Stability of stochastic approximation
under verifiable conditions

C. ANDRIEU, É. MOULINES & P. PRIOURET

FÉVRIER 2003

Prépublication n° 791

C. Andrieu : University of Bristol, School of Mathematics, University Walk, Bristol BS8 1TW, UK.

É. Moulines : ENST, URA CNRS 820, 46 rue Barrault, F-75634 Paris Cedex 13.

P. Priouret : Laboratoire de Probabilités et Modèles Aléatoires, CNRS-UMR 7599, Université Paris VI & Université Paris VII, 4 place Jussieu, Case 188, F-75252 Paris Cedex 05.
STABILITY OF STOCHASTIC APPROXIMATION UNDER VERIFIABLE CONDITIONS

C. ANDRIEU *, É. MOULINES †, AND P. PRIOURET ²

Abstract. In this paper we address the problem of the stability of the stochastic approximation procedure. The
stability of such algorithms is known to rely heavily on the growth of the mean field at the boundary of the parameter
set and the magnitude of the stepsize used in the procedure. The conditions typically required to ensure convergence
are either too difficult to check in practice or not satisfied at all, even for simple models. The most popular technique to
circumvent this problem consists of constraining the parameter to a compact subset in the parameter space. We propose
and analyze here an alternative, based on projection on adaptive truncation sets, extending previous works in this direction.
This procedure allows for the adaptive tuning of the magnitude of the stepsize, which is key to ensuring stability. The
stability - with probability one - of the scheme is proved under a set of verifiable assumptions. We illustrate these claims
in the so-called controlled Markovian setting and present two substantial examples. The first example is related to the
minimum prediction error estimation of the parameters of stable and invertible ARMA processes and the second example
is related to controlled Markov chain Monte Carlo algorithms.

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

AMS subject classifications. 62L20,90C15

1. Introduction. In many contexts of applied statistics it is of interest to find the roots of possibly
non linear equations of the form

\[ h(\theta) = 0, \quad \theta \in \Theta, \quad (1.1) \]

for some mapping \( h : \Theta \rightarrow \mathbb{R}^d \), where \( \Theta \) is an open subset of \( \mathbb{R}^d \). Most of the methods for solving the
previous equation are iterative, i.e. produce a sequence of iterates \( (\theta_n, n \geq 0) \) which eventually converges
to the set of solutions of (1.1),

\[ \mathcal{L}' = \{ \theta \in \Theta, h(\theta) = 0 \}. \quad (1.2) \]

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

\[ \theta_0 \in \Theta, \quad \theta_{n+1} = \theta_n + \gamma_n Y_{n+1}, \quad n = 0, 1, \ldots, \quad (1.3) \]

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

\[ Y_n = h(\theta_n) + \xi_n \quad (1.4) \]

---

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

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

²Université Pierre & Marie Curie, Laboratoire de Probabilités et Modélisation Aléatoire, URA CNRS 224, F 75252
PARIS Cédex 05 (priouret@ccr.jussieu.fr)
which will be referred to as the *noise sequence*. Convergence of SA has been studied under various sets of assumptions for the mean field \( \mathbf{h} \) and the noise \( \xi \) since the early work by [20]; see e.g. [4] (hereafter BMP), [15], [21], [14] and the references therein. Essentially, convergence of the SA sequence can be established toward an attractive subset provided that the sequence \( (\theta_n, n \geq 0) \) is w.p. 1 in a compact subset of \( \Theta \) (boundedness) and is w.p. 1 infinitely often in the domain of attraction of this attractive subset (stability-recurrence).

Showing in practice that \( (\theta_n, n \geq 0) \) satisfies the boundedness and the stability-recurrence conditions proves to be a difficult task. The available results hold under conditions which are still restrictive, despite recent advances (see [6], [5] and references therein). To the best of the authors’ knowledge, there are no generally applicable results which readily imply these conditions. As a result of this lack of generality and applicability, one is in practice faced with the problem of choosing these conditions on a case by case basis. This major drawback has motivated the design of modified Robbins-Monro recursions. Probably the most widely used method in practice consists of constraining the sequence \( (\theta_n, n \geq 0) \) to some compact set \( \mathcal{Q} \subset \Theta \) by means of a reprojection onto \( \mathcal{Q} \). This method has been thoroughly investigated in [21] (see also [7] and the references therein). Although relatively easy to implement, and sound when constraints about the system considered are available a priori, this approach becomes impractical and questionable in many situations of interest. This is in particular the case for high dimensional problems.

We propose here a modification of the Robbins-Monro procedure based on truncations on randomly varying boundaries, in the spirit of the procedure proposed in the seminal papers by [10] and [8]. Truncation on random sets has been applied with success to various stochastic approximation problems: see e.g. [27], [12] for SA for which the noise sequence is a martingale difference and [9] for dynamic SA. The application of random truncation to state dependent noise (see Section 2 for a definition) is a difficult problem and the results are scarce, and have been obtained under conditions more stringent than those considered in this paper, see [29], [28] and [11].

The algorithm that we propose differs from the original procedure proposed by [10] and [8] in the way the stepsizes are updated after each reinitialization, therefore allowing for more flexibility. The main result of this paper, stated and proved in Section 3, is that under the assumption that there exists a global Lyapunov function that controls the excursions of the parameter sequence, the proposed algorithm is stable w.p. 1 and is recurrent in the domain of attraction of the stationary points. Contrary to usual results, the conditions required on the growth of the Lyapunov functions and the mean field \( \mathbf{h} \) when \( \theta \) approaches the boundaries of the parameter set are minimal. As a consequence the results are applicable in quite general settings. A byproduct of the proof is an explicit bound for the tail probability of the number of re-projections, which is found to be super-exponential in most cases. Another byproduct of this work is a global convergence theorem, which stems almost immediately from the stability-recurrence results using classical theorems on SA convergence. To illustrate our findings, two substantial examples are presented. First (see Section 6), we analyze the recursive estimation of the parameters of a stable invertible ARMA process (using a variant of the recursive prediction error method). Secondly (see Section 7), we propose a novel (and more widely applicable) analysis of the convergence of an adaptive Markov chain Monte Carlo (MCMC) recently proposed and analysed in [16].

2. Setting and Algorithm description. The parameter set \( \Theta \) is assumed to be an open subset of \( \mathbb{R}^d \) equipped with the Borel \( \sigma \)-field \( \mathcal{B}(\mathbb{R}^d) \). \(^1\) Let \( X \) be the state space, assumed to be equipped with a countably generated \( \sigma \)-field, \( \mathcal{B}(X) \). Let \( (\rho_n, n \geq 0) \) be a sequence of non-negative numbers, the stepsizes. Let \( H : \Theta \times X \to \Theta \) be a function (measurability w.r.t the appropriate \( \sigma \)-fields is always implicitly

---

\(^1\) For any set \( A \) endowed with a topology, measurability is always taken to mean Borel measurability on the corresponding Borel \( \sigma \)-field, denoted by \( \mathcal{B}(A) \).
assumed). It is assumed that there is a family \((\pi_\theta, \theta \in \Theta)\) of probability measures on \((X, \mathcal{B}(X))\) such that, for every \(\theta \in \Theta\), \(\int_X H(\theta, x) \pi_\theta(dx) < \infty\) and we set \(h(\theta) = \int_X H(\theta, x) \pi_\theta(dx)\). Following the, perhaps, non-standard practice of [4], we often write \(H_\theta(x)\) as an equivalent expression for \(H(\theta, x), h_\theta\) for \(h(\theta)\), etc.. By convention, we set \(\inf(\emptyset) = \infty\). For a Markov chain with transition probability \(P\), any \(A \in \mathcal{B}(X)\) and \(x \in X\) we denote \(P^n(x, A)\) the \(n\) step probability transition function. For a probability measure \(\mu\), \(x \in X\) and a Borel function \(f : X \to \mathbb{R}\), we define

\[
P^n f(x) = P^n(x, f) = \int_X P^n(x, dy) f(y), \quad n = 1, 2, \ldots \quad \text{and} \quad \mu(f) = \int_X \mu(dx) f(x).
\]

Let \((P_\theta, \theta \in \Theta)\) be a family of Markov kernels indexed by the parameter \(\theta\) such that, for every \(\theta \in \Theta\), \(\pi_\theta\) is the stationary distribution of \(P_\theta\), i.e. for any \(A \in \mathcal{B}(X)\),

\[
\pi_\theta P_\theta(A) := \int_X \pi_\theta(dx) P_\theta(x, A) = \pi_\theta(A).
\]

As outlined in the introduction, one of the main difficulty with SA is identify practical conditions upon which the parameter \((\theta_n, n \geq 0)\) stays in a compact subset of \(\Theta\). In this paper, we introduce a practical procedure inspired from the truncation on randomly varying boundaries of [8]. This procedure amounts to monitoring the excursions of \((\theta_n, n \geq 0)\) outside an increasing sequence \((K_q, q \geq 0)\) of compact subsets of \(\Theta\):

\[
\bigcup_{q \geq 0} K_q = \Theta, \quad \text{and} \quad K_q \subset \text{int}(K_{q+1}), \quad q \geq 0, \quad (2.1)
\]

where \(\text{int}(A)\) denotes the interior of set \(A\). Let \((\gamma_k, k \geq 0)\) and \((\varepsilon_k, k \geq 0)\) be two monotone non-increasing sequences of positive numbers and \(K\) a subset of \(X\). Let \(\phi : \mathbb{N} \to \mathbb{Z}\) be a function such that, for all \(k \geq 0\), \(\phi(k) \geq 1 - k\). We construct a sequence \((\theta_n, X_n), n \geq 0\) as follows:

**Algorithm 1** SA with truncation on random boundaries

1. Set \(\kappa_0 = 0, \mu_0 = 0, \zeta_0 = 0, \theta_0 \in K_0\) and \(X_0 = x_0 \in K \subset X\).
2. For \(n \geq 1\), compute \(\bar{\theta} = \theta_{n-1} + \gamma_{\kappa_{n-1}+1} H(\theta_{n-1}, \bar{X})\), where \(\bar{X}\) is sampled from \(P_{\theta_{n-1}}(X_{n-1}, \cdot)\).
3. If \(\bar{\theta} \in K_{\kappa_{n-1}}\) and \(|\bar{\theta} - \theta_{n-1}| \leq \varepsilon_{\kappa_{n-1}}\) then set \((\theta_n, X_n) = (\bar{\theta}, \bar{X})\), \(\kappa_n = \kappa_{n-1}, \nu_n = \nu_{n-1} + 1\) and \(\zeta_n = \zeta_{n-1} + 1\).
4. Else set \((\theta_n, X_n) = (\bar{\theta}, \bar{X}) \in K_0 \times K, \kappa_n = \kappa_{n-1} + 1, \nu_n = 0, \zeta_n = \zeta_{n-1} + \phi(\nu_{n-1})\).

Note that (i) \((\kappa_n, n \geq 0)\) is the index of the active truncation set (also equal to the number of reinitializations before \(n\)) (ii) \((\nu_n, n \geq 0)\) is the number of iterations since the last reinitialization and (iii) \((\zeta_n, n \geq 0)\) is the current index in the stepsize sequence.

When a reprojection occurs, the r.v. \((\bar{\theta}, \bar{X})\) may be chosen in the following ways:

1. \((\bar{\theta}, \bar{X})\) can be random, independent on the past (random re-initialization)
2. \((\bar{\theta}, \bar{X})\) may be set to some fixed value \((\theta, x)\), in which case the algorithm is always started afresh from the same initial point
3. \((\bar{\theta}, \bar{X})\) can be chosen as the projection of \((\bar{\theta}, \bar{X})\) onto \(K_0 \times K\), i.e. \((\bar{\theta}, \bar{X}) = \Phi(\bar{\theta}, \bar{X})\), where \(\Phi\) is some measurable function from \(\Theta \times X \to K_0 \times K\).

It is easily checked that \((Z_n = (X_n, \theta_n, \kappa_n, \nu_n, \zeta_n), n \geq 0)\) is an homogeneous Markov chain on the product space \(Z = X \times \Theta \times \mathbb{N}\). We denote \(\mathcal{G} = (\mathcal{G}_k, k \geq 0)\) the filtration associated to this process. We denote
\( \mathbb{P}_z \) the probability measure associated to this chain started at \( z \in \mathbb{Z} \). By convention, we set
\[
\mathbb{P}_{x_0, \theta_0} = \mathbb{P}_{x_0, \theta_0, 0, 0, 0}.
\] (2.2)
This probability measure depends upon the deterministic sequences \( \gamma = (\gamma_n, n \geq 0) \) and \( \epsilon = (\epsilon_n, n \geq 0) \); the dependence will be implicit here.

This algorithm is reminiscent of the projection on random varying boundaries proposed in [10], [8]. When the current iterate wanders outside the active truncation set or when the difference between two successive values of the parameter is larger than a time-dependent threshold, then the algorithm is reinitialised with a smaller initial value of the stepsize and a larger truncation set.

Various choices for the function \( \phi \) can be considered. For example, the choice \( \phi(k) = 1 \) for all \( k \in \mathbb{N} \) coincides with the procedure proposed in [8]; in this case \( \zeta_n = n \). Another sensible choice consists of setting \( \phi(k) = 1 - k \) for all \( k \in \mathbb{N} \) in which case the number of iterations between two successive reinitialisations is not taken into account. In the latter case, we have \( \zeta_n = \kappa_n + \nu_n \). Naturally, many variations on this theme can be considered. We suggest here two possible extensions. First, the proposed procedure can be understood as a way of automatically learning the adequate scaling of the stepsize that ensures stability. In the light of this comment it is possible to suggest the following generalisation of the algorithm. Let \( (\gamma_{k,l} , k \geq 0, l \geq 0) \) be an array of stepsizes. Then, when a reprojection occurs, instead of jumping ahead in a unique sequence of stepsizes, it is possible to simply change the sequence of stepsizes. In this case, index \( l \) could for example coincide with the index of the current truncation set. This scheme allows obviously for an even greater flexibility. Another important variant of the proposed scheme consists of reinitialising the algorithm when \( |\theta - \theta_{n-1}| \geq \epsilon_{n-1} \) without changing the truncation set. In either cases the proof of convergence follows using the same type of arguments as those presented in this paper.

