$$S_{k,l}(\boldsymbol{\rho}, \mathcal{K}) \stackrel{\text{def}}{=} \mathbb{1}_{\{l \leq \sigma(\mathcal{K})\}} \sum_{j=k}^{l} \rho_j A_{\theta_{j-1}} \left\{H_{\theta_{j-1}}(X_j) - h(\theta_{j-1})\right\},$$

for $l \geq k \geq 1$, a stepsize sequence $\boldsymbol{\rho} = \{\rho_n, n \in \mathbb{N}\}$, a compact subset $\mathcal{K}$ of Θ, and $\sigma(\mathcal{K})$ defined by (4.11). Let $\theta \mapsto A_\theta$, $\theta \in \Theta$, be a measurable $d' \times d$ matrix function, where $d' \geq 1$; we will apply the result to $A_\theta = \mathrm{I}_{d\times d}$ where $\mathrm{I}_{d\times d}$ is the $d \times d$ identity matrix and $A_\theta = \nabla w(\theta)'$.

For a sequence $\boldsymbol{\rho} = \{\rho_n, n \in \mathbb{N}\}$, denote by $\mathbb{P}^{\boldsymbol{\rho}}_{x,\theta}$ (resp. $\mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}$) the probability (resp. the expectation) associated with the non-homogeneous Markov chain on $\mathsf{X} \times \Theta$ with $\delta_{(x,\theta)}$ as initial distribution and with transition mechanism given by line 3 and line 4 of Algorithm 1:

$$X_{n+1} \sim P_{\theta_n}(X_n, \cdot) \qquad \theta_{n+1} = \theta_n + \rho_{n+1} H_{\theta_n}(X_{n+1}).$$

Proposition 4.11 *Assume H1, H2, H3 and H4. Let $\boldsymbol{\rho}$ be a non-increasing positive sequence, $\boldsymbol{\psi} = \{\psi_n, n \in \mathbb{N}\}$ be a sequence such that $1 \leq \psi_n \leq n$ and $\mathcal{K}$ be a compact set of Θ. Let $\theta \mapsto A_\theta$, $\theta \in \Theta$ be a $(d' \times d)$-matrix valued function such that $\sup_{\theta \in \mathcal{K}} |A_\theta| < \infty$ and $\sup_{\theta,\theta' \in \mathcal{K}} |\theta - \theta'|^{-1} \ |A_\theta - A_{\theta'}| \leq C_A$. Then, there exists a constant C such that for any $\delta > 0$, $r \in (0,1]$ and any $(x,\theta) \in \mathsf{X} \times \mathcal{K}$,*

$$\mathbb{P}^{\boldsymbol{\rho}}_{x,\theta}\left(\sup_{l \geq k} |S_{k,l}(\boldsymbol{\rho}, \mathcal{K})| \geq \delta\right) \leq \delta^{-1} C \sum_{j \geq k} \rho_{j-\psi_j}^{1+r\alpha} \psi_j^{1+\alpha} + \delta^{-1} C W^p(x) \sum_{j \geq k} \rho_j^{p(1-r)}$$
$$+ \delta^{-1} W(x) \left\{\rho_k + \sum_{j \geq k} \left(\rho_j \lambda^{\psi_j} + \rho_{j-\psi_j}^{1+rv} \psi_j^3\right)\right\} + C_A \, C \delta^{-1} W^2(x) \sum_{j \geq k} \rho_j^2.$$

Proof. The proof is postponed to Section 4.4. □

Corollary 4.12 *Assume H1 to H6. Let $\mathcal{K}$ be a compact set of Θ and $(x_\star, \theta_\star) \in \mathsf{X} \times \mathcal{K}$. For any $\delta > 0$ and any $\varepsilon \in (0,1)$, there exists $i_\star$ such that for any $i \geq i_\star$*

$$\mathbb{P}^{\boldsymbol{\gamma}^{\leftarrow i}}_{x_\star,\theta_\star}\left(\sup_{n\geq 1}\left|S_{1,n}\left(\boldsymbol{\gamma}^{\leftarrow i}, \mathcal{K}\right)\right| \geq \delta\right) \leq \varepsilon.$$

For any $\delta > 0$ and any $i \geq 0$,

$$\lim_{k\to\infty} \mathbb{P}^{\boldsymbol{\gamma}^{\leftarrow i}}_{x_\star,\theta_\star}\left(\sup_{l\geq k}\left|S_{k,l}\left(\boldsymbol{\gamma}^{\leftarrow i}, \mathcal{K}\right)\right| \geq \delta\right) = 0.$$

Proof. By H1, $W(x_\star) < \infty$. Let α, β be resp. given by H4 and H6. We apply Proposition 4.11 with $r \in \left((\alpha \wedge \upsilon)^{-1}(\beta^{-1} - 1); 1 - (\beta p)^{-1}\right)$, $\boldsymbol{\rho} = \boldsymbol{\gamma}^{\leftarrow i}$ and $\psi_j \sim \tau \ln j$ when $j \to \infty$ for some $\tau \in ((\beta - 1)/\ln \lambda, +\infty)$. □

Proof of Theorem 2.1. Define the sequence of exit-times

$$T_0 = 0 \qquad T_m = \inf\{n \geq T_{m-1} + 1, I_n = I_{n-1} + 1\}, m \geq 1.$$

(i) Let M_0 be such that $\mathcal{L} \cup \mathcal{K}_0 \subset \{\theta \in \Theta : w(\theta) \leq M_0\}$ and set $\mathcal{W} \stackrel{\text{def}}{=} \{\theta \in \Theta : w(\theta) \leq M_0 + 1\}$. Since $\boldsymbol{\gamma}$ is decreasing, Lemma 4.1 shows that there exist $\delta_\star > 0$ and $i_\star \geq 0$ large enough such for any $i \geq i_\star$,

$$\mathbb{P}^{\boldsymbol{\gamma}^{\leftarrow i}}_{x_\star,\theta_\star}\left(\sigma\left(\mathcal{W}\right) < \infty\right) \leq \mathbb{P}^{\boldsymbol{\gamma}^{\leftarrow i}}_{x_\star,\theta_\star}\left(\sup_{n\geq 1}\left|S_{1,n}\left(\boldsymbol{\gamma}^{\leftarrow i}, \mathcal{W}\right)\right| > \delta_\star\right).$$

Corollary 4.12 shows that for any $\varepsilon \in (0,1)$, there exists $j_\star$ such that for any $j \geq j_\star$

$$\mathbb{P}^{\boldsymbol{\gamma}^{\leftarrow j}}_{x_\star,\theta_\star}\left(\sup_{n\geq 1}\left|S_{1,n}\left(\boldsymbol{\gamma}^{\leftarrow j}, \mathcal{W}\right)\right| > \delta_\star\right) \leq \varepsilon.$$

We can assume w.l.o.g. that $j_\star = i_\star$ and we do so. On the other hand, since $\mathcal{W}$ is a compact subset of Θ, by (2.1), there exists $m_\star$ such that for any $m \geq m_\star$, $\mathcal{W} \subset \mathcal{K}_m$. Hereagain, we can assume that $i_\star = m_\star$ and we do so. Hence, for any $i \geq i_\star$

$$\mathbb{P}^{\boldsymbol{\gamma}^{\leftarrow i}}_{x_\star,\theta_\star}\Big(\sigma\left(\mathcal{W}\right) < \infty\Big) \leq \varepsilon. \tag{4.13}$$

This yields for all $i \geq i_\star$,

$$\begin{aligned}\overline{\mathbb{P}}_{x_\star,\theta_\star,0}\left(T_{i+1} < \infty\right) &= \overline{\mathbb{E}}_{x_\star,\theta_\star,0}\left[\mathbb{1}_{\{T_i<\infty\}}\, \mathbb{P}^{\boldsymbol{\gamma}^{\leftarrow i}}_{x_\star,\theta_\star}\left(\sigma\left(\mathcal{K}_i\right) < \infty\right)\right] \\ &\leq \overline{\mathbb{P}}_{x_\star,\theta_\star,0}\left(T_i < \infty\right) \mathbb{P}^{\boldsymbol{\gamma}^{\leftarrow i}}_{x_\star,\theta_\star}\left(\sigma\left(\mathcal{W}\right) < \infty\right) \leq \varepsilon^{i-i_\star},\end{aligned}$$

where we used (4.13) and a trivial induction in the last inequality. Since $\varepsilon \in (0,1)$, we have $\sum_i \overline{\mathbb{P}}_{x_\star,\theta_\star,0}\left(T_{i+1} < \infty\right) < \infty$ which yields $\overline{\mathbb{P}}_{x_\star,\theta_\star,0}(\limsup_i\{T_i < \infty\}) = 0$ by the Borel-Cantelli lemma.