We define recursively \( (T_n, n \geq 0) \) the sequence of successive reinitialisation times
\[
T_{n+1} = \inf \{ k \geq T_n + 1, \nu_k = 0 \}, \quad \text{where} \quad T_0 = 0.
\] (2.3)
The key result (see Proposition 3.2), states that under appropriate conditions which are much weaker than those typically imposed to guarantee the convergence of unbounded stochastic approximation (see [20], [4], [21], [14] and the references therein)
\[
\inf_{(x, \theta) \in K \times \mathcal{K}_0} \mathbb{P}_{x, \theta} \left( \sup_{n \geq 0} \kappa_n < \infty \right) = \inf_{(x, \theta) \in K \times \mathcal{K}_0} \mathbb{P}_{x, \theta} \left( \bigcup_{n=0}^{\infty} \{ T_n = \infty \} \right) = 1,
\]
i.e., the number of reinitializations of the procedure described above is finite \( \mathbb{P}_{x, \theta} \)-a.e., for every \( (x, \theta) \in K \times \mathcal{K}_0 \). When \( \phi(k) = 1 \), it is even possible to derive an explicit bound for the tail probability
\[
\sup_{(x, \theta) \in K \times \mathcal{K}_0} \mathbb{P}_{x, \theta} \left[ \sup_{n \geq 0} I(T_n < \infty) \geq k \right],
\]
which allows one to derive \( L^p \) or exponential bounds, under additional technical assumptions. Convergence will then follow using standard results valid for bounded stochastic approximation algorithms.

3. Lyapunov functions and the key deterministic stability result. An element \( v \) of \( \mathbb{R}^d \) is denoted by its column vector \( v \) and its transpose is denoted by \( v^\top \). For elements \( v, w \) of \( \mathbb{R}^d \), we denote \( \langle v, w \rangle \) their Euclidian dot product, so that \( |v| = \sqrt{\langle v, v \rangle} \) denotes the Euclidean norm of \( v \).

(A1) The mean field \( h : \Theta \rightarrow \mathbb{R}^d \) is continuous and there exists a continuously differentiable function \( w : \Theta \rightarrow [0, \infty) \) such that
1. For any integer $M$, the level set $\mathcal{W}_M = \{ \theta \in \Theta, w(\theta) \leq M \} \subset \Theta$ is compact,
2. There exists $M_0$ such that
\[ \mathcal{L} = \{ \theta \in \Theta, \left\langle \nabla w(\theta), h(\theta) \right\rangle = 0 \} \subset \text{int}(\mathcal{W}_{M_0}) \] (3.1)
and for any $\theta \in \Theta \setminus \mathcal{W}_{M_0}$, \( \left\langle \nabla w(\theta), h(\theta) \right\rangle < 0. \)

By convention, we set $w(\theta) = \infty$ when $\theta \notin \Theta$. The function $w$ in (A1) is a Lyapunov function outside a compact set for the deterministic ODE
\[ \dot{\theta}(t) = h(\theta(t)), \quad t \in \mathbb{R}^+. \] (3.2)

Conditions of the form (A1) are standard to prove the asymptotic stability of the solution for this ODE. When the mean field $h = -\nabla J$ is the gradient of some mapping $J : \Theta \to (0, \infty)$, it is customary to set $w = J$, in which case the condition $\left\langle \nabla J, h \right\rangle = -|\nabla J|^2 \leq 0$ is automatically satisfied for all $\theta$. Because the stated assumptions share many similarities with this particular setting, they are often referred to as pseudo-gradient conditions. Though we will not strictly follow the so-called ODE approach outlined in [20], the existence of $w$ is crucial to control the stability of the iterates. The following elementary lemma (see proof in Appendix D) plays a key role in the sequel.

**Lemma 3.1.** Assume (A1), then
1. For any compact $\mathcal{K} \subset \Theta \setminus \mathcal{W}_{M_0}$, where $\mathcal{W}_{M_0}$ is defined in (3.1), and any $\delta_{\mathcal{K}} < \inf_{\mathcal{K}} |\left\langle \nabla w, h \right\rangle|$, there exist $\lambda_{\mathcal{K}} > 0, \beta_{\mathcal{K}} > 0$, such that, for all $\lambda, 0 \leq \lambda \leq \lambda_{\mathcal{K}}$, all $c, |c| \leq \beta_{\mathcal{K}}$, and all $a \in \mathcal{K}$, we have $w(a + \lambda h(a) + \lambda c) \leq w(a) - \lambda \delta_{\mathcal{K}}$.
2. For all $M > M_0$, there exist $\lambda_M > 0$, $\beta_M > 0$ such that, for all $a$ verifying $w(a) \leq M$, for all $\lambda, 0 \leq \lambda \leq \lambda_M$, for all $c, |c| \leq \beta_M$, we have $w(a + \lambda h(a) + \lambda c) \leq M$.

This simple lemma above implies a first deterministic stability result.

**Proposition 3.2.** There exist $\delta_0 > 0$, $\lambda_0 > 0$ such that, for all $n \geq 1$, for all $\theta_0 \in \mathcal{W}_{M_0}$, for all sequences $(\rho_k, \xi_k) \in \mathbb{R}^+ \times \mathbb{R}^d$ verifying, for $k \in \{1, \ldots, n\}$, $|\rho_k| \leq \lambda_0$ and $\sum_{j=1}^{k} \rho_j \xi_j \leq \delta_0$, we have, for $k \in \{1, \ldots, n\}$ and $\theta_k = \theta_{k-1} + \rho_k h(\theta_{k-1}) + \rho_k \xi_k$, $w(\theta_k) \leq M_0 + 1$.

**Proof.** Let $M$ be such that $M_0 < M < M_0 + 1$. Lemma 3.1 shows that there exists $\lambda_M < 0$, $\beta_M < 0$ such that, for all $a, \lambda, c$ verifying $w(a) \leq M$, $0 \leq \lambda \leq \lambda_M$, $|c| \leq \beta_M$,
\[ w(a + \lambda h(a) + \lambda c) \leq M. \] (3.3)

There exists $\delta_M > 0$ such that, for all $a, b$ verifying $w(a) \leq M_0 + 1$, $|b| \leq \delta_M$, we have
\[ |h(a + b) - h(a)| \leq \beta_M \quad \text{and} \quad |w(a + b) - w(a)| \leq M_0 + 1 - M. \] (3.4)

We set $\delta_0 = \delta_M$ and $\lambda_0 = \lambda_M$. We will show that, if $w(\theta_j) \leq M_0 + 1$ for $j \in \{1, \ldots, k - 1\}$ \((k \leq n)\) then $w(\theta_k) \leq M_0 + 1$. Define $\bar{\theta}_j$ for $j \in \{1, \ldots, k\}$ by
\[ \bar{\theta}_j = \bar{\theta}_{j-1} + \rho_j h(\theta_{j-1}), \quad \bar{\theta}_0 = \theta_0. \] (3.5)

By construction, $\theta_j - \bar{\theta}_j = \theta_{j-1} - \bar{\theta}_{j-1} + \rho_j \xi_j$, which implies that
\[ \theta_j - \bar{\theta}_j = \sum_{i=1}^{j} \rho_i \xi_i. \] (3.6)
Note also that
\[
\tilde{\theta}_j = \tilde{\theta}_{j-1} + \rho_j h(\tilde{\theta}_{j-1}) + \rho_j (h(\theta_{j-1}) - h(\tilde{\theta}_{j-1})).
\]
(3.7)
For \( j \in \{1, \ldots, k\} \), \( |\theta_{j-1} - \tilde{\theta}_{j-1}| \leq \delta_M \) and \( |h(\theta_{j-1}) - h(\tilde{\theta}_{j-1})| \leq \beta_M \) and (3.3) implies that, for \( j \in \{1, \ldots, k\} \), \( w(\tilde{\theta}_j) \leq M \) by an obvious induction. In particular, \( w(\tilde{\theta}) \leq M \). Thus, since \( |\theta_k - \tilde{\theta}| \leq \delta_M \), (3.4) shows that \( |w(\theta_k) - w(\tilde{\theta}_k)| \leq M_0 + 1 - M \) and we have \( w(\theta_k) \leq M_0 + 1 \), which concludes the proof.

The result above shows that it is possible under assumption (A1) that the parameter \( \theta_n \) always stays in \( \mathcal{W}_{M_0 + 1} \) provided that we are able to control the fluctuations of the sums \( \sum_{k=1}^n \rho_k \xi_k \) before the first exit time from \( \mathcal{W}_{M_0 + 1} \). This type of condition is well suited for the stability analysis of SA with state dependent noise, as will be seen below. There is however here an additional technical difficulty stemming from the fact that, in the state-dependent noise case, we have to control both that the parameter stays bounded and that the difference between two successive parameter values is “small”.

4. Control of the fluctuations. Before pursuing further, some additional definitions are needed. Let \( \rho = (\rho_k, k \geq 0) \) be a monotone non-increasing sequence. We extend both the parameter space and the state-space with two cemetery points, \( \theta_c \notin \Theta \) and \( x_c \notin X \), and define \( \bar{\Theta} = \Theta \cup \{\theta_c\}, \bar{X} = X \cup \{x_c\} \).

Let  \( \theta_0 = \theta \in \Theta, X_0 = x \in X \), and for \( k \geq 1 \) define recursively the sequences \( \theta_k \) and \( X_k \) as follows. If \( \theta_{k-1} = \theta_c \), then set \( \theta_k = \theta_c \) and \( X_k = x_c \). Otherwise, draw \( X_k \) according to \( P_{\theta_{k-1}}(X_{k-1}, \cdot) \) compute
\[
\eta = \theta_{k-1} + \rho_k H(\theta_{k-1}, X_k)
\]
and set:
\[
\theta_k = \begin{cases} 
\eta & \text{if } \eta \in \Theta, \\
\theta_c & \text{if } \eta \notin \Theta.
\end{cases}
\]

The sequence \( \{(X_n, \theta_n), n \geq 0\} \) is a non-homogeneous Markov chain on the product space \( \bar{\Theta} \times \bar{X} \). This non-homogeneous Markov chain defines a probability measure on the canonical state space \( \bar{\Theta} \times \bar{X} \) equipped with the canonical product \( \sigma \)-algebra. We denote \( \mathcal{F} = (F_n, n \geq 0) \) the canonical filtration of this Markov chain, \( F_n = \sigma((X_t, \theta_t), t \in \{0, \ldots, n\}) \). We denote \( \mathbb{P}_{x, \theta} \) the probability measure associated with this Markov chain starting from the initial condition \( X_0 = x \) and \( \theta_0 = \theta \), \((x, \theta) \in \bar{X} \times \bar{\Theta} \).

Let \( K \) be a compact subset of \( \Theta \) and let \( \epsilon = (\epsilon_n, n \geq 0) \) be a non-increasing sequence of positive numbers. Define, for \( 1 \leq l \leq n \) the partial sum
\[
S_{l,n}(\epsilon, \rho, \mathcal{K}) = I\{\sigma(\mathcal{K}) \land \nu(\epsilon) \geq n\} \sum_{k=l}^n \rho_k (H(\theta_{k-1}, X_k) - h(\theta_{k-1})).
\]
(4.1)
where
\[
\sigma(\mathcal{K}) = \inf\{k \geq 1, \theta_k \notin \mathcal{K}\},
\]
(4.2)
\[
\nu(\epsilon) = \inf\{k \geq 1, |\theta_k - \theta_{k-1}| \geq \epsilon_k\}.
\]
(4.3)

Denote, for any integer \( l \), and any \( \delta \geq 0 \),
\[
A_l(\delta, \epsilon, \rho, \mathcal{K}, x) = \sup_{\theta \in \mathcal{K}_0} \mathbb{P}_{x, \theta}^\rho \left( \sup_{n \geq l} |S_{l,n}(\epsilon, \rho, \mathcal{K})| > \delta \right),
\]
(4.4)
\[
B(\epsilon, \rho, \mathcal{K}, x) = \sup_{\theta \in \mathcal{K}_0} \mathbb{P}_{x, \theta}^\rho (\nu(\epsilon) < \sigma(\mathcal{K})),
\]
(4.5)
where \( \mathcal{K}_0 \) is defined in (2.1). From now on, for the sake of simplicity and clarity, we specialize our results to the case where \( \phi(k) = 1 \). For a sequence \( a = (a_k, k \geq 0) \) and an integer \( l \), define \( a^{-l} = \cdots \)
as \(a_k^{s,l} = a_k^{k+1}\). For simplicity, we denote \(A_1(\delta, \epsilon, \rho, x) = A_1(\delta, \epsilon, \rho, W_{m+1}, x), B(\epsilon, \rho, x) = B(\epsilon, \rho, W_{m+1}, x), \sigma = \sigma(W_{m+1}), \) where \(m+1\) is given in (A1).

Set \(q_0 \geq 1\) such that \(W_{m+1} \in K_{q_0}\) and \(\gamma_{q_0} \leq \lambda_0\), where \(\lambda_0\) is given in Proposition 3.2. The existence of such a \(q_0\) follows from (i) the sets \(W_M\) are compact and \(\bigcup_{q} K_q\) is an increasing covering of \(\Theta\) (ii) \(\gamma_k \downarrow 0\).

**Proposition 4.1.** Assume (A1). Let \(\gamma = (\gamma_k, k \geq 0)\) and \(\epsilon = (\epsilon_k, k \geq 0)\) be monotone non-increasing sequences and let \((\theta_k, X_k)\) be defined by Algorithm 1. For \(n \geq q_0\), we have

\[
\sup_{(x, \theta) \in K \times K_0} P_{x, \theta}(T_n < \infty) \leq \prod_{l=0}^{l=n-1} \sup_{x \in K, k \geq l} (A_1(\delta_0, \epsilon^-_k, \gamma^-_{k}, x) + B(\epsilon^-_k, \gamma^-_{k}, x)),
\]

where \(K\) is defined in Algorithm 1, the sequence of stopping times \(T_n\) is given in (2.3), the constant \(\delta_0\) is defined in Proposition 3.2, and \(P_{x, \theta}\) is defined in (2.2).

**Proof.** By the strong Markov property, we have, for \(n \geq q_0\),

\[
\sup_{(x, \theta) \in K \times K_0} P_{x, \theta}(T_n < \infty|G_{T_{n-1}}) I(T_{n-1} < \infty) \leq \sup_{(x, \theta) \in K \times K_0} P_{x, \theta}(T_{n-1} < \infty|G_{T_{n-1}}) I(T_{n-1} < \infty),
\]

where we have used that, according to Algorithm 1, on \(\{T_{n-1} < \infty\}, X_{T_{n-1}} \in K, \theta_{T_{n-1}} \in K_0, \tau_{T_{n-1}} = n-1\) and \(\nu_{T_{n-1}} = 0\). For any integers \(p\) and \(q\), any \(x \in X\) and any \(\theta \in \Theta\), we have

\[
P_{x, \theta, n, p}(T_1 < \infty) = P_{x, \theta}^\tau(n\nu < p) (\sigma(K) \wedge \nu(\epsilon^\tau) < \infty).
\]

For \(q \geq q_0\), we have \(W_{m+1} \subset K_{q_0} \subset K_q\), showing that (recall that \(\sigma = \sigma(W_{m+1})\))

\[
P_{x, \theta}^\tau(n\nu < p) (\sigma(K) \wedge \nu(\epsilon^\tau) < \infty) \leq P_{x, \theta}^\tau(n\nu(\epsilon^\tau) < \infty) \leq P_{x, \theta}^\tau(\sigma < \infty, \sigma \wedge \nu(\epsilon^\tau) < \infty) \leq P_{x, \theta}^\tau(\sigma < \infty, \sigma \leq \nu(\epsilon^\tau)) + P_{x, \theta}^\tau(\nu(\epsilon^\tau) < \sigma).
\]

By Proposition 3.2, \(\{\sigma = n, n \leq \nu(\epsilon^\tau)\} \subset \left\{\sigma_{n} \in \{1, \ldots, n\} \mid S_{1,n}(\epsilon^\tau) \gamma^\tau > \delta_0\right\}\), where \(S_{1,n}(\epsilon, \rho)\) is a shorthand notation for \(S_{1,n}(\epsilon, \rho, W_{m+1})\) (see the definition in (4.1)) which implies that

\[
P_{x, \theta}^\tau(\sigma \wedge \nu(\epsilon^\tau) < \infty) \leq P_{x, \theta}^\tau(\sup_{n \geq 1} |S_{1,n}(\epsilon^\tau, \gamma^\tau)| > \delta_0) + P_{x, \theta}^\tau(\nu(\epsilon^\tau) < \sigma).
\]

Combining the results above, we have,

\[
\sup_{(x, \theta) \in K \times K_0} P_{x, \theta}(T_n < \infty|G_{T_{n-1}}) I(T_{n-1} < \infty) \leq \sup_{x \in K, \theta \in \Theta} (A_1(\delta_0, \epsilon^{-}\tau_{n-1}, \gamma^{-}\tau_{n-1}, x) + \epsilon^{-}\tau_{n-1}, \gamma^{-}\tau_{n-1}, x)) I(T_{n-1} < \infty).
\]

Since \(\tau_{n-1} = n - 1\), the proof follows by a straightforward induction. \(\square\)

**Corollary 4.2.** We have

\[
\limsup_{k \to \infty} k^{-1} \log \left(\sup_{(x, \theta) \in K \times K_0} P_{x, \theta}(\sup_{\kappa \geq k} \kappa)\right) = \limsup_{k \to \infty} k^{-1} \log \left(\sup_{(x, \theta) \in K \times K_0} P_{x, \theta}(T_k < \infty)\right) \leq \log C(\delta_0, \epsilon, \gamma).
\]
where
\[
C(\delta_0, \varepsilon, \gamma) = \lim_{k \to \infty} \sup_{x \in \mathbb{R}} \left( A_1(\delta_0, e^{-\gamma k}, x) + B(e^{-\gamma k}, \gamma e^{-k}, x) \right)
\] (4.6)

In particular, if there exist sequences \( \varepsilon = (\varepsilon_k, k \geq 0) \) and \( \gamma = (\gamma_k, k \geq 0) \) such that \( C(\delta_0, \varepsilon, \gamma) < 1 \), the tail distribution of the number of reinitialisations is geometrically bounded. When \( C(\delta_0, \varepsilon, \gamma) = 0 \), the distribution of the number of restarts is super-exponential.

As seen below, bounds for \( C(\delta_0, \varepsilon, \gamma) \) stem from Wald’s maximal inequalities for martingales combined with Burkholder inequality (see [1], [24], Chapter 17). We will derive below conditions on the family of Markov kernels \( (P_{\theta}, \theta \in \Theta) \) for that purpose.

Define, for \( V : X \to [1, \infty) \) and \( g : X \to \mathbb{R} \) the norm
\[
\|g\|_V = \sup_{x \in X} \frac{|g(x)|}{V(x)}.
\] (4.7)

Consider the following assumptions

(A2) For any \( \theta \in \Theta \), the Poisson equation \( g - P_{\theta} g = H_{\theta - \pi_{\theta}}(H_{\theta}) \) has a solution \( g_0 \). There exist a function \( W : X \to [1, \infty] \) such that \( \{x \in X, W(x) < \infty\} \neq \emptyset \), constants \( \alpha \in (0, 1] \), \( p \geq 2 \) such that for any compact subset \( K \subset \Theta 

- (i) \sup_{\theta \in K} \|H_{\theta}\|_W < \infty, \quad (4.8)

- (ii) \sup_{\theta \in K} (\|g_0\|_W + \|P_{\theta} g_0\|_W) < \infty, \quad (4.9)

- (iii) \sup_{(\theta, \theta') \in W} \|\theta - \theta'\|^{-\alpha} \left( \|g_0 - g_{\theta'}\|_W + \|P_{\theta} g_0 - P_{\theta'} g_{\theta'}\|_W \right) < \infty. \quad (4.10)

- (ii) there exist constants \( (C_k, k \geq 0) \) such that, for any \( k \in \mathbb{N} \), for any sequence \( \rho = (\rho_k, k \geq 0) \) and for any \( x \in X, \)
\[
\sup_{\theta \in K} \mathbb{E}^{\rho}_{x, \theta}[W^p(X_k)I(\sigma(K) \geq k)] \leq C_k W^p(x), \quad (4.11)
\]

- (iii) there exist \( \epsilon > 0 \) and a constant \( C \) such that for any sequence \( \rho = (\rho_k, k \geq 0) \) and for any \( x \in X, \)
\[
\sup_{\theta \in K} \mathbb{E}^{\rho}_{x, \theta}[W^p(X_k)I(\sigma(K) \land \nu_\epsilon \geq k)] \leq CW^p(x). \quad (4.12)
\]

where
\[
\nu_\epsilon = \inf\{k \geq 1, |\theta_k - \theta_{k-1}| > \epsilon\}. \quad (4.13)
\]

Assumption (A2) states the existence and the regularity of the solutions of the Poisson equation. These conditions are standard and are implied under a variety of ergodicity conditions for the Markov chain with transition kernel \( (P_\theta, \theta \in \Theta) \) and the regularity of the function \( \theta \to P_\theta \). We stress that the function \( W \) is global but that the bounds (4.8), (4.9), (4.10), (4.11), (4.12) depend on the particular compact \( K \) under consideration.

In Section 5 conditions that guarantee (A2) are given. We have
Lemma 4.3. Assume (A2-iii). Let $\mathcal{K}$ be a compact subset of $\Theta$ and $s \in \mathbb{N}$. There exists a constant $C$ such that for any sequence $\epsilon = (\epsilon_k, k \geq 0)$ verifying $0 < \epsilon_k \leq \epsilon$ for all $k \geq s$ (where $\epsilon$ is defined in (A2-iii)), for any sequence $\rho = (\rho_k, k \geq 0)$ and for any $x \in \mathcal{X}$,

$$\sup_{\theta \in \mathcal{K}} \sup_{k \geq 0} \mathbb{E}_{\theta}^{\rho} [W^p(X_k)I\{\sigma(\mathcal{K}) \land \nu(\epsilon) \geq k\}] \leq CW^p(x).$$

Proof. Under (A2), there exists a constant $C$ such that, for any sequence $\rho = (\rho_k, k \geq 0)$ and any $x \in \mathcal{X}$ we have

$$\sup_{\theta \in \mathcal{K}} \mathbb{E}_{x, \theta}^{\rho} [W^p(X_k)I\{\sigma(\mathcal{K}) \land \nu_s \geq k\}] \leq CW^p(x),$$

where $\nu_s$ is defined in (4.13). For any sequence $\epsilon = (\epsilon_k, k \geq 0)$ such that $\epsilon_k \leq \epsilon$ for any $k \geq s$,

$$\mathbb{E}_{x, \theta}^{\rho} [W^p(X_{k+s})I\{\sigma(\mathcal{K}) \land \nu(\epsilon) \geq k+s\}] = \mathbb{E}_{x, \theta}^{\rho} \left[ \sup_{\theta' \in \mathcal{K}} \mathbb{E}_{x, \theta'}^{\rho_{k-s}} [W^p(X_k)I\{\sigma(\mathcal{K}) \land \nu(\epsilon^{k-s}) \geq k]I(\sigma(\mathcal{K}) \land \nu(\epsilon) \geq s)) \right]$$

$$\leq \mathbb{E}_{x, \theta}^{\rho} \left[ \sup_{\theta' \in \mathcal{K}} \mathbb{E}_{x, \theta'}^{\rho_{k-s}} [W^p(X_k)I\{\sigma(\mathcal{K}) \land \nu_s \geq k]I[\sigma(\mathcal{K}) \geq s)] \right]$$

$$\leq C \mathbb{E}_{x, \theta}^{\rho} [W^p(X_s)I(\sigma(\mathcal{K}) \geq s)],$$

and the proof is concluded by (A2-ii). \[\Box\]

Proposition 4.4. Assume (A2). Let $\mathcal{K}$ be a compact subset of $\Theta$ and let $\rho = (\rho_k, k \geq 0)$, $\epsilon = (\epsilon_k, k \geq 0)$ be two non-increasing sequences of positive numbers, such as $\lim_{k \to \infty} \epsilon_k = 0$. Then, for $p$ as defined in (A2)-(iii),

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

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

(4.14)

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

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

(4.15)

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

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

Using $I\{\sigma(\mathcal{K}) \land \nu(\epsilon) \geq k\} = I\{\sigma(\mathcal{K}) \land \nu(\epsilon) \geq k+1\} + I\{\sigma(\mathcal{K}) \land \nu(\epsilon) = k\}$, we may write $T_n = \sum_{n}^{\infty} T_n^{(i)}$...
where

\[
T_{n}^{(1)} = \sum_{k=1}^{n} \rho_{k} \left( g_{\theta_{k-1}}(X_k) - P_{\theta_{k-1}} g_{\theta_{k-1}}(X_{k-1}) \right) I\{\sigma(K) \land \nu(\epsilon) \geq k\},
\]

\[
T_{n}^{(2)} = \sum_{k=1}^{n-1} \rho_{k+1} \left( P_{\theta_k} g_{\theta_k}(X_k) - P_{\theta_{k-1}} g_{\theta_{k-1}}(X_k) \right) I\{\sigma(K) \land \nu(\epsilon) \geq k + 1\},
\]

\[
T_{n}^{(3)} = \sum_{k=1}^{n-1} (\rho_{k+1} - \rho_{k}) P_{\theta_{k-1}} g_{\theta_{k-1}}(X_k) I\{\sigma(K) \land \nu(\epsilon) \geq k + 1\},
\]

\[
T_{n}^{(4)} = \rho_{n} P_{\theta_{n-1}} g_{\theta_{n-1}}(X_n) I\{\sigma(K) \land \nu(\epsilon) \geq n\},
\]

\[
T_{n}^{(5)} = -\sum_{k=1}^{n-1} \rho_{k} P_{\theta_{k-1}} g_{\theta_{k-1}}(X_k) I\{\sigma(K) \land \nu(\epsilon) = k\}.
\]

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

\[
\sup_{\theta \in K} \mathbb{E}_{\theta}^{\rho} \left\{ \sup_{n \geq 1} |T_{n}^{(1)}|^{p} \right\} \leq C \left( \sum_{k=1}^{\infty} \rho_k^p \right)^{p/2} \sup_{\theta \in K} \mathbb{E}_{\theta}^{\rho} \left[ W_{p}(X_k) I\{\sigma(K) \land \nu(\epsilon) \geq k\} \right],
\]

\[
\sup_{\theta \in K} \mathbb{E}_{\theta}^{\rho} \left\{ \sup_{n \geq 1} |T_{n}^{(2)}|^{p} \right\} \leq C \left( \sum_{k=1}^{\infty} \rho_k^p \right)^{p} \sup_{\theta \in K} \mathbb{E}_{\theta}^{\rho} \left[ W_{p}(X_k) I\{\sigma(K) \land \nu(\epsilon) \geq k\} \right],
\]

\[
\sup_{\theta \in K} \mathbb{E}_{\theta}^{\rho} \left\{ \sup_{n \geq 1} |T_{n}^{(3)}|^{p} \right\} \leq C \rho_{n}^{p} \sup_{\theta \in K} \mathbb{E}_{\theta}^{\rho} \left[ W_{p}(X_k) I\{\sigma(K) \land \nu(\epsilon) \geq n\} \right],
\]

\[
\sup_{\theta \in K} \mathbb{E}_{\theta}^{\rho} \left\{ \sup_{n \geq 1} |T_{n}^{(4)}|^{p} \right\} \leq C \left( \sum_{k=1}^{\infty} \rho_k^p \right)^{p/2} \sup_{\theta \in K} \mathbb{E}_{\theta}^{\rho} \left[ W_{p}(X_k) I\{\sigma(K) \land \nu(\epsilon) \geq k\} \right],
\]

where \( \mathbb{E}_{\theta}^{\rho} \) is the expectation associated to the probability \( \mathbb{P}_{\theta}^{\rho} \). The proof of these inequalities can be adapted from Lemma 4-6 of BMP - see also [2], Chapter 6, Lemma 6.2-6.4. This is deferred to Appendix A. Since \( T_{n}^{(i)} I\{\sigma(K) \land \nu(\epsilon) \geq n\} = 0 \), we have

\[
S_{n}(\epsilon, \rho, K) = T_{n} I\{\sigma(K) \land \nu(\epsilon) \geq n\} = \sum_{i=1}^{4} T_{n}^{(i)} I\{\sigma(K) \land \nu(\epsilon) \geq n\}.
\]