(ii) By Theorem 2.1(i), $I \stackrel{\text{def}}{=} \sup_n I_n$ is finite $\overline{\mathbb{P}}_{x_\star,\theta_\star,0}$-a.s. and $\overline{\mathbb{P}}_{x_\star,\theta_\star,0}(\forall n \geq 1, \theta_n \in \mathcal{K}_I) = 1$. Since I is finite a.s., it is equivalent to prove that for any $i \geq 0$, on the set $\{I = i\}$, $\lim_k \mathrm{d}(\theta_k, \mathcal{L}) = 0$ a.s. Let i be fixed. We apply Proposition 4.6 with $\boldsymbol{\rho} = \boldsymbol{\gamma}^{\leftarrow i}$, $\vartheta_k = \theta_{T_i+k}$ and $\xi_j \leftarrow H_{\theta_j}(X_{j+1}) - h\left(\theta_j\right)$.
h is Holder-continuous under H2, H3 and H4. Since $\sum_k \gamma_k = +\infty$ and $\lim_k \gamma_k = 0$

by assumptions, then $\sum_k \rho_k = \infty$ and $\lim_k \rho_k = 0$. For any $\delta > 0$, by applying the strong Markov property with the stopping-time T_i, we have

$$\begin{aligned}
&\overline{\mathbb{P}}_{x_\star,\theta_\star,0}\left(\limsup_k \sup_{l\geq k}\left|\sum_{j=k}^{l}\gamma_{i+j+1}A_{\theta_{T_i+j}}\left\{H_{\theta_{T_i+j}}(X_{T_i+j+1})-h\left(\theta_{T_i+j}\right)\right\}\right|\geq\delta, I=i\right)\\
&\leq \mathbb{P}^{\boldsymbol{\gamma}^{\leftarrow i}}_{x_\star,\theta_\star}\left(\limsup_k \sup_{l\geq k}\left|\sum_{j=k}^{l}\gamma_{i+j+1}A_{\theta_{T_i+j}}\left\{H_{\theta_j}(X_{j+1})-h\left(\theta_j\right)\right\}\right|\geq\delta, \sigma(\mathcal{K}_i)=+\infty\right)\\
&\leq \mathbb{P}^{\boldsymbol{\gamma}^{\leftarrow i}}_{x_\star,\theta_\star}\left(\limsup_k \sup_{l\geq k}\left|S_{k,l}\left(\boldsymbol{\gamma}^{\leftarrow i},\mathcal{K}_i\right)\right|\geq\delta\right)\\
&\leq \lim_k \mathbb{P}^{\boldsymbol{\gamma}^{\leftarrow i}}_{x_\star,\theta_\star}\left(\sup_{l\geq k}\left|S_{k,l}\left(\boldsymbol{\gamma}^{\leftarrow i},\mathcal{K}_i\right)\right|\geq\delta\right).
\end{aligned}$$

The RHS is zero by Corollary 4.12. Hence, by choosing $A_{\vartheta_j} = \mathrm{I}_{d\times d}$, the Proposition 4.6-(C-iv) holds; note also that with this choice of A_θ, $C_A = 0$. Let us check Proposition 4.6-(C-v). Under Theorem 2.1-(i), Proposition 4.6- (C-v)-(A) holds.
Assume now that Theorem 2.1-(ii) is satisfied; we prove that Proposition 4.6-(C-v)-(B) holds. Along the same lines as above, and choosing A_{θ_j} equal to the transpose of $\nabla w(\theta_j)$, we establish that $\lim_k \sum_{j=1}^k \rho_j \langle\nabla w(\vartheta_j), \xi_j\rangle$ exists. By H1 and since $\sup_{\mathcal{K}} |h| < \infty$, there exists a constant C such that

$$\begin{aligned}
&\overline{\mathbb{P}}_{x_\star,\theta_\star,0}\left(\limsup_k \sup_{l\geq k}\sum_{j=k}^{l}\gamma^2_{i+j+1}\left|H_{\theta_{T_i+j}}(X_{T_i+j+1})-h\left(\theta_{T_i}\right)\right|^2\geq\delta, I=i\right)\\
&\leq \lim_k \mathbb{P}^{\boldsymbol{\gamma}^{\leftarrow i}}_{x_\star,\theta_\star}\left(\sup_{l\geq l}\sum_{j=k}^{k}\gamma^{\leftarrow i}_{j+1}(W^2(X_{j+1})+1)\geq\delta/C\right).
\end{aligned}$$

The RHS tends to zero since $\sup_j \mathbb{E}^{\boldsymbol{\gamma}^{\leftarrow i}}_{x_\star,\theta_\star}\left[W^2(X_j)\right] < \infty$ (we assumed $p \geq 2$ in H2-(c)) and $\sum_j \gamma_j^2 < \infty$ (we assumed that $\beta > 1/2$). Hence, Proposition 4.6-(C-v)-(B) holds. We then conclude by Proposition 4.6 that $\overline{\mathbb{P}}_{x_\star,\theta_\star,0}$-a.s. on the set $I = i$, the sequence $\{\theta_k, k \geq 0\}$ converges to a connected component of $\mathcal{L}$.

4.4. Proof of Proposition 4.11. Let $\boldsymbol{\rho}$ be a non-increasing positive sequence and $\mathcal{K}$ be a compact subset of Θ such that $\mathcal{K}_0 \subseteq \mathcal{K}$. Throughout this section, $\boldsymbol{\rho}$ and $\mathcal{K}$ are fixed; we will therefore use the notations $S_{k,l}$ and σ instead of $S_{k,l}(\boldsymbol{\rho},\mathcal{K})$ and $\sigma(\mathcal{K})$. Set

$$C_\star \stackrel{\text{def}}{=} \sup_{\theta\in\mathcal{K}}\{|g_\theta|_W + |P_\theta g_\theta|_W\} \vee \sup_{\theta\in\mathcal{K}} |A_\theta|\,, \tag{4.14}$$

where g_θ is the solution to the Poisson equation (4.6). $C_\star$ is finite by (4.7), H1 and H2-(b). By (4.6), $H_{\theta_{j-1}}(X_j) - h(\theta_{j-1}) = g_{\theta_{j-1}}(X_j) - P_{\theta_{j-1}} g_{\theta_{j-1}}(X_j)$ for any $j \geq 1$.

We then write $S_{k,l} = \mathbb{1}_{\{\sigma \geq l\}} \sum_{i=1}^{4} T_{k,l}^{(i)}$ with

$$T_{k,l}^{(1)} \stackrel{\text{def}}{=} \sum_{j=k}^{l} \rho_j A_{\theta_{j-1}} \left(g_{\theta_{j-1}}(X_j) - P_{\theta_{j-1}} g_{\theta_{j-1}}(X_{j-1}) \right) \mathbb{1}_{\{j \leq \sigma\}},$$

$$T_{k,l}^{(2)} \stackrel{\text{def}}{=} \sum_{j=k}^{l} \rho_{j+1} A_{\theta_j} \left(P_{\theta_j} g_{\theta_j}(X_j) - P_{\theta_{j-1}} g_{\theta_{j-1}}(X_j) \right) \mathbb{1}_{\{j+1 \leq \sigma\}},$$

$$T_{k,l}^{(3)} \stackrel{\text{def}}{=} \sum_{j=k}^{l} \left(\rho_{j+1} A_{\theta_j} \mathbb{1}_{\{j+1 \leq \sigma\}} - \rho_j A_{\theta_{j-1}} \mathbb{1}_{\{j \leq \sigma\}} \right) P_{\theta_{j-1}} g_{\theta_{j-1}}(X_j),$$

$$T_{k,l}^{(4)} \stackrel{\text{def}}{=} \rho_k A_{\theta_{k-1}} P_{\theta_{k-1}} g_{\theta_{k-1}}(X_{k-1}) - \rho_{l+1} A_{\theta_l} P_{\theta_l} g_{\theta_l}(X_l) \mathbb{1}_{\{l+1 \leq \sigma\}}.$$

For any measurable set $\mathcal{A}$, we can bound by Markov's and Jensen's inequality

$$\mathbb{P}_{x,\theta}^{\boldsymbol{\rho}} \left(\sup_{l \geq k} |S_{k,l}| \geq \delta \right) \leq \frac{3}{\delta} \Bigg(\mathbb{E}_{x,\theta}^{\boldsymbol{\rho}} \left[\sup_{l \geq k} \left| T_{k,l}^{(1)} \right|^p \right]^{1/p} + \mathbb{E}_{x,\theta}^{\boldsymbol{\rho}} \left[\sup_{k \leq l \leq \sigma} \left| T_{k,l}^{(3)} + T_{k,l}^{(4)} \right| \right] \\ + \mathbb{E}_{x,\theta}^{\boldsymbol{\rho}} \left[\sup_{k \leq l \leq \sigma} \left| T_{k,l}^{(2)} \mathbb{1}_{\{\mathcal{A}\}} \right| \right] \Bigg) + \mathbb{P}_{x,\theta}^{\boldsymbol{\rho}}(\mathcal{A}^c).$$

The terms on the right are bounded individually by the three lemmas below, concluding the proof of Proposition 4.11.

Lemma 4.13 *Assume H1, H2 and* $\sup_{\theta \in \mathcal{K}} |A_\theta| < \infty$. *There exists a constant C - which does not depend on $\boldsymbol{\rho}$ - such that for any $x \in \mathsf{X}$, $\theta \in \mathcal{K}_0$ and $k \geq 1$,*

$$\mathbb{E}_{x,\theta}^{\boldsymbol{\rho}} \left[\sup_{l \geq k} \left| T_{k,l}^{(1)} \right|^p \right] \leq C\, W^p(x) \sum_{\ell \geq k} \rho_\ell^p.$$

Proof. Let $k \geq 1$ be fixed. Note that $\{T_{k,l}^{(1)}, l \geq k\}$ is a $\mathcal{F}_l$-martingale under the probability $\mathbb{P}_{x,\theta}^{\boldsymbol{\rho}}$ which implies that for any $p > 1$, there exists a constant C such that (see e.g. [17, Theorems 2.2 and 2.10])

$$\begin{aligned} \mathbb{E}_{x,\theta}^{\boldsymbol{\rho}} \left[\sup_{l \geq k} \left| T_{k,l}^{(1)} \right|^p \right] &\leq \lim_{L \to \infty} \mathbb{E}_{x,\theta}^{\boldsymbol{\rho}} \left[\mathbb{E}_{x,\theta}^{\boldsymbol{\rho}} \left[\sup_{k \leq l \leq k+L} \left| T_{k,l}^{(1)} \right|^p \Big| \mathcal{F}_k \right] \right] \\ &\leq C \lim_{L \to \infty} \mathbb{E}_{x,\theta}^{\boldsymbol{\rho}} \left[\mathbb{E}_{x,\theta}^{\boldsymbol{\rho}} \left[\left| T_{k,k+L}^{(1)} \right|^p \Big| \mathcal{F}_k \right] \right] \\ &\leq C\, W^p(x) \left(\sum_{l \geq k} \rho_l^p \right). \end{aligned}$$

□

Lemma 4.14 *Assume H1, H2,* $\sup_{\theta \in \mathcal{K}} |A_\theta| < \infty$ *and* $\sup_{\theta, \theta' \in \mathcal{K}} |\theta - \theta'|^{-1} |A_\theta - A_{\theta'}| \leq C_A$. *There exists a constant C - which does not depend on $\boldsymbol{\rho}$ - such that for any $x \in \mathsf{X}$, $\theta \in \mathcal{K}_0$, $k \geq 1$,*

$$\mathbb{E}_{x,\theta}^{\boldsymbol{\rho}} \left[\sup_{k \leq l \leq \sigma} \left| T_{k,l}^{(3)} + T_{k,l}^{(4)} \right| \right] \leq C \left(\rho_k W(x) + C_A \sum_{j \geq k} \rho_j^2 W^2(x) \right).$$

Proof. By (4.14), we have $|P_{\theta_i} g_{\theta_i}|_W \mathbb{1}_{\{i<\sigma\}} \leq C_\star$ and $\sup_{\theta \in \mathcal{K}} |A_\theta| \leq C_\star$. Since $\boldsymbol{\rho}$ is non-increasing, this yields $\mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[\sup_{k\leq l\leq\sigma}\left|T^{(4)}_{k,l}\right|\right] \leq 2C_\star^2 \rho_k W(x)$. We write $T^{(3)}_{k,l} = T^{(3,a)}_{k,l} - T^{(3,b)}_{k,l}$ with

$$T^{(3,a)}_{k,l} \stackrel{\text{def}}{=} \sum_{j=k}^{l} \left(\rho_{j+1} A_{\theta_j} - \rho_j A_{\theta_{j-1}}\right) P_{\theta_{j-1}} g_{\theta_{j-1}}(X_j) \mathbb{1}_{\{j+1\leq\sigma\}},$$

$$T^{(3,b)}_{k,l} \stackrel{\text{def}}{=} \sum_{j=k}^{l} \rho_j A_{\theta_{j-1}} P_{\theta_{j-1}} g_{\theta_{j-1}}(X_j) \mathbb{1}_{\{\sigma=j\}}.$$

Note that $\{j=\sigma\}\cap\{l\leq\sigma\}=\emptyset$ for any $j<l$. Hence

$$\mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[\sup_{k\leq l\leq\sigma}\left|T^{(3,b)}_{k,l}\right|\right] = \mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[\rho_l \left|A_{\theta_{l-1}}\right| \left|P_{\theta_{l-1}} g_{\theta_{l-1}}(X_l)\right| \mathbb{1}_{\{l=\sigma\}}\right] \leq C_\star^2 \rho_k W(x),$$

where in the inequality we used that $\boldsymbol{\rho}$ is non-increasing. Finally, along the same lines, we get

$$\begin{aligned}\left|\rho_{j+1} A_{\theta_j} - \rho_j A_{\theta_{j-1}}\right| \left|P_{\theta_{j-1}} g_{\theta_{j-1}}(X_j)\right| \mathbb{1}_{\{j+1\leq\sigma\}} &\leq (\rho_j - \rho_{j+1}) C_\star^2 W(X_j) \mathbb{1}_{\{j<\sigma\}} \\ &\quad + \rho_j \left|A_{\theta_j} - A_{\theta_{j-1}}\right| C_\star W(X_j) \mathbb{1}_{j<\sigma}\end{aligned}$$

Since $\left|A_{\theta_j} - A_{\theta_{j-1}}\right| \mathbb{1}_{j<\sigma} \leq C_A C_\star \rho_j W(X_j)$, this yields $\mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[\sup_{k\leq l\leq\sigma}\left|T^{(3,a)}_{k,l}\right|\right] \leq C_\star^2\left(\rho_k W(x) + C_A \rho_k^2 W^2(x)\right)$. □

For any $r\in(0,1)$ and $0\leq k-\psi_k\leq n$, set

$$\mathcal{A}_r(k,n) \stackrel{\text{def}}{=} \bigcap_{\ell=k-\psi_k}^{n} \{|\theta_\ell - \theta_{\ell-1}| \leq \rho_\ell^r\} = \left\{\sup_{k-\psi_k\leq\ell\leq n} \frac{|\theta_\ell - \theta_{\ell-1}|}{\rho_\ell^r} \leq 1\right\}.$$

Proposition 4.15 *Assume H1, H2, H3, H4 and* $\sup_{\theta\in\mathcal{K}}|A_\theta|<\infty$.