The Markov inequality and Lemma 4.3 imply that

\[
\mathbb{E}_{\theta}^{\rho} \left( \sup_{n \geq 1} |S_{n}(\epsilon, \rho, K)| \geq \delta \right) \leq C \delta^{-p} \left\{ \left( \sum_{k=1}^{\infty} \rho_k^p \right)^{p/2} + \left( \sum_{k=1}^{\infty} \rho_k^p \right)^p \right\} W_{p}(x),
\]
STABILITY OF STOCHASTIC APPROXIMATION

11

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

\[
\mathbb{P}\left( \sup_{n \geq 1} \left| S_{i+1,n}(\epsilon, \rho, \mathcal{K}) \right| \geq \delta \right) = \mathbb{P}\left( \mathbb{P}_{X, \theta}^{\epsilon_{l-1}} \left( \sup_{n \geq 1} \left| S_{i,n}(\epsilon^{l-1}, \rho^{l-1}, \mathcal{K}) \right| \geq \delta \right) \mathbb{I}(\sigma(\mathcal{K}) \land \nu(\epsilon) \geq l) \right)
\]

\[
\leq \mathbb{P}\left( \mathbb{P}_{X, \theta}^{\epsilon_{l-1}} \left( \sup_{n \geq 1} \left| S_{i,n}(\epsilon^{l-1}, \rho^{l-1}, \mathcal{K}) \right| \geq \delta \right) \mathbb{I}(\sigma(\mathcal{K}) \land \nu(\epsilon) \geq l) \right).
\]

Since the sequence \( \epsilon \) is non-increasing, there exists an integer \( s \) such that for all \( s \) and \( k \geq s \), \( \epsilon_{k-1}^{l-1} \leq \epsilon \), for all \( k \geq s \) (where \( \epsilon \) is defined in (A2)-iii) and Lemma 4.3 shows that there exists a constant \( C \) such that for any \( l \), for any \( x \in \mathcal{X} \), and any monotone non-increasing sequence \( \rho \),

\[
\sup_{\theta \in \mathcal{K}} \mathbb{E}_{X, \theta}^{\epsilon_{l-1}}[W^p(X_k \mathbb{I}(\sigma(\mathcal{K}) \land \nu(\epsilon) \geq k)] \leq CW^p(x).
\]

The proof follows from (4.14).

It remains to bound \( \mathbb{P}_{X, \theta}^{\epsilon_{l-1}}(\nu(\epsilon) \geq \sigma(\mathcal{K})) \).

\[
\mathbb{P}_{X, \theta}^{\epsilon_{l-1}}(\nu(\epsilon) \geq \sigma(\mathcal{K})) = \sum_{k=1}^{\infty} \mathbb{P}_{X, \theta}^{\epsilon_{l-1}}(\nu(\epsilon) = k, \sigma(\mathcal{K}) \geq k)
\]

\[
= \sum_{k=1}^{\infty} \mathbb{P}_{X, \theta}^{\epsilon_{l-1}}(\left| H(\theta_{k-1}, X_k) \right| \geq \epsilon_k \rho_k^{-1}, \sigma(\mathcal{K}) \geq k, \nu(\epsilon) = k)
\]

\[
\leq C \sum_{k=1}^{\infty} (\epsilon_k^{-1} \rho_k)^p \sup_{\theta \in \mathcal{K}} \mathbb{E}_{X, \theta}^{\epsilon_{l-1}}[W^p(X_k \mathbb{I}(\sigma(\mathcal{K}) \land \nu(\epsilon) \geq k)]
\]

The proof follows from Lemma 4.3. \( \square \)

We finally need a condition on the stepsize sequences,

(A3) The sequences \( \{\gamma_k, k \geq 0\} \) and \( \{\epsilon_k, k \geq 0\} \) are non-increasing, positive, \( \lim_{k \to \infty} \epsilon_k = 0 \) and

(i) \( \sum_{k=1}^{\infty} \gamma_k^2 < \infty \), \( \sum_{k=1}^{\infty} \gamma_k \epsilon_k^2 < \infty \) and \( \sum_{k=1}^{\infty} (\epsilon_k^{-1} \gamma_k)^p < \infty \),

(ii) \( \sum_{k=1}^{\infty} \gamma_k = \infty \),

where \( p \) and \( \alpha \) are defined in (A2).

For instance, assume that the sequence \( \gamma = (\gamma_k, k \geq 0) \) verifies \( \sum_{k=0}^{\infty} \gamma_k = \infty \) and \( \sum_{k=0}^{\infty} \gamma_k^p < \infty \) with \( 1 < p(1 + \alpha)/(p + \alpha) \). Then, (A3) is verified by setting \( \epsilon_k = C \gamma_k^\alpha \) for some constant \( C \) and some \( \eta \) such that

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

We have

Proposition 4.5. Assume (A2) and (A3)-(i). Then, for any subset \( \mathcal{K} \subset \mathcal{X} \) such that \( \sup_{x \in \mathcal{K}} W(x) < \infty \) and for any \( \delta > 0 \), we have for any fixed \( l \)

\[
\lim_{k \to \infty} \sup_{x \in \mathcal{K}} \left( A_l(\delta \epsilon^{-k}, \gamma^{-k}, x) + B(\epsilon^{-k}, \gamma^{-k}, x) \right) = 0,
\]
and for any fixed $k$, 
\[
\lim_{t \to \infty} \sup_{x \in K} \left( A_t(\delta, \epsilon^{-k}, \gamma^{-k}, x) + B(\delta, \epsilon, \gamma, x) \right) = 0,
\]
where $A_t(\delta, \epsilon, \gamma, x)$ and $B(\delta, \epsilon, \gamma, x)$ are given by (4.4) and (4.5), respectively. We may now summarize the discussion above to obtain the following stability result.

**Theorem 4.6.** Assume (A1) to (A3). Then, for any subset $K \subset X$ such that $\sup_{x \in K} W(x) < \infty$, we have
\[
\lim_{k \to \infty} \sup_{x \in K} k^{-1} \log \left( \sup_{n} \mathbb{P}_{x, \theta}(\sup_{n} \kappa_n \geq k) \right) = -\infty.
\]

### 4.1. A general convergence result.

By strengthening (A1), it is possible to formulate w.p. 1 convergence results for the sequence $(\theta_k, k \geq 0)$. For simplicity, we assume here the existence of a global Lyapunov function. More general results can be obtained under weaker assumptions (see e.g. [3] or [21]), but we do not pursue this here. To obtain such a result, it is required to strengthen Assumption (A1).

For a function $w$ verifying (A1) we consider the following additional conditions

**(A4)** For all $\theta \in \Theta$, $\left( \nabla w(\theta), h(\theta) \right) \leq 0$ and $\text{int}(w(\mathcal{L})) = \emptyset$, where $\mathcal{L}$ is defined in (3.1).

The latter condition $\text{int}(w(\mathcal{L})) = \emptyset$ has been introduced in [12]. When $w = \nabla J$, this property follows from Sard’s Theorem, under easily verified smoothness conditions for $J$.

W.p.1 convergence follows from the following result on deterministic sequences, which is adapted from [12], Theorem 2. For $A$ a subset of $\mathbb{R}^d$, define $d(x, A) = \inf \{ y \in A | x - y \}$.

**Proposition 4.7.** Assume (A1) and (A4). Let $(\rho_k, k \geq 0)$ be a monotone non-increasing sequence such that $\sum_{k=1}^{\infty} \rho_k = \infty$, let $(\xi_k, k \geq 0)$ be a sequence of vectors in $\mathbb{R}^d$ and let $K$ be a compact subset of $\Theta$. Assume that $\lim_{t \to \infty} \sup_{1 \leq i \leq j} \left| \sum_{k=1}^{j} \rho_k \xi_k \right| = 0$ and that for any $k \geq 0, \theta_k \in K$, where
\[
\theta_k = \theta_{k-1} + \rho_k h(\theta_{k-1}) + \rho_k \xi_k, \quad k \geq 1.
\]

Then $\lim \sup_{k \to \infty} d(\theta_k, \mathcal{L}) = 0$.

Combining this result with Corollary 4.2, it is possible to obtain the following global convergence result.

**Theorem 4.8.** Assume (A1) to (A4). Let $K \subset X$ such that $\sup_{x \in K} W(x) < \infty$ and let $(\theta_k, k \geq 0)$ be the sequence defined in Algorithm 1. Then, for all $x_0 \in K$ and $\theta_0 \in K_0$, we have $\lim_{k \to \infty} d(\theta_k, \mathcal{L}) = 0$, $\mathbb{P}_{x_0, \theta_0} - a.e$, where $\mathbb{P}_{x_0, \theta_0}$ is defined in (2.2).

**Proof.** Define, for $k \geq 1$,
\[
Z_k = \lim_{t \to \infty} \sup_{T_k \in T_k^{-1}} \left| \sum_{j=u}^{v} \gamma_{j+1} (H(\theta_j, X_{j+1}) - h(\theta_j)) I(v < T_k) \right| I(T_k < \infty),
\]
where $\gamma_j$ and $T_k$ are defined in Algorithm 1 and (2.3), respectively. We first show that, for any $k$ and any $\delta > 0$, $\sup_{(x, \theta_0) \in K_0} \mathbb{P}_{x, \theta_0}(|Z_k| \geq \delta) = 0$. We have, by the strong Markov property and the definition
(4.1) that

\[
P_{\varepsilon_0, \delta_0}(|Z_k| \geq \delta) 
\leq 2 \lim_{t \to \infty} P_{\varepsilon_0, \delta_0} \left( \sup_{0 \leq t_{k+1} \leq n} \frac{1}{n} \sum_{j=1}^{n} \gamma_{c_j}(H(\theta_j, X_{j+1}) - h(\theta_j)) I(n < T_k) I(T_{k-1} < \infty) \geq \delta/2 \right) 
\]

\[
\leq 2 \lim_{t \to \infty} P_{\varepsilon_0, \delta_0} \left( \sup_{0 \leq t_{k+1} \leq n} \left| S_{t/\kappa}(e^{-\gamma \epsilon T_k}, \gamma^{e^{-\gamma \epsilon T_k}}, K_{k-1}) \right| \geq \delta/2 \right) I(T_{k-1} < \infty) \right). 
\]

By Proposition 4.4, there exists a constant \( C \) such that

\[
P_{\varepsilon_0, \delta_0} \left( \sup_{0 \leq t_{k+1} \leq n} |S_{t/\kappa}(e^{-\gamma \epsilon T_k}, \gamma^{e^{-\gamma \epsilon T_k}}, K_{k-1})| \geq \delta/2 \right) I(T_{k-1} < \infty) 
\]

\[
\leq C \delta^{-p} \left\{ \left( \sum_{k=1}^{\infty} \gamma_k^{2p} \right)^{p/2} + \left( \sum_{k=1}^{\infty} \gamma_k \epsilon_k^{2p} \right)^{p} \right\} I(T_{k-1} < \infty). \tag{4.26} 
\]

Thus \( P_{\varepsilon_0, \delta_0}(|Z_k| \geq \delta) = 0 \). Corollaries 4.2 and 4.5 show that, for all \((x_0, \theta_0) \in K \times K_0, k = sup_k \kappa_k < \infty P_{x, \theta} \text{ a.e. } S_{k} = \theta_k = \theta^{e^{-\gamma \epsilon T_k}}, \gamma_k = \gamma^{e^{-\gamma \epsilon T_k}} \) and

\[
\xi_k = H(\theta_k^{e^{-\gamma \epsilon T_k}}, X_k^{e^{-\gamma \epsilon T_k}}) - h(\theta_k^{e^{-\gamma \epsilon T_k}}), \quad k \geq 1. 
\]

Then, \( \tilde{\theta}_k = \theta_{k-1} + \gamma_k h(\theta_{k-1}) + \kappa_k \xi_k \) and, since \( T_k = \infty \), for all \((x, \theta) \in K \times K_0, k \geq 1, \theta_k = \theta_{k-1} + \gamma_k h(\theta_{k-1}) + \kappa_k \xi_k \), and a.e.

The proof follows from Proposition 4.7. \[ \square \]

5. Alternative and simpler conditions. We present in this section two sets of simple conditions that imply (A2). The first set of conditions (DRP1-3) is particularly adapted to the case where the Markov chain is irreducible and aperiodic for any \( \theta \in \Theta \). This is the case for example in the context of coupled optimisation and simulation, where the Markovian dynamic is induced by the imputation of missing data using Markov chain Monte Carlo. The second set of conditions, (BMP1-4) is suited to the case where the Markov chain is naturally expressed as a random iterative system.

5.1. Drift conditions. We preface this section by recalling some results on \( \psi \)-irreducible and aperiodic Markov chain (see [24] for an extensive treatment). Let \( P \) be a transition kernel on \( X \) and \( \Phi = (\Phi_n, n \in \mathbb{N}) \) a Markov chain on \( X \) with transition \( P \).