(i) There exists a constant $C<\infty$ - which does not depend on $\boldsymbol{\rho}$ - such that for any $r\in(0,1)$, $x\in\mathsf{X}$, $\theta\in\Theta$ and $k\geq 1$,

$$\mathbb{P}^{\boldsymbol{\rho}}_{x,\theta}\Big(\mathcal{A}_r(k,\sigma)\Big) \geq 1 - C\left(\sum_{\ell\geq k-\psi_k} \rho_\ell^{p(1-r)}\right) W^p(x). \tag{4.15}$$

(ii) There exists a constant $C>0$ - which does not depend on $\boldsymbol{\rho}$ - such that for any $r\in(0,1)$, $x\in\mathsf{X}$, $\theta\in\Theta$, $k\geq 1$, and for any positive sequence $\boldsymbol{\psi}=\{\psi_j, j\in\mathbb{N}\}$ such that $1\leq\psi_j\leq j$,

$$\begin{aligned} C\,\mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}&\left[\sup_{k\leq l\leq\sigma}\left|T^{(2)}_{k,l}\right| \mathbb{1}_{\{\mathcal{A}_r(k,\sigma)\}}\right] \\ &\leq \sum_{j\geq k} \rho_{j-\psi_j}^{1+r\alpha} \psi_j^{1+\alpha} + W(x) \sum_{j\geq k}\left(\rho_j \lambda^{\psi_j} + \rho_{j-\psi_j}^{1+rv} \psi_j^3\right). \end{aligned}$$

Proof. Throughout this proof, C denotes a constant which may change upon each appearance and only depends on $\mathcal{K}$.

(i) By H1, there exists a constant C such that $\mathbb{P}^{\boldsymbol{\rho}}_{x,\theta}$-a.s. , on the set $\{k-\psi_k \leq \ell \leq \sigma\}$, $|\theta_\ell - \theta_{\ell-1}| \leq C\rho_\ell W(X_\ell)$. Hence, by the Markov inequality

$$1 - \mathbb{P}^{\boldsymbol{\rho}}_{x,\theta}\left(\mathcal{A}_r(k,\sigma)\right) \leq C\mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[\sup_{k-\psi_k \leq \ell \leq \sigma} \rho_\ell^{p(1-r)} W^p(X_\ell)\right]$$
$$\leq C\left(\sum_{\ell \geq k-\psi_k} \rho_\ell^{p(1-r)}\right) W^p(x).$$

(ii) We use Proposition 4.10 with $\theta \leftarrow \theta_j$, $\theta' \leftarrow \theta_{j-1}$, $\vartheta \leftarrow \theta_{j-\psi_j}$, $x \leftarrow X_j$, $n \leftarrow \psi_j$ and $L \leftarrow L_j \stackrel{\text{def}}{=} \psi_j \rho^r_{j-\psi_j+1}$. Since $\boldsymbol{\rho}$ is non-increasing, observe that

$$\left(\left|\theta_j - \theta_{j-\psi_j}\right| \vee \left|\theta_{j-1} - \theta_{j-\psi_j}\right|\right) \mathbb{1}_{\{\mathcal{A}_r(k,j)\}} \leq \psi_j\, \rho^r_{j-\psi_j+1},$$

thus justifying that with the above definitions, we have $|\theta - \vartheta| \vee |\theta' - \vartheta| \leq L$. By using $D_W(\theta,\theta'') \leq D_W(\theta,\theta') + D_W(\theta',\theta'')$ and $W \geq 1$, we write $\sup_{k \leq l \leq \sigma}\left|T^{(2)}_{k,l}\right| \mathbb{1}_{\{\mathcal{A}_r(k,\sigma)\}} \leq C\sum_{i=1}^2 \Xi^{(i)}_k$ with

$$\Xi^{(1)}_k \stackrel{\text{def}}{=} \sum_{j\geq k} \rho_j \psi_j A_{\theta_{j-1}} \left\{|\theta_j - \theta_{j-1}|^\alpha + 2D_W(\theta_j,\theta_{j-1})W(X_j)\right.$$
$$\left.+2D_W(\theta_{j-1},\theta_{j-\psi_j})W(X_j)\right\} \mathbb{1}_{\{j\leq\sigma\}} \mathbb{1}_{\{\mathcal{A}_r(k,j)\}} + \sum_{j\geq k} \rho_j A_{\theta_{j-1}} \lambda^{\psi_j} W(X_j) \mathbb{1}_{\{j\leq\sigma\}},$$

$$\Xi^{(2)}_k \stackrel{\text{def}}{=} \sum_{j\geq k} \rho_j A_{\theta_{j-1}} \sum_{l=1}^{\psi_j} P^l_{\theta_{j-\psi_j}} \mathcal{H}_{\theta_{j-\psi_j},L_j}(X_j) \mathbb{1}_{\{j\leq\sigma\}} \mathbb{1}_{\{\mathcal{A}_r(k,j-1)\}}.$$

Let us consider $\Xi^{(1)}_k$. By (4.14), H3 and the monotonicity of $\boldsymbol{\rho}$, we have

$$\mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[\Xi^{(1)}_k\right] \leq C_\star \sum_{j\geq k} \rho_j^{1+r\alpha}\psi_j + C_\star W(x)\sum_{j\geq k}\rho_j\lambda^{\psi_j} + CW(x)\sum_{j\geq k}\psi_j^2 \rho^{1+r\upsilon}_{j-\psi_j}.$$

Let us now consider $\Xi^{(2)}_k$. Set $\mathbb{B}_{l,j} \stackrel{\text{def}}{=} \mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[D_W(\theta_l,\theta_{j-\psi_j})W(X_l)\mathbb{1}_{\{l+1\leq\sigma\}}\mathbb{1}_{\{\mathcal{A}_r(k,l)\}}\right]$. We write

$$\mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[P^l_{\theta_{j-\psi_j}}\mathcal{H}_{\theta_{j-\psi_j},L_j}(X_j)\mathbb{1}_{\{j\leq\sigma\}}\mathbb{1}_{\{\mathcal{A}_r(k,j-1)\}}\right]$$
$$= \mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[P_{\theta_{j-1}}P^l_{\theta_{j-\psi_j}}\mathcal{H}_{\theta_{j-\psi_j},L_j}(X_{j-1})\mathbb{1}_{\{j\leq\sigma\}}\mathbb{1}_{\{\mathcal{A}_r(k,j-1)\}}\right]$$
$$\leq \mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[P^{l+1}_{\theta_{j-\psi_j}}\mathcal{H}_{\theta_{j-\psi_j},L_j}(X_{j-1})\mathbb{1}_{\{j\leq\sigma\}}\mathbb{1}_{\{\mathcal{A}_r(k,j-1)\}}\right] + C\,\mathbb{B}_{j-1,j},$$

where in the last inequality, we used that

$$\left|P^l_{\theta_{j-\psi_j}}\mathcal{H}_{\theta_{j-\psi_j},L_j}\right|_W \mathbb{1}_{\{j-\psi_j<\sigma\}} \leq 2\sup_{\theta\in\mathcal{K}}|H_\theta|_W \;\sup_{l\geq1}\sup_{\theta\in\mathcal{K}}\left|P^l_\theta W\right|_W$$

which is finite by H1 and H2-(b). Since $\mathbb{1}_{\{j\leq\sigma\}}\mathbb{1}_{\{\mathcal{A}_r(k,j-1)\}} \leq \mathbb{1}_{\{j-1\leq\sigma\}}\mathbb{1}_{\{\mathcal{A}_r(k,j-2)\}}$, we have by a trivial induction

$$\mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[P^l_{\theta_{j-\psi_j}}\mathcal{H}_{\theta_{j-\psi_j},L_j}(X_j)\mathbb{1}_{\{j\leq\sigma\}}\mathbb{1}_{\{\mathcal{A}_r(k,j-1)\}}\right] \leq \mathbb{A}_{l,j} + C\sum_{i=1}^{\psi_j-1}\mathbb{B}_{j-i,j},$$

where $\mathbb{A}_{l,j} \stackrel{\text{def}}{=} \mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[P^{l+\psi_j}_{\theta_{j-\psi_j}} \mathcal{H}_{\theta_{j-\psi_j},L_j}(X_{j-\psi_j}) \mathbb{1}_{\{j-\psi_j+1\leq\sigma\}}\mathbb{1}_{\{\mathcal{A}_r(k,j-\psi_j)\}}\right]$. Finally, using again $|\mathcal{H}_{\vartheta,L}|_W \, \mathbb{1}_{\{\vartheta\in\mathcal{K}\}} \leq 2\sup_{\theta\in\mathcal{K}}|H_\theta|_W$ and H2-(b), we obtain

$$\mathbb{A}_{l,j} \leq C\,\lambda^{l+\psi_j} W(x) + \sup_{\theta\in\mathcal{K}} \pi_\theta \mathcal{H}_{\theta,L_j}.$$

By H4, the last term in the RHS is upper bounded by L_j^α. Combining the above inequalities and using (4.14), we have

$$\mathbb{E}^{\boldsymbol{\rho}}_{x,\theta}\left[\Xi^{(2)}_k\right] \leq C\sum_{j\geq k}\rho_j\psi_j\sum_{i=1}^{\psi_j-1}\mathbb{B}_{j-i,j} + C_\star\sum_{j\geq k}\rho_j\psi_j L_j^\alpha + C\mathbb{W}_{x,\theta}\sum_{j\geq k}\rho_j\lambda^{\psi_j}.$$

The result follows upon noting that $\mathbb{B}_{l,j} \leq W(x)(l-j+\psi_j)\rho^{rv}_{j-\psi_j}$.
□

5. Proofs of Section 3.

5.1. Proof of Proposition 3.1. Let $\Theta = \mathbb{R}$. From (3.1), $\sup_{\theta\in\Theta}\sup_{x\in\mathbb{R}}|H_\theta(x)| \leq 1$ so that H1 is satisfied with the constant function $W = 1$.

We have $|H_{\theta_1}(x) - H_{\theta_2}(x)| = \mathbb{1}_{\{\theta_1\wedge\theta_2\leq\phi(x)<\theta_1\vee\theta_2\}}$ and under the stated assumptions,

$$\sup_{\theta\in\mathbb{R}}\int \sup_{\theta'\in\mathcal{B}(\theta,\delta)}|H_\theta(x) - H_{\theta'}(x)|\,\pi(\mathrm{d}x) \leq \sup_{\theta\in\mathbb{R}}\int \mathbb{1}_{\{\theta-\delta\leq\phi(y)\leq\theta+\delta\}}\pi(y)\mathrm{d}y \leq C\delta.$$

Therefore, H4 is satisfied with $\alpha = 1$.

Since the weak derivative of $\theta\mapsto|\theta - x|$ is $\mathrm{sign}(\theta - x)$ almost-everywhere, the dominated convergence theorem implies that w is differentiable and its derivative is

$$w'(\theta) = \frac{1}{2}\left(\int\mathbb{1}_{\phi(y)\leq\theta}\pi(y)\mathrm{d}y - \int\mathbb{1}_{\phi(y)\geq\theta}\pi(y)\mathrm{d}y\right) + \left(\frac{1}{2} - q\right) = \int\mathbb{1}_{\phi(y)\leq\theta}\pi(y)\mathrm{d}y - q; \tag{5.1}$$

we also have w continuously differentiable. Since $\int|\phi(y)|\,\pi(y)\mathrm{d}y < \infty$,

$$w(\theta) \geq \frac{|\theta|}{2} + \left(\frac{1}{2} - q\right)\theta - \frac{1}{2}\int|\phi(y)|\,\pi(y)\mathrm{d}y \underset{|\theta|\to\infty}{\longrightarrow} \infty;$$

since w is continuous, this implies that the level sets of w are compact, thus showing H5-(a) holds. By definition of h (see (2.3)), we have $h(\theta) = -w'(\theta)$. Therefore, the set $\mathcal{L}$ in H5-(b) is given by (3.3) and it is compact. In addition, H5-(c) is satisfied. Finally, $w(\theta)$ reaches its minimum at $\theta_\star\in\mathcal{L}$ (see (5.1)). Since the Lyapunov function w is defined up to an additive constant, we can assume with no loss of generality that w is non-negative, which concludes the proof of H5.
Note that w is constant on $\mathcal{L}$ since $w'(\theta) = 0$ for any $\theta\in\mathcal{L}$ and $\mathcal{L}$ is an interval. Hence $w(\mathcal{L})$ has an empty interior.

5.2. Proof of Proposition 3.2. Let $\mathcal{K}_1$ and $\mathcal{K}_2$ be resp. a compact of $\mathbb{R}$ and $\mathcal{V}$; set $\mathcal{K} \stackrel{\text{def}}{=} \mathcal{K}_1\times\mathcal{K}_2$. Let $\tau > 0$ be such that $C\tau^\alpha \leq \delta$ where C,δ are given by E1. For any $\vartheta = (\theta, s)\in\mathcal{K}$, $\vartheta' = (\theta', s')$ with $|\theta - \theta'| \leq \tau$ and $|s - s'| \leq \tau$, and $x = (y,z)\in\mathsf{X}\times\mathsf{X}$ it holds

$$\begin{aligned}|H_\vartheta(x) - H_{\vartheta'}(x)| &\leq |s - s'| + \frac{\mu(z)}{g_{\hat\nu(s)}(z)}\left\{\left|\mathbb{1}_{\{\phi(z)\geq\theta\}} - \mathbb{1}_{\{\phi(z)\geq\theta'\}}\right| + \left|1 - \frac{g_{\hat\nu(s)}(z)}{g_{\hat\nu(s')}(z)}\right|\right\}\\ &\leq \tau + \frac{\mu(z)}{g_{\hat\nu(s)}(z)}\left\{\mathbb{1}_{\{\theta\wedge\theta'\leq\phi(z)\leq\theta\vee\theta'\}} + \psi(s,s',z)\exp\left(\psi(s,s',z)\right)\right\},\end{aligned}$$

where $\psi(s, s', z) \stackrel{\text{def}}{=} |B(\hat{\nu}(s)) - B(\hat{\nu}(s'))| + |\hat{\nu}(s) - \hat{\nu}(s')| \, |S(z)|$. By E1, there exists a constant C - depending only upon $\mathcal{K}$ - such that for any $\vartheta \in \mathcal{K}$

$$\int \pi_\vartheta(\mathrm{d}x) \sup_{|\theta - \theta'| \leq \tau} |H_\vartheta(x) - H_{\vartheta'}(x)| \leq \tau + C\tau^\alpha .$$

5.3. Proofs of Section 3.3.

Lemma 5.1 *Under E2, for any* $0 \leq \kappa < d$, $\sup_{\theta \in \mathbb{R}^d} \int |x - \theta|^{-\kappa} \pi(x)\mathrm{d}x < \infty$.

Proof. Let $0 < \kappa < d$.

$$\begin{aligned} \int |x - \theta|^{-\kappa} \pi(x)\mathrm{d}x &= \int_0^{+\infty} \mathrm{d}t \int_{\{x : |x-\theta|^\kappa \leq 1/t\}} \pi(x)\mathrm{d}x \\ &\leq 1 + \sup_{x \in \mathbb{R}^d} \pi(x) \int_1^{+\infty} \mathrm{d}t \sup_{\theta \in \mathbb{R}^d} \int_{\{x : |x-\theta| \leq t^{-\kappa}\}} \mathrm{d}x \\ &\leq 1 + C \int_1^{+\infty} t^{-d/\kappa} \mathrm{d}t, \end{aligned}$$

for a finite constant C, which does not depend on θ. □

Proof. [Proposition 3.3] As $|H_\theta(x)| = 1$, H1 is satisfied with the constant function $W = 1$. By [6, Lemma 19-(ii)], there exists C such that for any x, θ, t with $x \neq \theta$, we have

$$\left| |x - \theta + t| - |x - \theta| + \left\langle t, \frac{x - \theta}{|x - \theta|} \right\rangle \right| \leq C \frac{|t|^2}{|x - \theta|}$$

Since $\sup_{\theta \in \Theta} \int |x - \theta|^{-1} \pi(x)\mathrm{d}x < \infty$ (see Lemma 5.1 below), then this inequality implies that w is differentiable and $\nabla w(\theta) = -\int (x - \theta)/|x - \theta| \, \pi(x)\mathrm{d}x$. The dominated convergence theorem implies that w is continuously differentiable. H5-(a) follows from the lower bound $w(\theta) \geq |\theta| - \int |x| \, \pi(x)\mathrm{d}x$ and the continuity of w. We have $\nabla w = -h$ from which H5-(c) trivially follows. Finally, by E2 and [30], $\mathcal{L}$ contains a single point, and H5-(b) is satisfied.