Recall that a Markov chain on a state space \( X \) is said to be \( \psi \)-irreducible if there exists a measure \( \psi \) on \( B(X) \) such that, whenever \( \psi(A) > 0 \), \( P_{\tau_A} < \infty \) \( \tau_A \) is the first return time to \( A \), \( \tau_A = \inf \{ n \geq 1, \Phi_n \in A \} \). Denote by \( \psi \) a maximal irreducibility measure for \( P \) (see [24] Chapter 4 for the definition and the construction of such measure). If \( P \) is \( \psi \)-irreducible, aperiodic and has an invariant probability measure \( \pi \), it is well known that \( \pi \) is a maximal irreducibility measure. A chain is said to be \( \psi \) uniformly ergodic ([24], Chapter 16) if there exists constants \( 0 < \rho < 1, C \) and a function \( V : X \to [1, \infty) \), such that

\[
\|P^n(x, - \pi(\cdot))\|_V \leq C\rho^n. 
\]
When the chain is \( \psi \)-irreducible and aperiodic, \( V \)-uniform ergodicity can be established under drift and minorization conditions. A subset \( C \subset \mathcal{X} \) is said to be an \((m, \delta)\)-small set if there exist a probability measure \( \nu \) on \( \mathcal{X} \), a positive integer \( m \) and \( \delta > 0 \) such that
\[
P^m(x, A) \geq \delta \nu(A), \quad \forall x \in C, \quad \forall A \in \mathcal{B}(X).
\] (5.1)

A function \( V \geq 1 \) is said to be a drift function outside \( C \) if there exists a constants \( \lambda < 1, b \) and a function \( V \geq 1 \) such that
\[
P V \leq \lambda V + b I_C.
\]

Denote \( \mathcal{L}_V = \{ g : \mathcal{X} \to \mathbb{R}, \sup_{x \in \mathcal{X}} ||g||_V < \infty \} \).

(DRI) For any \( \theta \in \Theta \), \( P_\theta \) is irreducible and aperiodic. In addition there exist a function \( V : \mathcal{X} \to [1, \infty) \), constants \( p \) and \( \beta, p \geq 2 \) and \( 0 \leq \beta \leq 1 \), such that, for any compact subset \( K \subset \Theta \),
(DRI1) there exist an integer \( m \), constants \( 0 < \lambda < 1, b, \kappa, \delta > 0 \) and a probability measure \( \nu \) such that
\[
\sup_{\theta \in K} P^m_\theta V^p(x) \leq \lambda V^p(x) + b I_C, \quad \forall x \in \mathcal{X}, \tag{5.2}
\]
\[
\sup_{\theta \in K} P^m_\theta V^p(x) \leq \kappa V^p(x) \quad \forall x \in \mathcal{X}, \tag{5.3}
\]
\[
\inf_{\theta \in K} P^m_\theta (x, A) \geq \delta \nu(A) \quad \forall x \in C, \quad \forall A \in \mathcal{B}(X). \tag{5.4}
\]

(DRI2) there exists \( C \) such that, for all \( x \in \mathcal{X} \),
\[
\sup_{\theta \in K} ||H_\theta(x)|| \leq CV(x), \tag{5.5}
\]
\[
\sup_{(\theta, \theta') \in K} ||\theta - \theta'||^{\beta} ||H_\theta(x) - H_{\theta'}(x)|| \leq CV(x). \tag{5.6}
\]

(DRI3) there exists \( C \) such that, for all \( (\theta, \theta') \in K \times K \),
\[
||P_\theta g - P_{\theta'} g||_V \leq C ||g||_V ||\theta - \theta'||^\beta \quad \forall g \in \mathcal{L}_V, \tag{5.7}
\]
\[
||P_\theta g - P_{\theta'} g||_V \leq C ||g||_V ||\theta - \theta'||^\beta \quad \forall g \in \mathcal{L}_V. \tag{5.8}
\]

Proposition 5.1. Assume (DRI). Then, for any \( \theta \in \Theta \), \( P_\theta \) admits a single stationary distribution \( \pi_\theta \), verifying \( \pi_\theta(V^p) < \infty \). The Poisson equation \( u - P_\theta u = H_\theta - h_\theta \) has a solution \( g_\theta \) verifying,
\[
\sup_{\theta \in K} (||g_\theta||_V + ||P_\theta g_\theta||_V) < \infty. \tag{5.9}
\]

For any \( 0 < \alpha < \beta \),
\[
\sup_{(\theta, \theta') \in K \times K} ||\theta - \theta'||^{-\alpha} |h(\theta) - h(\theta')| < \infty, \tag{5.10}
\]
\[
\sup_{(\theta, \theta') \in K \times K} ||\theta - \theta'||^{-\alpha} \{|g_\theta - g_{\theta'}||_V + ||P_\theta g_\theta - P_{\theta'} g_{\theta'}||_V\} < \infty. \tag{5.11}
\]

In addition, there exist constants \( C \) and \( \epsilon > 0 \) such that, for every non increasing sequence \( \rho = (\rho_k, k \geq 0) \), and any \( x \in \mathcal{X} \),
\[
\sup_{\theta \in K} E^\rho_{x, \theta}[V^p(X_k)I(\sigma(K) \land \nu_k \geq k)] \leq CV^p(x). \tag{5.12}
\]

The proof is in Appendix B.
5.2. Lipshitz conditions. In [4] a set of sufficient conditions are given, allowing to obtain computable bounds for kernels which are contractive in some weighted Lipshitz space. Here, the state space \(X = \mathbb{R}^r\) which is equipped with its Borel \(\sigma\)-field. For any \(q \geq 0\) and \(g : \mathbb{R}^r \rightarrow \mathbb{R}^d\) we define the following norm
\[
[g]_{q} = \sup_{x \neq y \in \mathbb{R}^r} \frac{|g(x) - g(y)|}{|x - y|(1 + |x|^q + |y|^q)}.
\]
and the associated space \(\text{Li}(q) = \{g : \mathbb{R}^r \rightarrow \mathbb{R}^d, [g]_{q} < \infty\}\). We also define, for any \(q \geq 0\),
\[
V_q(x) = (1 + |x|^q), \quad \text{and} \quad \|g\|_q = \|g\|_{V_q} = \sup_{x \in \mathbb{R}^r} |g(x)|/(1 + |x|^q).
\]
(5.12)
Note that \([g]_{q} < \infty\) implies that \(\|g\|_{q+1} < \infty\).
(BMP) There exist constants \(q, p, q \geq 0, p \geq 2\) such that, for any compact subset \(K \subset \Theta\),

(BMP1) there exist an integer \(m \geq 1\), constants \(0 \leq \lambda < 1\) and \(b\) such that
\[
\sup_{\theta \in K} P_{\theta}^m V_{q+1}^p \leq \lambda V_{q+1}^p + b, \quad \text{and} \quad \sup_{\theta \in K} \|P_{\theta} V_{q+1}^p\|_{p(q+1)} < \infty.
\]
(5.13)
(BMP2)
\[
\left( \sup_{\theta \in K} (|H_{\theta}|)^{q} \right) \left( \sup_{\theta \in K} \|H_{\theta}\|_{q+1} \right) < \infty.
\]
(5.14)
\[
\sup_{(\theta, \theta') \in K \times K} \{ |\theta - \theta'|^{-\beta} \|H_{\theta} - H_{\theta'}\|_{q+1} \} < \infty.
\]
(5.15)
(BMP3) there exist an integer \(m \geq 1\) and constants \(C\) and \(\rho, 0 < \rho < 1\) such that, for \(r = m(q+1) - 1\),
\[
\sup_{\theta \in K} |P_{\theta} g|_q \leq C|g|_q, \quad \text{and} \quad \sup_{\theta \in K} |P_{\theta}^m g|_q \leq \rho |g|_q, \quad \forall g \in \text{Li}(q)
\]
(5.16)
\[
\sup_{\theta \in K} |P_{\theta} g|_r \leq C|g|_r, \quad \text{and} \quad \sup_{\theta \in K} |P_{\theta}^m g|_r \leq \rho |g|_r, \quad \forall g \in \text{Li}(r).
\]
(5.17)
(BMP4) there exists a constant \(C\) such that
\[
|P_{\theta} g - P_{\theta'} g|_{q+1} \leq C|g|_q |\theta - \theta'|^{\beta} \quad \forall g \in \text{Li}(q),
\]
(5.18)
\[
|P_{\theta} g - P_{\theta'} g|_{r+1} \leq C|g|_r |\theta - \theta'|^{\beta} \quad \forall g \in \text{Li}(r).
\]
(5.19)

PROPOSITION 5.2. Assume (BMP). Then, for any \(\theta \in \Theta\), \(P_\theta\) has a unique stationary distribution \(\pi_\theta, \pi_\theta(|H_\theta|) < \infty\) and for any compact subset \(K \subset \Theta\),
\[
\sup_{(\theta, \theta') \in K \times K} |\theta - \theta'|^{-\alpha} |h(\theta) - h(\theta')| < \infty,
\]
(5.20)
with \(h_\theta = \pi_\theta(H_\theta)\). The Poisson equation \(g - P_\theta g = H_\theta - h_\theta\) has a solution \(g_\theta\) and
\[
\sup_{(\theta, \theta') \in K \times K} |\theta - \theta'|^{-\alpha} \{ ||g_0 - g_\theta||_{q+1} + ||P_\theta g_0 - P_{\theta'} g_\theta||_{q+1} \} < \infty.
\]
(5.21)
In addition, there exist a constant \(C\) and \(\epsilon > 0\) such that, for every non increasing sequence \(\rho = (\rho_k, k \geq 0)\), and any \(x \in X\),
\[
\sup_{\theta \in K} \mathbb{E}^\rho_{x, \theta}[V^p(X_k)I(\sigma(K) \land \nu_s \geq k)] \leq CV^p(x).
\]
(5.22)
The proof is in Appendix C.
6. Recursive prediction error method for ARMA estimation. We restrict our attention to the ARMA$(1,1)$ case in order to avoid unnecessary notational difficulties. Let $(Z_k, k \in \mathbb{Z})$ be a real valued stationary ARMA$(1,1)$ sequence

$$Z_k - \alpha Z_{k-1} = U_k + \beta V_{k-1}, \quad |\alpha| < 1, |\beta| < 1, \alpha + \beta \neq 0,$$

where $(V_k, k \in \mathbb{Z})$ is an i.i.d. zero mean sequence such that, for some $p \geq 1$, $E|V|^p < \infty$.

The recursive prediction error algorithm (see [22]) consists of a stochastic approximation procedure for minimizing the (stationary) innovation variance $W(\alpha, \beta) = E_{\alpha, \beta}[V_k^2(\alpha, \beta)]$, where $V_k(\alpha, \beta)$ is recursively defined as

$$V_k(\alpha, \beta) = Z_k - \alpha Z_{k-1} - \beta V_{k-1}(\alpha, \beta). \quad (6.1)$$

The partial derivatives of $V_k(\alpha, \beta)$ w.r.t $\alpha$, $\beta$ and the gradient of the innovation variance are respectively given by

$$\partial_\alpha V_k(\alpha, \beta) = -Z_{k-1} - \beta \partial_\alpha V_{k-1}(\alpha, \beta),$$
$$\partial_\beta V_k(\alpha, \beta) = -V_{k-1}(\alpha, \beta) - \beta \partial_\beta V_{k-1}(\alpha, \beta),$$
$$\nabla W(\alpha, \beta) = E_{\alpha, \beta}[V_k(\alpha, \beta)(\partial_\alpha V_k(\alpha, \beta) \partial_\beta V_k(\alpha, \beta))^T].$$

This suggests the following stochastic approximation procedure

$$\hat{V}_k = Z_k - \alpha_{k-1} Z_{k-1} - \beta_{k-1} \hat{V}_{k-1},$$
$$\partial_\alpha \hat{V}_k = -Z_{k-1} - \beta_{k-1} \partial_\alpha \hat{V}_{k-1},$$
$$\partial_\beta \hat{V}_k = -\hat{V}_{k-1} - \beta_{k-1} \partial_\beta \hat{V}_{k-1},$$
$$\alpha_k = \alpha_{k-1} - \gamma_k \hat{V}_k \partial_\alpha \hat{V}_k,$$
$$\beta_k = \beta_{k-1} - \gamma_k \hat{V}_k \partial_\beta \hat{V}_k,$$

which can be cast into the framework outlined in (1.3) above by denoting

- $\theta_k = (\alpha_k, \beta_k)$ the values of the ARMA parameters at iteration $k$, assumed to belong to $\Theta = (-1, +1)^2$. These constraints naturally correspond to causal and invertible ARMA processes,
- $X_k = [V_k, Z_k, \hat{V}_k, \partial_\alpha \hat{V}_k, \partial_\beta \hat{V}_k]^T \in X = \mathbb{R}^6$ and $H(\theta, x) = H(x) = -(x_2 x_4, x_4 x_2)^T$, where $x = (x_1, x_2, x_3, x_4, x_5)$. It follows from the definitions above that $(X, k \geq 0)$ is a controlled Markov chain whose transition probability at iteration $k$ depends on the current value of the parameter $\theta_k$ and is defined by the following linear state space,

$$X_{k+1} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ \beta & \tilde{\alpha} & 0 & 0 & 0 \\ \beta & (\tilde{\alpha} - \alpha_k) & -\beta_k & 0 & 0 \\ 0 & -1 & 0 & -\beta_k & 0 \\ 0 & 0 & -1 & 0 & -\beta_k \\ \end{bmatrix} X_k + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ \end{bmatrix} V_{k+1}, \quad (6.2)$$

where $(\tilde{\alpha}, \tilde{\beta})$ is the true value of the parameter.