Let $\theta, \theta' \in \Theta$. For any $x \notin \{\theta, \theta'\}$,

$$|H_{\theta'}(x) - H_\theta(x)| = \left| \frac{x - \theta'}{|x - \theta'| \, |x - \theta|} \left(|x - \theta| - |x - \theta'|\right) + \frac{\theta - \theta'}{|x - \theta|} \right| \leq 2 \frac{|\theta' - \theta|}{|x - \theta|}.$$

Define $\mathcal{H}_{\theta,\delta}(x) \stackrel{\text{def}}{=} \sup_{\theta' \in \mathcal{B}(\theta,\delta)} |H_{\theta'}(x) - H_\theta(x)|$. Let $0 < \beta < 1/d$. Then

$$\begin{aligned} &\int \pi(x)\mathcal{H}_{\theta,\delta}(x)\mathrm{d}x = \int_{x \in \mathcal{B}(\theta, \delta + \delta^\beta)} \pi(x)\mathcal{H}_{\theta,\delta}(x)\mathrm{d}x + \int_{x \notin \mathcal{B}(\theta, \delta + \delta^\beta)} \pi(x)\mathcal{H}_{\theta,\delta}(x)\mathrm{d}x \\ &\leq 2 \sup_{x \ in \mathbb{R}^d} \pi(x) \int_{x \in \mathcal{B}(\theta, \delta + \delta^\beta)} \mathrm{d}x + 2 \int_{x \notin \mathcal{B}(\theta, \delta + \delta^\beta)} \sup_{\theta' \in \mathcal{B}(\theta,\delta)} \frac{|\theta' - \theta|}{|x - \theta|} \pi(x)\mathrm{d}x \\ &\leq C\delta^{\beta d} + 4\delta^{1-\beta}, \end{aligned}$$

for a constant C which is finite by E2. Hence, and H4 is satisfied with $\alpha = (\beta d) \wedge (1 - \beta) < 1$. This holds true with $\beta = 1/(1 + d)$ for which $\beta d = 1 - \beta$. □

5.4. Proofs of Section 3.4. We start with a preliminary lemma which gives a control on the intersection of two Voronoi cells associated with $\theta, \bar{\theta} \in (\mathbb{R}^d)^N$.

Lemma 5.2 *For any compact set $\mathcal{K}$ of Θ, there exists $\delta_{\mathcal{K}} > 0$ such that for any $\theta \in \mathcal{K}$ and any $i \neq j$:*

(i)

$$\sup_{\delta \leq \delta_K} \frac{1}{\sqrt{\delta}} \sup_{\bar\theta \in \mathcal{B}(\theta,\delta)^N \cap \Theta} \left| \frac{\bar\theta^{(j)} - \bar\theta^{(i)}}{\left|\bar\theta^{(j)} - \bar\theta^{(i)}\right|} - \frac{\theta^{(j)} - \theta^{(i)}}{\left|\theta^{(j)} - \theta^{(i)}\right|} \right| < \infty.$$

(ii) for any $\delta \leq \delta_{\mathcal{K}}$, there exists a measurable set $R_{i,j}(\theta,\delta)$ such that

$$\sup_{\bar\theta \in \mathcal{B}(\theta,\delta) \cap \Theta} \mathbb{1}_{\{C_i(\theta) \cap C^{(j)}(\bar\theta) \cap \mathcal{B}(0,\Delta)\}} \leq \mathbb{1}_{\{R_{i,j}(\theta,\delta)\}},$$
$$\sup_{\delta \leq \delta_K} \frac{1}{\sqrt{\delta}} \int \mathbb{1}_{\{R_{i,j}(\theta,\delta)\}}(x)\,\mathrm{d}x < \infty.$$

Proof. Let $\mathcal{K}$ be a compact set of Θ. The function on $(\mathbb{R}^d)^N$ given by $\theta \mapsto \min_{i\neq j} \left|\theta^{(i)} - \theta^{(j)}\right|$ is continuous. Since $\mathcal{K}$ is a compact subset of Θ, there exists $b_{\mathcal{K}} > 0$ such that for any $\theta \in \mathcal{K}$, $\min_{i \neq j} \left|\theta^{(i)} - \theta^{(j)}\right| \geq b_{\mathcal{K}}$. Choose $\delta_{\mathcal{K}} \in (0, b_{\mathcal{K}}/2 \wedge 1)$. Let $i \neq j \in \{1, \cdots, N\}$ and $\theta \in \mathcal{K}$ be fixed. For any $\delta \leq \delta_{\mathcal{K}}$ and $\bar\theta \in \mathcal{B}(\theta,\delta)$, it holds

$$\begin{aligned}\left|\bar\theta^{(j)} - \bar\theta^{(i)}\right| &\geq \left|\theta^{(j)} - \theta^{(i)}\right| - \left|\bar\theta^{(j)} - \theta^{(j)}\right| - \left|\bar\theta^{(i)} - \theta^{(i)}\right| \geq \left|\theta^{(j)} - \theta^{(i)}\right| - 2\delta \\ &\geq b_{\mathcal{K}} - 2\delta > 0.\end{aligned} \tag{5.2}$$

Similarly,

$$\left|\bar\theta^{(j)} - \bar\theta^{(i)}\right| \leq \left|\theta^{(j)} - \theta^{(i)}\right| + 2\delta. \tag{5.3}$$

Define $n = (\theta^{(j)} - \theta^{(i)})/\left|\theta^{(j)} - \theta^{(i)}\right|$ and $n' = (\bar\theta^{(j)} - \bar\theta^{(i)})/\left|\bar\theta^{(j)} - \bar\theta^{(i)}\right|$.

(i) We have $|n - n'|^2 = 2\,(1 - \langle n, n' \rangle)$. In addition, for any $\delta \leq \delta_{\mathcal{K}}$ and $\bar\theta \in \mathcal{B}(\theta,\delta)$,

$$\begin{aligned}\langle n, n' \rangle &= \left|\theta^{(j)} - \theta^{(i)}\right|^{-1} \left\langle \theta^{(j)} - \theta^{(i)}, n' \right\rangle \\ &= \left|\theta^{(j)} - \theta^{(i)}\right|^{-1} \left\langle \left|\bar\theta^{(j)} - \bar\theta^{(i)}\right| n' + \theta^{(j)} - \bar\theta^{(j)} + \bar\theta^{(i)} - \theta^{(i)}, n' \right\rangle \\ &\geq \frac{\left|\bar\theta^{(j)} - \bar\theta^{(i)}\right|}{\left|\theta^{(j)} - \theta^{(i)}\right|} - \frac{2\delta}{\left|\theta^{(j)} - \theta^{(i)}\right|} \geq 1 - \frac{4\delta}{\left|\theta^{(j)} - \theta^{(i)}\right|} \geq 1 - \frac{4\delta}{b_{\mathcal{K}}},\end{aligned}$$

where we used (5.2) in the last equation. Therefore

$$|n - n'|^2 \leq 8\delta/b_{\mathcal{K}}, \tag{5.4}$$

(ii) Let $x \in C^{(i)}(\theta)$. We write $x - \theta^{(i)} = \left\langle x - \theta^{(i)}, n \right\rangle n + m$ where $\langle m, n \rangle = 0$. Using $\left|x - \theta^{(i)}\right|^2 = \left|\left\langle x - \theta^{(i)}, n\right\rangle\right|^2 + |m|^2$ and $x - \theta^{(j)} = \left\langle x - \theta^{(i)}, n \right\rangle n - \left|\theta^{(i)} - \theta^{(j)}\right| n + m$ we get

$$\left|x - \theta^{(j)}\right|^2 = \left|\left\langle x - \theta^{(i)}, n \right\rangle - \left|\theta^{(i)} - \theta^{(j)}\right|\right|^2 + |m|^2.$$

Since $x \in C^{(i)}(\theta)$, $\left|x - \theta^{(i)}\right| \leq \left|x - \theta^{(j)}\right|$ so that $\left|\left\langle x - \theta^{(i)}, n\right\rangle\right|^2 \leq \left|\left\langle x - \theta^{(i)}, n \right\rangle - \left|\theta^{(i)} - \theta^{(j)}\right|\right|^2$. This implies that $\left\langle x - \theta^{(i)}, n \right\rangle \leq \left|\theta^{(j)} - \theta^{(i)}\right|/2$. Therefore,

$$C^{(i)}(\theta) \subset \left\{ x \in \mathbb{R}^d, \left\langle x - \theta^{(i)}, n \right\rangle \leq \frac{1}{2} \left|\theta^{(j)} - \theta^{(i)}\right| \right\}.$$

Let now $x \in C_j(\bar{\theta}) \cap \mathcal{B}(0,\Delta)$. Following the same lines as above and using (5.3)

$$\left\langle x - \bar{\theta}^{(j)}, n' \right\rangle \geq -\frac{1}{2}\left|\bar{\theta}^{(j)} - \bar{\theta}^{(i)}\right| \geq -\frac{1}{2}\left|\theta^{(j)} - \theta^{(i)}\right| - \delta. \tag{5.5}$$

Moreover

$$\begin{aligned}\left\langle x - \theta^{(i)}, n \right\rangle &= \left\langle x - \theta^{(i)}, n - n' \right\rangle + \left\langle x - \bar{\theta}^{(j)}, n' \right\rangle + \left\langle \bar{\theta}^{(j)} - \bar{\theta}^{(i)}, n' \right\rangle + \left\langle \bar{\theta}^{(i)} - \theta^{(i)}, n' \right\rangle \\ &= \left\langle x - \theta^{(i)}, n - n' \right\rangle + \left\langle x - \bar{\theta}^{(j)}, n' \right\rangle + \left|\bar{\theta}^{(j)} - \bar{\theta}^{(i)}\right| + \left\langle \bar{\theta}^{(i)} - \theta^{(i)}, n' \right\rangle .\end{aligned}$$

Since $x, \theta^{(i)} \in \mathcal{B}(0,\Delta)$, we have by (5.2), (5.4) and (5.5)

$$\begin{aligned}\left\langle x - \theta^{(i)}, n \right\rangle &\geq -2\Delta \left|n - n'\right| - \frac{1}{2}\left|\theta^{(j)} - \theta^{(i)}\right| - \delta + \left|\theta^{(j)} - \theta^{(i)}\right| - 2\delta - \delta \\ &\geq \frac{1}{2}\left|\theta^{(j)} - \theta^{(i)}\right| - 4\delta - 4\Delta\sqrt{2/b_{\mathcal{K}}}\sqrt{\delta}.\end{aligned}$$

Therefore,

$$C^{(j)}(\bar{\theta}) \cap \mathcal{B}(0,\Delta) \subset \left\{ x \in \mathbb{R}^d, \left\langle x - \theta^{(i)}, n \right\rangle \geq \frac{1}{2}\|\theta^{(j)} - \theta^{(i)}\| - 4\delta - 4\Delta\sqrt{2/b_{\mathcal{K}}}\sqrt{\delta} \right\}.$$

Hence,

$$\begin{aligned}&C^{(i)}(\theta) \cap C^{(j)}(\bar{\theta}) \cap \mathcal{B}(0,\Delta) \\ &\subset \left\{ x \in \mathcal{B}(0,\Delta), \frac{1}{2}\left|\theta^{(j)} - \theta^{(i)}\right| - 4\delta - 4\Delta\sqrt{2/b_{\mathcal{K}}}\sqrt{\delta} \leq \left\langle x - \theta^{(i)}, n \right\rangle \leq \frac{1}{2}\left|\theta^{(j)} - \theta^{(i)}\right| \right\}.\end{aligned}$$

Finally, since $\delta_{\mathcal{K}} < 1$, we have $\delta \leq \sqrt{\delta}$, and this concludes the proof, by noticing that this last set is independent of $\bar{\theta}$.
□

Proof. [Proof of Lemma 3.4] For any compact set $\mathcal{K} \subset \Theta$, there exists C such that $\sup_{\theta \in \mathcal{K}} |H_\theta(u)| \leq C(|u| + 1)$. Therefore, H1 is satisfied with $W(u) = 1 + |u|$.
w is nonnegative and continuously differentiable on Θ; since $\nabla w = -h$, H 5-(c) is satisfied.
We now prove H5-(b); the proof is by contradiction. Assume that $\mathcal{L}$ is not included in a level set of w: then there exists a sequence $\{\theta_q, q \geq 1\}$ of $\mathcal{L}$ such that $\lim_q w(\theta_q) = +\infty$. Since $\widetilde{w}$ is bounded on $(\mathcal{B}(0,\Delta))^N$, then $\lim_q \sum_{i \neq j} \left|\theta_q^{(i)} - \theta_q^{(j)}\right|^{-2} = +\infty$ which implies that there exist a subsequence (still denoted $\{\theta_q, q \geq 1\}$) and indices $i \neq j$ such that $\lim_q \left|\theta_q^{(i)} - \theta_q^{(j)}\right| = 0$. Since $\mathcal{L}$ is closed, we proved that there exists a point $\lim_q \theta_q$ in $\mathcal{L}$ such that $\lim_q \theta_q^{(i)} = \lim_q \theta_q^{(j)}$. This is a contradiction since $\mathcal{L} \subset \Theta$.
Let us prove H4. Let $\mathcal{K} \subset \Theta$ be a compact set. We write

$$|H_\theta(x) - H_{\bar{\theta}}(x)| \leq \left|\widetilde{H}_\theta(x) - \widetilde{H}_{\bar{\theta}}(x)\right| + \lambda \sum_{i=1}^{N} \sum_{j \neq i} \left(\left| \frac{\theta^{(i)} - \theta^{(j)}}{\left|\theta^{(i)} - \theta^{(j)}\right|^4} - \frac{\bar{\theta}^{(i)} - \bar{\theta}^{(j)}}{\left|\bar{\theta}^{(i)} - \bar{\theta}^{(j)}\right|^4} \right| \right).$$

Since $\mathcal{K}$ is a compact of Θ, there exists a constant C such that

$$|H_\theta(x) - H_{\bar{\theta}}(x)| \leq \left|\widetilde{H}_\theta(x) - \widetilde{H}_{\bar{\theta}}(x)\right| + C\left|\theta - \bar{\theta}\right|.$$

For any $\theta, \bar\theta \in \mathcal{K}$ and any $x \in \mathbb{R}^d$,

$$\left|\widetilde{H}_{\bar\theta}(x) - \widetilde{H}_\theta(x)\right|^2 /4 = \sum_{i=1}^N \left[\left|\bar\theta_i - \theta^{(i)}\right|^2 \mathbb{1}_{\{C^{(i)}(\theta)\cap C^{(i)}(\bar\theta)\}}(x) \right.$$
$$\left. + \left|\theta^{(i)} - x\right|^2 \mathbb{1}_{\{C_i(\theta)\cap C^{(i)}(\bar\theta)^c\}}(x) + \left|\bar\theta^{(i)} - x\right|^2 \mathbb{1}_{\{C^{(i)}(\theta)^c \cap C^{(i)}(\bar\theta)\}}(x) \right].$$

Therefore, for any $x \in \mathcal{B}(0,\Delta)$, any $\theta \in \mathcal{K}$, and any $\bar\theta \in \mathcal{B}(0,\delta)$,

$$\begin{aligned}\left|\widetilde{H}_{\bar\theta}(x) - \widetilde{H}_\theta(x)\right|/2 &\le \sqrt{\sum_{i=1}^N \left|\bar\theta^{(i)} - \theta^{(i)}\right|^2} + \sum_{i=1}^N \sum_{j=1, j\neq i}^N \left|\theta^{(i)} - x\right| \mathbb{1}_{\{C^{(i)}(\theta)\cap C^{(j)}(\bar\theta)\}}(x) \\ &\quad + \sum_{i=1}^N \sum_{j=1, j\neq i}^N \left|\bar\theta^{(i)} - x\right| \mathbb{1}_{\{C^{(j)}(\theta)\cap C^{(i)}(\bar\theta)\}}(x) \\ &\le \delta + 2\Delta N^2 \sup_{i\neq j} \mathbb{1}_{\{C^{(i)}(\theta)\cap C^{(j)}(\bar\theta)\cap \mathcal{B}(0,\Delta)\}}(x).\end{aligned}$$

By Lemma 5.2, there exists $\delta_{\mathcal{K}}$ such that for any $\delta \le \delta_{\mathcal{K}}$, there exist a measurable set $R_{i,j}(\theta,\delta)$ such that

$$\sup_{\bar\theta \in \mathcal{B}(0,\delta)} \mathbb{1}_{\{C^{(i)}(\theta)\cap C^{(j)}(\bar\theta)\cap \mathcal{B}(0,\Delta)\}}(x) \le \mathbb{1}_{\{R_{i,j}(\theta,\delta)\}}(x).$$