We now check the required Assumption (A1) that guarantees that Algorithm 1 is stable.
Cond. (A1). This algorithm is a recursive approximation of the minimum prediction error and the natural Lyapunov function is here the stationary innovation variance \( W(\alpha, \beta) = \mathbb{E}_\alpha, \beta [V_k(\alpha, \beta)^2] \), which can be alternatively expressed as

\[
W(\alpha, \beta) = \frac{\sigma^2}{2\pi} \int_{-\pi}^{\pi} \frac{1 - \alpha e^{i\lambda} \beta}{1 + \beta e^{i\lambda}} \frac{1 + \beta e^{i\lambda}}{|1 - \alpha e^{i\lambda}|^2} d\lambda,
\]

The moduli of the eigenvalues of \( F(\theta) = F(\alpha, \beta) \) are less than \( |\alpha| \vee |\beta| \). Therefore, for any compact set \( \mathcal{K} \subset \Theta \), there exist constants \( K \) and \( \rho < 1 \) such that for any \( k \geq 1 \),

\[
\sup_{\theta \in \mathcal{K}} |F^k(\theta)| \leq K \rho^k. \tag{6.3}
\]

Hence, for any \( \beta, |\beta| < 1 \), the linear control model is stable, and thus admits a single stationary distribution \( \pi_{\alpha, \beta} \). The mean field is given by \( h(\alpha, \beta) = \frac{\partial W(\alpha, \beta)}{\partial \alpha, \partial W(\alpha, \beta)} \partial \beta \). It is easily seen that, when \( \alpha \neq -\beta, \theta \) is the unique solution of the equation \( \nabla W(\theta) = 0 \) on the set \( \Theta \). We now point out a couple of properties of \( F \) useful to check the (BMP) assumptions. First, \( \theta \mapsto F(\theta) \) is a Lipschitz function on \( \Theta \), i.e. there exists a constant \( \hat{K} \) such that,

\[
|F(\theta) - F(\theta')| \leq \hat{K}|\theta - \theta'|, \quad \forall (\theta, \theta') \in \Theta \times \Theta. \tag{6.4}
\]

Now if we define for \( k = 0, 1, \ldots \)

\[
X_{k+1}^\theta = F(\theta) + G \mathcal{E}_{k+1} \quad \text{and} \quad X_0^\theta = x,
\]

\[
F^k(\theta)x + U^{\theta}_{k+1} \quad \text{with} \quad U^\theta_{k+1} = \sum_{j=0}^{k} F^j(\theta) G \mathcal{E}_{k+1-j},
\]

then Minkowski inequality and Eq. (6.3) imply that

\[
\sup_{\theta \in \mathcal{K}} \mathbb{E} \left[ |U^\theta_k|^{p+1} \right] \leq \sup_{\theta \in \mathcal{K}} \left( \sum_{j=0}^{\infty} |F^j(\theta)| ||G|| ||V||^{p+1} \right) < \infty. \tag{6.5}
\]

Cond. (BMP1). From Eq. (6.5) there exists a constant \( b \) such that for any \( m \geq 1 \)

\[
\int_X F^m_\theta (x, dy) |y|^{p+1} = \mathbb{E} \left[ |F^m_\theta x + U^\theta_m|^{p+1} \right] \leq 2^p |F^m_\theta|^{p+1} |x|^{p+1} + b,
\]

and we conclude by setting \( m \) in such a way that

\[
\sup_{\theta \in \mathcal{K}} |F^m_\theta| \leq K \rho^m < 1/2.
\]

Cond. (BMP2). Follows from the fact that \( H \) does not depend upon parameter \( \theta \) and that \( H(x) \) is Lipschitz function.

Cond. (BMP3). For any \( 1 \leq q \leq (p+1), g \in L^q(q), k = 1, 2, \ldots \) and \( (x, y) \in X \times X \),

\[
|P^k_\theta g(x) - P^k_\theta g(y)| = |\mathbb{E} \left[ g(F^k_\theta(x) + U^\theta_k) - g(F^k_\theta(y) + U^\theta_k) \right]|,
\]

\[
\leq |g|^{q} |F^k_\theta(x - y)| (1 + \mathbb{E} \left[ |F^k_\theta(x) + U^\theta_k|^{q} \right] + \mathbb{E} \left[ |F^k_\theta(y) + U^\theta_k|^{q} \right]),
\]

and we conclude using the Minkowski inequality, (6.3) and (6.5).
Fig. 6.1. The top plot corresponds to $\hat{\theta}$. The bottom plot corresponds to $\hat{\beta}$. The vertical lines display the time instants when $m$-initializations occur. The horizontal dashed lines correspond to the successive re-projection boundaries and the horizontal dotted lines correspond to the true values of the parameters.

Cond. (BMP4). For any $1 \leq q \leq p$ there exists a constant $C_q$ (which depends only on $q$) such that for any $g \in \text{Li}(q)$,

$$
|\mathbb{E}[g(F(\theta)x + GV_0)] - \mathbb{E}[g(F(\theta')x + GV_0)]| \leq C_q|\theta - \theta'| |x| \times (1 + |F(\theta)|^q |x|^q + |F(\theta')|^q |x|^q + \mathbb{E}[|V_0|^q]).
$$

(6.6)

Consequently, from (6.4), (6.3) and $||V||_{p+1} < \infty$ we have that for any $1 \leq q \leq p$,

$$
|P_qg(x) - P_qg(x)| \leq C_q|\theta - \theta'| (1 + |x|^{q+1}),
$$

which shows that assumption (BMP4) is satisfied with $\beta = 1$.

Here we consider an illustrative example. We consider an ARMA$(1, 1)$ process with a root close to the unit circle. We chose to simply reinitialize the algorithm at a predetermined value of the parameters. Results of a computer simulation are presented on Fig. 6. The series of compact sets where chosen to be of the form $K_q = [-1 + .5/(q + 1), 1 - .5/(q + 1)]$.

7. Controlled MCMC algorithm. Here is a more substantial example, drawn from [16]. Hastings-Metropolis algorithms ([23],[18]) are a class of algorithms whose aim is to produce samples distributed according to predefined probability distributions, say $\pi$, which are only known up to a scale factor. Here we will assume that $\pi$ is defined on $\mathbb{R}^d$. This technique is especially relevant when $\pi$ is the posterior distribution in a Bayesian context, but also finds many applications in computational physics. The general mechanism of these algorithms is as follows. First consider a candidate transition kernel $Q(x, A)$,
$x \in \mathbb{R}^n$, $A \in \mathcal{B}(\mathbb{R}^n)$, which generates potential transitions for a Markov chain on $\mathbb{R}^n$ equipped with the Borel $\sigma$-algebra $\mathcal{B}(\mathbb{R}^n)$. We will assume that $Q(x, \cdot)$ is absolutely continuous with density $q(x, y)$ with respect to the Lebesgue measure. A candidate transition generated according to the law of $Q$ is then accepted with probability $\alpha(x, y)$ given by

$$\alpha(x, y) = \begin{cases} 1 & \text{if } \pi(x)q(x, y) > 0 \\ 1/2 & \text{if } \pi(x)q(x, y) = 0. \end{cases} \quad (7.1)$$

The class of the Hastings-Metropolis algorithms is very broad. We consider in this section perhaps the simplest version of this algorithm, namely the symmetric random-walk Hastings-Metropolis algorithm or simply Metropolis algorithm. In this case $q(x, y) = q(|x - y|)$ for some density $q$ on $\mathbb{R}^+$. The transition kernel of the Metropolis algorithm is then given by

$$P_q(x, A) = \int_A \alpha(x, x + z)q(|z|) \, dz + I_A(x) \int \left(1 - \alpha(x, x + z)\right)q(|z|) \, dz. \quad (7.2)$$

The idea behind controlled MCMC is to tune the transition kernel $q$ in the light of the past simulations. Such a scheme has been recently proposed by [6] and, as pointed out in [1], this is a particular case of a more general methodology that can be cast in the stochastic approximation framework developed in this paper. Instead of using a fixed proposal, the algorithm uses a parametric family of probability densities, here a $d$-dimensional zero mean normal density with covariance matrix $\Gamma$ belonging to the cone $C^+_d$ of positive definite matrices on $\mathbb{R}^d$,

$$q_r(z) = \frac{1}{(\det 2\pi \Gamma)^{1/2}} \exp \left(-\frac{1}{2} z^\top \Gamma^{-1} z\right).$$

The covariance matrix $\Gamma$ is adapted according to the following stochastic approximation procedure

$$\mu_k = \mu_{k-1} + \gamma_k (X_k - \mu_k) \quad k \geq 1 \quad (7.3)$$

$$\Gamma_k = \Gamma_{k-1} + \gamma_k ((X_k - \mu_{k-1})(X_k - \mu_{k-1})^\top - \Gamma_{k-1}) \quad (7.4)$$

where $X_k$ is sampled from $P_q(X_{k-1}, \cdot)$ ($P_q$ is here a shorthand notation for the random-walk Metropolis kernel $P_{\pi_q}$ (7.2)). To make the connection with the derivations above, denote $\theta = (\mu, \Gamma) \in \Theta$ where $\Theta = \mathbb{R}^d \times C^+_d$. This stochastic approximation procedure is associated to the minimisation of the Kullback-Leibler divergence between the target density $\pi$ and the family $(q_r, \Gamma \in C^+_d)$. The aim of the procedure is thus to find the “best” approximation (in the Kullback-Leibler sense) of the target density by a multivariate Gaussian density. Assume that $\mathbb{E}_\pi |X|^2 = \int_{\mathbb{R}^d} |x|^2 \, dx < \infty$. Define, for $(\mu, \Gamma) \in \mathbb{R}^d \times C^+_d$,

$$J(\mu, \Gamma) = \log \det \Gamma + (\mu - \mu_\pi)^\top \Gamma^{-1} (\mu - \mu_\pi) + \text{Tr}(\Gamma^{-1} \Gamma_\pi),$$

where $\mu_\pi = \mathbb{E}_\pi [X]$ and $\Gamma_\pi = \mathbb{E}_\pi [(X - \mu_\pi)(X - \mu_\pi)^\top]$. $J$ is, up to an additive and a multiplicative constants, the Kullback-Leibler divergence between $\pi$ and $q_r$. It is assumed here that $\Gamma_\pi \in C^+_d$. We have for $(\mu, \Gamma) \in \mathbb{R}^d \times C^+_d$,

$$H(x; \mu, \Gamma) = (x - \mu)(x - \mu)^\top - \Gamma \quad (7.5)$$

$$h(\mu, \Gamma) = (\mu_\pi - \mu)(\mu_\pi - \mu)^\top + \Gamma_\pi - \Gamma. \quad (7.6)$$

We now check that all the conditions required to ensure convergence of this scheme are satisfied.
**Condition (A1)-(A4).** The regularity assumptions are obviously checked but we have to verify that
\[ \langle \nabla J(\mu, \Gamma), h(\mu, \Gamma) \rangle \leq 0 \text{ for any } (\mu, \Gamma) \in \Theta, \text{ with strict inequality when } (\mu, \Gamma) \neq (\mu_\pi, \Gamma_\pi). \] It follows from straightforward algebra that
\[ \langle \nabla J(\mu, \Gamma), h(\mu, \Gamma) \rangle = -2(\mu - \mu_\pi) \Gamma^{-1}(\mu - \mu_\pi) - \text{Tr}(\Gamma^{-1}(\Gamma - \Gamma_\pi)(\Gamma - \Gamma_\pi)) - ((\mu - \mu_\pi) \Gamma^{-1}(\mu - \mu_\pi))^2 \]
which implies (A1)-(A4).

**Condition (A2).** Now we have to verify (A2). In this example, it is appropriate to use the conditions outlined for \( \psi \)-irreducible and aperiodic kernels (DRI). To check (DRI) we need to introduce additional conditions on the target distribution. These conditions are not minimal but easy to check in practice (see [19]).

(ADM1) The probability distribution \( \pi \) has the following properties:
1. It is positive on every compact set and continuously differentiable.
2. It is super-exponential, i.e.
\[ \lim_{|x| \to +\infty} \frac{x}{|x|} \nabla \log \pi(x) = -\infty. \quad (7.7) \]
3. The contours \( \partial A(x) = \{ y : \pi(y) = \pi(x) \} \) are asymptotically regular, i.e.
\[ \lim_{|x| \to +\infty} \sup_{|z| < \infty} \frac{x}{|x|} \frac{\nabla \pi(x)}{|\nabla \pi(x)|} < 0. \quad (7.8) \]

**Lemma 7.1.** Assume (ADM1), let \( W_\eta(x) = c\pi^{-\eta}(x) \) with \( \eta \in (0,1) \) and where \( c \) is such that \( W_\eta \geq 1 \). Then for any compact set \( K \subset C_0^q \),
1. For any non-empty bounded set \( C \subset \mathbf{X} \) there exist \( \varepsilon(K, C) > 0 \) and a probability measure \( \nu_{K, C} \) such that for any \( A \in \mathcal{B}(\mathbf{X}) \),
\[ \inf_{\Gamma \in K} P_{\Gamma}(x, A) \geq \varepsilon(K, C) \nu_{K, C}(A) \quad (7.9) \]
2. In addition,
\[ \lim_{|x| \to +\infty} \sup_{\Gamma \in K} \frac{P_{\Gamma} W_\eta(x)}{W_\eta(x)} < 1, \]
\[ \sup_{z \in \mathbf{X} \times K} \inf_{\Gamma \in K} \frac{P_{\Gamma} W_\eta(x)}{W_\eta(x)} < +\infty. \]

**Proof.** For a compact set \( K \subset C_0^q \) we introduce \( \lambda_{\min}(K) \) and \( \lambda_{\max}(K) \) the minimum and maximum eigenvalues of matrices in \( K \) and we will need the following sets, \( A(x) = \{ z; \pi(x + z) \geq \pi(x) \} \) and \( R(x) = \{ z; \pi(x + z) < \pi(x) \} \). Let \( I \) be the identity matrix, then for any \( \Gamma \in K \),
\[ q_{\Gamma}(z) \geq \left( \frac{\lambda_{\min}(K)}{\lambda_{\max}(K)} \right)^{n/2} q_{\lambda_{\max}(K)}(I)(z). \]
Now Theorem 2.2 of [26] applies to \( q_{\lambda_{\min}(K)}(I)(z) \) and from the inequality above we conclude that (7.9) is satisfied. Again, using this inequality, we have
\[ \inf_{\Gamma \in K} \int_{A(x)} q_{\Gamma}(z)dz \geq \left( \frac{\lambda_{\min}(K)}{\lambda_{\max}(K)} \right)^{n/2} \int_{A(x)} q_{\lambda_{\min}(K)}(I)(z)dz. \]
From the conclusion of the proof of Theorem 4.3 of [19] we have
\[
\liminf_{|x| \to +\infty} \int_{A(x)} q_{\lambda_{\min}(\mathcal{K})}(z)dz > 0,
\]
but from the conclusion of the proof of Theorem 4.1 of [19],
\[
\limsup_{|x| \to +\infty} \sup_{\Gamma \in \mathcal{K}} \frac{P_{\Gamma} W_n(x)}{W_n(x)} = 1 - \liminf_{|x| \to +\infty} \inf_{\Gamma \in \mathcal{K}} \int_{A(x)} q_{\Gamma}(z)dz,
\]
and we conclude. Finally, for any \( \Gamma \in \mathcal{K} \),
\[
\frac{P_{\Gamma} W_n(x)}{W_n(x)} = \int_{A(x)} \frac{\pi(x + z) - \eta}{\pi(x) - \eta} q_{\Gamma}(z)dz + \int_{R(x)} \left( 1 - \frac{\pi(x + z)}{\pi(x)} + \frac{\pi(x + z)^{1-\eta}}{\pi(x)^{1-\eta}} \right) q_{\Gamma}(z)dz,
\]
where all the ratios are at most 1, so that this quantity is bounded independently of \( \Gamma \).