Therefore, $\left|\widetilde{H}_{\bar\theta}(x) - \widetilde{H}_\theta(x)\right|/2 \le \delta + 2\Delta N^2 \sup_{i\neq j} \mathbb{1}_{\{R_{i,j}(\theta,\delta)\}}(x)$. Under E3, π is bounded on Θ. In addition, Lemma 5.2 shows that

$$\sup_{\delta\le\delta_{\mathcal{K}}} \frac{1}{\sqrt{\delta}} \sup_{\theta\in\mathcal{K}} \sup_{i\neq j} \int \mathbb{1}_{\{R_{i,j}(\theta,\delta)\}}(x)\mathrm{d}x < \infty.$$

Then, there exists C' such that for any $\delta \le \delta_{\mathcal{K}}$,

$$\sup_{\theta\in\mathcal{K}} \int \pi(\mathrm{d}x) \sup_{\{\bar\theta, \left|\bar\theta-\theta\right|\le\delta\}} \left|\widetilde{H}_{\bar\theta}(x) - \widetilde{H}_\theta(x)\right| \le C'\sqrt{\delta}.$$

Moreover, as $\sup_{\theta\in\Theta}\sup_{x\in\mathcal{B}(0,\Delta)} \left|\widetilde{H}_\theta(x)\right| < \infty$, for any $\delta \ge \delta_{\mathcal{K}}$,

$$\sup_{\theta\in\mathcal{K}} \int \pi(\mathrm{d}x) \sup_{\{\bar\theta, \left|\bar\theta-\theta\right|\le\delta\}} \left|\widetilde{H}_{\bar\theta}(x) - \widetilde{H}_\theta(x)\right| \le 2 \sup_{\Theta\times \mathrm{supp}(\pi)} \left|\widetilde{H}_\theta(x)\right| \frac{\sqrt{\delta}}{\min(1, \sqrt{\delta_{\mathcal{K}}})}.$$

Therefore H4 is satisfied with $\alpha = 1/2$. $\square$

Acknowledgments. M. Vihola was supported by the Academy of Finland (grants 250575 and 274740).

REFERENCES

[1] S. Andradóttir. A stochastic approximation algorithm with varying bounds. *Oper. Res.*, 43(6):1037–1048, 1995.

[2] C. Andrieu and E. Moulines. On the ergodicity property of some adaptive MCMC algorithms. *Ann. Appl. Probab.*, 16(3):1462–1505, 2006.
[3] C. Andrieu, E. Moulines, and P. Priouret. Stability of Stochastic Approximation under Verifiable Conditions. *SIAM J. Control Optim.*, 44(1):283–312, 2005.
[4] C. Andrieu, V. B. Tadić, and M. Vihola. On the stability of some controlled Markov chains and its applications to stochastic approximation with Markovian dynamic. *Ann. Appl. Probab.*, 25(1):1–45, 2015.
[5] C. Andrieu and M. Vihola. Markovian stochastic approximation with expanding projections. *Bernouilli*, 20(2):545–585, 2014.
[6] M.A. Arcones. Asymptotic Theory for M-Estimators over a Convex Kernel. *Econometric Theory*, 14(4):387–422, 1998.
[7] O. Bardou, N. Frikha, and G. Pags. Computing VaR and CVaR using Stochastic Approximation and Adaptive Unconstraines Importance Sampling. *Monte Carlo Methods and Applications*, 15(3):173–210, 2009.
[8] M. Benaïm, J.C. Fort, and G. Pagès. Convergence of the one-dimensional Kohonen algorithm. *Adv. in Appl. Probab.*, 30(3):850–869, 1998.
[9] A. Benveniste, M. Métivier, and P. Priouret. *Adaptive Algorithms and Stochastic Approximations*. Springer-Verlag, 1990.
[10] V.S. Borkar. *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge University Press, 2008.
[11] H. Cardot, P. Cénac, and P.A. Zitt. Recursive estimation of the conditional geometric median in Hilbert spaces. *Electronic Journal of Statistics*, 6:2535–2562, 2012.
[12] H. Cardot, P. Cénac, and P.A. Zitt. Efficient and fast estimation of the geometric median in Hilbert spaces with an averaged stochastic gradient algorithm. *Bernoulli*, 19:18–43, 2013.
[13] H. Chen and Y.M. Zhu. Stochastic Approximation procedures with random varying truncations. *Scientia Sinica (Series A)*, 29:914 – 926, 1986.
[14] M. Duflo. *Random Iterative Models*, volume 34. Springer Berlin Heidelberg, 1997.
[15] D. Egloff and M. Leippold. Quantile estimation with adaptive importance sampling. *Ann. Statist.*, 38(2):1244–1278, 2010.
[16] G. Fort, E. Moulines, and P. Priouret. Convergence of adaptive and interacting Markov chain Monte Carlo algorithms. *Ann. Statist.*, 39(6):3262–3289, 2012.
[17] P. Hall and C. Heyde. *Martingale Limit Theory and its Application*. Academic Press, 1980.
[18] B. Jourdain and J. Lelong. Robust adaptive importance sampling for normal random vectors. *Ann. Appl. Probab.*, 19(5):1687–1718, 2009.
[19] S. Kamal. Stabilization of stochastic approximation by step size adaptation. *Systems and Control Letters*, 61(4):543–548, 2012.
[20] T. Kohonen. Analysis of simple self-organising process. *Biological Cybernetics*, 44:135–140, 1982.
[21] D.P. Kroese, T. Taimre, and Z.I. Botev. *Handbook of Monte Carlo methods*. Wiley Series in Probability and Statistics, 2011.
[22] H. J. Kushner. Stochastic approximation with discontinuous dynamics and state dependent noise: w.p. 1 and weak convergence. *J. Math. Anal. Appl.*, 81(2):524 – 542, 1981.
[23] H. J. Kushner and D. Clark. *Stochastic Approximation for constrained and unconstrained systems*. Springer-Verlag, 1978.
[24] H. J. Kushner and G. Yin. *Stochastic Approximation and Recursive Algorithms and Applications*. Springer-Verlag, 2003.
[25] B. Lapeyre and J. Lelong. A framework for adaptive Monte Carlo procedures. *Monte Carlo Methods Appl.*, 17(1):77–98, 2011.
[26] S. Laruelle and G. Pagès. Stochastic approximation with averaging innovation applied to Finance. *Monte Carlo Methods Appl.*, 18(1):1 – 52, 2012.
[27] J. Lelong. Asymptotic normality of randomly truncated stochastic algorithms. *ESAIM: Probab. Stat.*, 17:105–119, 2013.
[28] V. Lemaire and G. Pagès. Unconstrained recursive importance sampling. *Ann. Appl. Probab.*, 20(3):1029–1067, 2010.
[29] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Springer, London, 1993.
[30] P. Milasevic and G.R. Ducharme. Uniqueness of the spatial median. *Ann. Statist.*, 15(3):1332–1333, 1987.
[31] G. Pagès. A space quantization method for numerical integration. *J. Comput. Appl. Math.*, 89(1):1–38, 1997.
[32] G. Pagès and H. Pham. Optimal quantization methods for nonlinear filtering with discrete-time observations. *Bernoulli*, 11(5):893–932, 2005.

[33] G. Pagès, H. Pham, and J. Printems. Optimal quantization methods and applications to numerical problems in finance . In S.T. Rachev and G.A. Anastassiou, editors, *Handbook on Numerical Methods in Finance*, pages 253–298. Birkhäuser, Boston, MA, 2004.

[34] H. Pham, W. Runggaldier, and A. Sellami. Approximation by quantization of the filter process and applications to optimal stopping problems under partial observation. *Monte Carlo methods and Applications*, 11(1):57–81, 2005.

[35] H. Robbins and S. Monro. A stochastic approximation method. *Ann. Math. Statist.*, 22:400 – 407, 1951.

[36] E. Saksman and M. Vihola. On the ergodicity of the adaptive Metropolis algorithm on unbounded domains. *Ann. Appl. Probab.*, 20(6):2178–2203, November 11 2010.

[37] V. Tadić. Stochastic approximation with random truncations, state-dependent noise and discontinuous dynamics. *Stochastics Stochastics Rep.*, 64:283 –326, 1998.

[38] D.Y. Wong, B.H. Juang, and A.H. Gray. Recent developments in vector quantization for speech processing. In *Proc. Int. Conf. Acoust., Speech, Signal Processing*, 1981.

[39] L. Younes. On the convergence of Markovian stochastic algorithms with rapidly decreasing ergodicity rates. *Stochastics Stochastics Rep.*, 65:177 – 228, 1999.