**Lemma 7.2.** Let \( \mathcal{K} \) be a compact subset of \( \Theta \) and \( \eta \in (0,1) \). For \( \Gamma, \Gamma' \in \mathcal{K} \times \mathcal{K} \), \( g \in L_W \), we have
\[
||P_{\Gamma} g - P_{\Gamma'} g||_{W_n} \leq \frac{2a}{\lambda_{\min}(\mathcal{K})} ||g||_{W_n} ||\Gamma - \Gamma'||,
\]
where \( \lambda_{\min}(\mathcal{K}) \) is the minimum possible eigenvalue for matrices in \( \mathcal{K} \).

**Proof.** Let \( \delta_{\Gamma, \Gamma'}(z) = q_{\Gamma}(z) - q_{\Gamma'}(z) \) and consider
\[
|P_{\Gamma} g(x) - P_{\Gamma'} g(x)| = \left| \int_{X} \alpha(x, x + z) \delta_{\Gamma, \Gamma'}(z) g(x + z)dz + g(x) \right| \int_{X} \alpha(x, x + z) \delta_{\Gamma, \Gamma'}(z) dz \right|.
\]
Now
\[
\left| \int_{X} \alpha(x, x + z) \delta_{\Gamma, \Gamma'}(x, x + z) g(x + z)dz \right| \leq ||g||_{W_n} \int_{X} \alpha(x, x + z) ||\delta_{\Gamma, \Gamma'}(z)|| W_n(x + z) dy
\]
and notice that we have
\[
\int_{X} \alpha(x, x + z) W_n(x + z)dz = \int_{A(x)} \frac{\pi(x + z) - \eta}{\pi(x) - \eta} \left| \delta_{\Gamma, \Gamma'}(x, x + z) \right| dz + \int_{R(x)} \frac{\pi(x + z)^{1-\eta}}{\pi(x)^{1-\eta}} \left| \delta_{\Gamma, \Gamma'}(x, x + z) \right| dz \leq \int_{X} \left| \delta_{\Gamma, \Gamma'}(x, x + z) \right| dz.
\]
Then, since for \( z \in A(x) \), \( \pi(x + z) \geq \pi(x) \) whereas for \( z \in R(x) \) \( \pi(x + z) < \pi(x) \), this implies
\[
\left| \int_{X} \delta_{\Gamma, \Gamma'}(x, x + z) g(x + z)dz \right| \leq ||g||_{W_n} \int_{X} \left| \delta_{\Gamma, \Gamma'}(z) \right| dz.
\]
Note that the RHS is the total variation distance between \( q_{\Gamma} \) and \( q_{\Gamma'} \) and that the result that we have derived above remains valid for any probability densities \( q \) and \( q' \) defined on \( \mathbb{R}^n \). Now,
\[
\int_{X} \left| \delta_{\Gamma, \Gamma'}(z) \right| dz = \int_{0}^{1} \left| \frac{\partial q_{\Gamma + v(z - \delta(z))}}{\partial v} \right| dz
\]
and let \( \Gamma_v = \Gamma + v(\Gamma' - \Gamma) \), so that
\[
\frac{\partial \log q_{\Gamma + v(\Gamma' - \Gamma)}}{\partial v} = -\frac{1}{2} \text{Tr} \left[ (\Gamma - \Gamma) + \Gamma^{-1} zz^T(\Gamma - \Gamma) \right]
\]
and consequently

$$\int_x \left| \int_0^1 \frac{\partial g_{\Gamma_{\tau + v}}(x,y)}{dv} dv \right| dz \leq |\Gamma' - \Gamma| \int_0^1 |\Gamma_{\tau + v}^{-1}| dv \leq \frac{a}{\lambda_{\min}(K)} |\Gamma' - \Gamma|,$$

where we have used the following inequality,

$$|\text{Tr}[\Gamma_{\tau + v}^{-1} z z^T \Gamma_{\tau + v}^{-1} (\Gamma' - \Gamma)]| \leq |\Gamma' - \Gamma| |\text{Tr}[\Gamma_{\tau + v}^{-1} z z^T]|.$$

\[\square\]

**Appendix A. Proof of Proposition 4.4.** Denote $D(\epsilon, \rho, K; x) = \sup_{k \geq 1} \sup_{\theta \in K} E_{x, \theta}^D [W^p(X_k) I(\sigma(K) \wedge \nu(\epsilon) \geq k + 1)]$.

**Proof.** [Proof of (4.21)] Under (A2),

$$\sup_{\theta \in K} E_{x, \theta}^D \{ ((g_{\theta_k}(X_{k+1}))^p + |P_{\theta_k} g_{\theta_k}(X_{k+1})|^p) I(\sigma(K) \wedge \nu(\epsilon) \geq k + 1) \} \leq CD(\epsilon, \rho, K; x).$$

Since

$$E_{x, \theta}^D \{ (g_{\theta_k}(X_{k+1})) - P_{\theta_k} g_{\theta_k}(X_k)) I(\sigma(K) \wedge \nu(\epsilon) \geq k + 1) \} = (P_{\theta_k} g_{\theta_k}(X_k) - P_{\theta_k} g_{\theta_k}(X_k)) I(\sigma(K) \wedge \nu(\epsilon) \geq (k + 1)) = 0,$$

$T_n^{(1)}$ is a $(\mathbb{R}^d$-valued) martingale. Using the Burkholder inequality ([17], Theorem 2.10), we have

$$E_{x, \theta}^D \left\{ \left| T_n^{(1)} \right|^p \right\} \leq C_p E_{x, \theta}^D \left( \sum_{k=0}^{n-1} \rho_{k+1}^2 |g_{\theta_k}(X_{k+1}) - P_{\theta_k} g_{\theta_k}(X_k)|^2 I(\sigma(K) \wedge \nu(\epsilon) \geq k + 1) \right)^{p/2}.$$

where $C_p$ is a universal constant. Using Minkowski's inequality and $\sup_{\theta \in K}(||g_{\theta}||_V + ||P_{\theta_k} g_{\theta_k}||_V) < \infty$, (A2), we have

$$E_{x, \theta}^D \left\{ \left| T_n^{(1)} \right|^p \right\} \leq C \left( \sum_{k=1}^{n} \rho_k^2 \right)^{p/2} D(\epsilon, \rho, K; x).$$

Since $T_n^{(1)}$ is a martingale in $L^p$, then $|T_n^{(1)}|$ is a non-negative submartingale in $L^p$ and Doob’s $L^p$ inequality implies that

$$E_{x, \theta}^D \left\{ \sup_{n \geq 1} \left| T_n^{(1)} \right|^p \right\} \leq C \left( \sum_{k=1}^{n} \rho_k^2 \right)^{p/2} D(\epsilon, \rho, K; x),$$

which concludes the proof. \[\square\]

**Proof.** [Proof of (4.22)] Under (A2), we have

$$\sup_{n \geq 1} \left| \sum_{k=1}^{n-1} \rho_{k+1} (P_{\theta_k} g_{\theta_k}(X_{k+1}) - P_{\theta_{k-1}} g_{\theta_{k-1}}(X_k)) \right| I(\sigma(K) \wedge \nu(\epsilon) \geq k + 1) \leq \sum_{k=0}^{\infty} \rho_{k+1} W(X_k) |\theta_k - \theta_{k-1}|^a I(\sigma(K) \wedge \nu(\epsilon) \geq k + 1),$$

$$\leq \sum_{k=0}^{\infty} \rho_{k+1} e_k^2 W(X_k) I(\sigma(K) \wedge \nu(\epsilon) \geq k + 1).$$
We conclude the proof by applying Minkowski's inequality. \[\]

**Proof.** [Proof of (4.23)] Under (A2),

\[
\sup_{n \geq 1} \left| \sum_{k=1}^{n} (\rho_{k+1} - \rho_{k}) P_{\theta_{k-1}} g_{\theta_{k-1}}(X_k) I(\sigma(K) \land \nu(\epsilon) \geq k + 1) \right| \leq \sum_{k=1}^{\infty} (\rho_k - \rho_{k+1}) W(X_k) I(\sigma(K) \land \nu(\epsilon) \geq k + 1),
\]

and the proof follows from Minkowski's inequality. \[\]

**Proof.** [Proof of (4.24)] Under (A2),

\[
\sup_{n \geq 1} \left| \rho_{n} P_{\theta_{0}} g_{\theta_{0}}(X_0) I(\sigma(K) \land \nu(\epsilon) \geq 1) - \rho_{n} P_{\theta_{n-1}} g_{\theta_{n-1}}(X_n) I(\sigma(K) \land \nu(\epsilon) \geq n) \right|^p \leq 2^{p-1} \left( \rho_{n} W_{P}(X_0) I(\sigma(K) \land \nu(\epsilon) \geq 1) + \sup_{n \geq 1} \rho_{n} W_{P}(X_n) I(\sigma(K) \land \nu(\epsilon) \geq n) \right) \leq 2^{p-1} \sum_{k=1}^{\infty} \rho_{k} W_{P}(X_k) I(\sigma(K) \land \nu(\epsilon) \geq k).
\]

The proof follows from (A2), the inequality \( \sum_{k=1}^{n} \rho_{k} \leq (\sum_{k=1}^{n} \rho_{k}^2)^{p/2} \) for \( p \geq 2 \). \[\]

**Appendix B. Proof of Proposition 5.1.** The following proposition is an obvious extension of Theorem 2.3 in [25] (see also [13] for a similar statement).

**Proposition B.1.** Suppose that \( P \) is irreducible and aperiodic and that there exists a set \( C \in \mathcal{B}(X) \) such that (5.1) is satisfied for some integer \( m \) and \( \delta > 0 \) and that there is a drift to \( C \) in the sense that, for some \( \lambda < 1, \beta \) and a function \( V: X \rightarrow [1, \infty) \),

\[
PV(x) \leq \lambda V(x) \quad \forall x \notin C \quad \text{and} \quad \sup_{x \in C} (V(x) + PV(x)) \leq b.
\]

Then, there exist constants \( K \) and \( \rho < 1 \), depending only upon \( m, \delta, \lambda, b \), such that, for all \( x \in X \), and all \( g \in L_V \)

\[
\| P^k g - \pi(g) \|_V \leq K \rho^k \| g \|_V.
\]

In addition, \( u = \sum_{n \geq 0} (P^k g - \pi(g)) \) is a solution of the Poisson equation \( u - Pu = g - \pi(g) \). Theorem 2.3 in [25] is stated in the strongly aperiodic case, i.e. where \( C \) is a \( (1, \delta) \) small set. Explicit but intricate expressions for \( K \) and \( \rho \) (in terms of the constants \( m, \delta, \lambda, b \)) are given in this reference. Partial extensions to the general aperiodic case is considered in Theorem 2.4. Meyn and Tweedie results are based on splitting and regeneration techniques, Sharper and simpler bounds have been recently obtained in [13] using a coupling technique. These results have been derived in the strongly aperiodic case; extensions to the general aperiodic case can be considered in the same framework.

In some applications the drift condition (B.2) is expressed on iterates of the kernel instead of the kernel itself. The following simple lemma can then be used to verify the conditions of the proposition above.

**Lemma B.2.** Assume that, there exist integer \( m \) and constants \( \lambda < 1 \) and \( \kappa \) such that

\[
P^m V(x) \leq \lambda V(x), \quad \forall x \notin C, \quad \text{and} \quad PV(x) \leq \kappa V(x), \forall x \in X,
\]

\( \tag{B.3} \)
for some function $V : X \to [1, \infty]$. Then, there exists a function $W$ and constants $\rho < 1$, $c$, $C$ (depending only upon $m, n, \lambda$) such that
\[
P W(x) \leq \rho W(x), \quad x \notin C \quad \text{and} \quad cV \leq W \leq CV.
\]

Proof. Define $W = \lambda^{1-1/m} V + \lambda^{1-2/m} PV + \ldots + \lambda^{m-1} V$. For $x \notin C$, we have
\[
P W(x) < \lambda V(x) + \lambda^{1-1/m} PV(x) + \ldots + \lambda^{1/m} PV^{m-1}(x) \leq \lambda^{1/m} W(x)
\]
Note finally that
\[
\lambda^{1-1/m} V \leq W \leq \left( \sum_{i=1}^{m-1} \lambda^{1-i/m} \kappa^{i-1} \right) V.
\]

\begin{corollary}
Assume that $P$ is \( \psi \)-irreducible and aperiodic and that (B.3) holds for $m \geq 1$, $0 < \lambda < 1$, $\kappa < \infty$, and some $(m, \delta)$ small set $C$ ($\delta > 0$). Then, $P$ has a unique invariant distribution $\pi$ and there exist constant $K$ and $\rho < 1$ (depending only upon $m$, $\lambda$, $\kappa$, $\delta$) such that, for all $g \in \mathcal{L}_V$,
\[
\sup_{k \geq 0} \| P^k g \|_V \leq K \| g \|_V \quad , \quad |\pi(g)| \leq K \| g \|_V \quad \text{and} \quad \| P^k g - \pi(g) \|_V \leq K \rho^k \| g \|_V.
\]
\end{corollary}

\begin{proposition}
Assume (DRI1)-(DRI3). Then, there exist a constant $C$ and $\rho < 1$ such that, for all $g \in \mathcal{L}_V$, with $q = 1$ or $q = p$
\[
\sup_{\theta \in K} \| P^k_{\Theta} g - \pi_{\Theta}(g) \|_{V^q} \leq C \rho^k \| g \|_{V^q},
\]
\[
\sup_{(\theta, \theta') \in K \times K} |\theta - \theta'|^{-\beta} \| P^k_{\Theta} g - P^k_{\Theta'} g \|_{V^q} \leq C \| g \|_{V^q}.
\]
\end{proposition}

\begin{proof}
Eq. (B.5) follows from Proposition B.1. To prove (B.6) write, for all $(\theta, \theta') \in \Theta \times \Theta$, all $n \in \mathbb{N}$, and all $g \in \mathcal{L}_V$
\[
P^n_{\Theta}(x) - P^n_{\Theta}(x) = \sum_{k=0}^{n-1} P^k_{\Theta}(x) (P^{n-j-1}_{\Theta}(x) - \pi_{\Theta}(x)).
\]
Eq. (B.5) shows that there exists a constant $C$ such that, for any $l \geq 0$,
\[
\sup_{\theta \in K} \| P^l_{\Theta} g - \pi_{\Theta}(g) \|_{V^q} \leq C \| g \|_{V^q} \quad \rho^l
\]
Under assumption (DRI3) we thus have, for any $l \geq 0$,
\[
\| (P_{\Theta} - P^n_{\Theta})(x) - \pi_{\Theta}(x) \|_{V^q} \leq C |\theta - \theta'|^{\beta} \| (P^n_{\Theta}(x) - \pi_{\Theta}(x)) \|_{V^q} \leq C |\theta - \theta'|^{\beta} \| g \|_{V^q}\rho^l,
\]
which concludes the proof.
\end{proof}
Proof of Proposition 5.1. Under (DRI), $P_\theta$ is positive recurrent and admits a single stationary measure $\pi_\theta$, which verifies $\sup_{\theta \in K} \pi_\theta (V^*) < \infty$ which implies that $\sup_{\theta \in K} |h(\theta)| < \infty$.

Proof. [Proof Eq. (5.8)] Let $x_0 \in X$ and $k \in \mathbb{N}$. Write $h(\theta) - h(\theta') = A(\theta, \theta') + B(\theta, \theta') + C(\theta, \theta')$ where

$$A(\theta, \theta') = (h(\theta) - P^k_\theta H_\theta(x_0)) + (P^k_\theta H_{\theta'}(x_0) - h(\theta')),$$

$$B(\theta, \theta') = P^k_\theta H_{\theta}(x_0) - P^k_\theta H_{\theta'}(x_0),$$

$$C(\theta, \theta') = P^k_\theta H_{\theta}(x_0) - P^k_\theta H_{\theta'}(x_0).$$

Propositions B.1 and B.4 show that there exist constants $C$ and $\rho < 1$ such that, for all $(\theta, \theta') \in K \times K$,

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

$$|B(\theta, \theta')| \leq C \sup_{\theta \in K} ||H_{\theta}|| V(\theta - \theta'|^3 V(x_0),$$

$$|C(\theta, \theta')| \leq \int_X P^k_\theta \rho(x_0, dy) |H_{\theta}(y) - H_{\theta'}(y)|,$$

$$\leq C|\theta - \theta'|^3 \int P^k_\theta \rho(x_0, dy) V(y) \leq C|\theta - \theta'|^3 V(x_0).$$

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

$$|h(\theta) - h(\theta')| \leq C V(x_0) (\rho^k + |\theta - \theta'|^3).$$

(B.10)

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

Proof of Eq. (5.9). Using Eq. (5.8), Proposition B.1 and B.4, there exists a constant $C$ such that, for all $(\theta, \theta') \in K$ we have

$$|((P^k_\theta H_{\theta}(x) - h(\theta)) - (P^k_\theta H_{\theta'}(x) - h(\theta'))|$$

$$\leq |P^k_\theta H_{\theta}(x) - P^k_\theta H_{\theta'}(x)| + |P^k_\theta H_{\theta'}(x) - P^k_\theta H_{\theta'}(x)| + |h(\theta) - h(\theta')|,$$

$$\leq C|\theta - \theta'|^3 V(x).$$

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

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

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

$$|P^k_\theta g_{\theta}(x) - P^k_\theta g_{\theta'}(x)| \leq \sum_{k=s}^{\infty} |((P^k_\theta H_{\theta}(x) - h(\theta)) - (P^k_\theta H_{\theta'}(x) - h(\theta'))|$$

$$\leq CV(x) \left\{ N|\theta - \theta'|^3 + \sum_{k=N+s}^{\infty} \rho^k \right\}$$

$$\leq CV(x) \left\{ N|\theta - \theta'|^3 + \frac{\rho^{N+s}}{1 - \rho} \right\}.$$

The proof follows by setting $N = [\beta \log |\theta - \theta'| / \log \rho]$, for $|\theta - \theta'| \leq \delta < 1$, $\theta \neq \theta'$, $N = s$ otherwise, and using the fact that for any $0 < \alpha < \beta$, $|\theta - \theta'|^\alpha \log |\theta - \theta'| = o(|\theta - \theta'|^\alpha)$. \qed
Proof. [Proof Eq. (5.10)] Let \( \rho = (\rho_k, k \geq 0) \) be a non-increasing sequence of positive numbers and let \( \mathcal{K} \) be a compact subset of \( \Theta \). (DR1) (5.3) shows that, for all \( k \geq 0 \), \( l \geq 0 \), all \( x \in X \),
\[
\sup_{\theta \in \mathcal{K}} \mathbb{E}^\rho[V^\rho(X_{k+l})I(\sigma(\cdot) \geq k + l)] \leq \kappa^l V^\rho(X_k) I(\sigma(\cdot) \geq k).
\] (B.11)
We will show that there exist constants \( \epsilon > 0, 0 < \rho < 1 \) and \( C \) such that, for all \( k \)
\[
\mathbb{E}^\rho[V^\rho(X_{k+m})I(\sigma(\cdot) \land \nu \geq k + m)] \leq \mathbb{E}^\rho[V^\rho(X_k)I(\sigma(\cdot) \land \nu \geq k)] + C.
\] (B.12)
For \( n \in \mathbb{N} \), write \( n = um + v \), where \( v \in \{0, \ldots, m - 1\} \). (B.12) shows that
\[
\mathbb{E}^\rho_{\epsilon,0}[V^\rho(X_{um+v})I(\sigma(\cdot) \land \nu \geq um + v)] \leq \rho^m \mathbb{E}^\rho[X_{\epsilon,0}[V^\rho(X_v)I(\sigma(\cdot) \land \nu \geq v)] + \frac{C}{1 - \rho}
\]
and the proof follows from (B.11). It remains to prove (B.12). We repeatedly use the following lemma adapted from [4] (Lemma 3, p. 292.)

**Lemma B.5.** Assume (DR1). Let \( \psi : \Theta \times X \to \mathbb{R} \) be a function verifying \( \sup_{\theta \in \mathcal{K}} \|\psi(\theta)\|_{V^\rho} < \infty \). Then, for any \( \epsilon > 0 \), for any \( l \geq 1 \) there exist a constant \( C_l \) such that, for all \( k \geq 0 \),
\[
\mathbb{E}^\rho[|\psi_k(\theta, X_{k+l})I(\sigma(\cdot) \land \nu \geq k + l)] \leq \mathbb{E}^\rho[|P_{\theta_k} \psi_k(X_{k+l-1})I(\sigma(\cdot) \land \nu \geq k + l - 1)] + C_l \epsilon^\alpha \sup_{\theta \in \mathcal{K}} \|\psi(\theta)\|_{V^\rho} V^\rho(X_k) I(\sigma(\cdot) \land \nu \geq k).
\]
**Proof.**
\[
\mathbb{E}^\rho[|\psi_k(\theta, X_{k+l})I(\sigma(\cdot) \land \nu \geq k + l)] \leq \mathbb{E}^\rho[|P_{\theta_k} \psi_k(X_{k+l-1})I(\sigma(\cdot) \land \nu \geq k + l)] + R_l
\]
where
\[
R_l = \mathbb{E}^\rho[|P_{\theta_k} \psi_k(X_{k+l-1})I(\sigma(\cdot) \land \nu \geq k + l)] \leq R_l \epsilon^\alpha \sup_{\theta \in \mathcal{K}} \|\psi(\theta)\|_{V^\rho} V^\rho(X_k) I(\sigma(\cdot) \land \nu \geq k).
\]
Under (DR3), there exists a constant \( C \) such that for all \( x \in X \)
\[
|P_{\theta_k} \psi_k(x)I(\sigma(\cdot) \land \nu \geq k + l)| \leq C \sup_{\theta \in \mathcal{K}} \|\psi(\theta)\|_{V^\rho} V^\rho(x) I(\sigma(\cdot) \land \nu \geq k + l).
\]
Finally, (DR1) implies that
\[
\mathbb{E}^\rho[V^\rho(X_{k+l-1})I(\sigma(\cdot) \land \nu \geq k + l)] \leq \kappa^l V^\rho(X_k) I(\sigma(\cdot) \land \nu \geq k),
\]
which implies
\[
R_l \leq C \epsilon^\alpha \sup_{\theta \in \mathcal{K}} \|\psi(\theta)\|_{V^\rho} V^\rho(X_k) I(\sigma(\cdot) \land \nu \geq k).
\]
Using repeatedly the lemma above, we may write
\[
\mathbb{E}^\rho[V^\rho(X_{k+m})I(\sigma(\cdot) \land \nu \geq k + m)] \leq \mathbb{E}^\rho[P_{\theta_k} V^\rho(X_{k+m-1})I(\sigma(\cdot) \land \nu \geq k + m - 1)] + C_m \epsilon^\alpha V^\rho(X_k) I(\sigma(\cdot) \land \nu \geq k)
\]
\[
\mathbb{E}^\rho[P_{\theta_k} V^\rho(X_{k+m-2})I(\sigma(\cdot) \land \nu \geq k + m - 2)] + C_{m-1} \epsilon^\alpha V^\rho(X_k) I(\sigma(\cdot) \land \nu \geq k)
\]
\[
\vdots
\]
\[
P_{\theta_k} V^\rho(X_k) I(\sigma(\cdot) \land \nu \geq k) + \left( \sum_{i=0}^{m-1} C_{m-i} \epsilon^\alpha \right) V^\rho(X_k) I(\sigma(\cdot) \land \nu \geq k).
\]
The proof follows from (DR1) for $\epsilon$ sufficiently small. □

Appendix C. Proof of Proposition 5.2. We first recall two results from BMP, which are key to the following developments. These results are important as they establish computable bounds for geometric ergodicity over the space $\mathcal{L}(q)$ without assuming irreducibility conditions (and hence the existence of small sets). $C$ is a generic constant which may take different values upon each appearance.

**Lemma C.1** ([1], Lemma 1, pp. 252). Let $g : \mathcal{X} \rightarrow \mathcal{X}$. Assume that there exist constants $K$, $K_g$, $\rho < 1$ and $s \geq 0$ such that, for any $n \in \mathbb{N}$,

\[ |P^ng(x) - P^ng(y)| \leq K_g \rho^n (1 + |x|^s + |y|^s), \quad \forall (x, y) \in \mathcal{X} \times \mathcal{X}, \tag{C.1} \]

\[ \int_{\mathcal{X}} P(x, dy) |y|^s \leq K (1 + |x|^s). \tag{C.2} \]

Then, there exist $\gamma$ and a constant $C$ (depending only on $K$ and $\rho$ but not on $K_g$) such that, for any $x \in \mathcal{X}$ and any $n \in \mathbb{N}$,

\[ |P^n g(x) - \gamma| \leq CK_g \rho^n (1 + |x|^s), \quad \forall x \in \mathcal{X}. \tag{C.3} \]

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

**Proof.** We have,

\[
|P^{n+1} g(x) - P^n g(x)| = \left| \int_{\mathcal{X}} P(x, dy) (P^n g(y) - P^n g(x)) \right| \\
\leq K_g \rho^n \int_{\mathcal{X}} P(x, dy)(1 + |x|^s + |y|^s) \leq K_g (1 + K) \rho^n (1 + |x|^s).
\]

This shows that, for any $x \in \mathcal{X}$, \(\sum_{n \geq 0} |P^{n+1} g(x) - P^n g(x)| \leq C K_g (1 + |x|^s)\) for some constant $C$, and thus that \(\lim_{n \to \infty} P^n g(x)\) exists. Eq (C.1) shows that this limit does not depend on $x$. Denote $\gamma = \lim_{n \to \infty} P^n g(x)$. Now for any $k \geq 1$,

\[ |P^{n+k} g(x) - P^n g(x)| \leq C K_g \rho^n (1 + |x|^s), \quad \forall x \in \mathcal{X}, \]

and Eq. (C.3) follows by letting $k \to \infty$. Set $u(x) = \sum_{n \geq 0} (P^n g(x) - \gamma)$. Since $|u(x)| \leq C (1 + |x|^s)$ for some constant $C$, the dominated convergence theorem implies that

\[ \int_{\mathcal{X}} P(x, dy) u(y) = \sum_{n \geq 1} (P^n g(x) - \gamma) = u(x) - (g(x) - \gamma), \]

and thus $u - Pu = g - \gamma$. □

**Proposition C.2** ([1], Proposition 2, pp. 253). Assume that there exist constants $C$, $\rho < 1$ and $s \geq 0$ such that, for any $g \in \mathcal{L}(s)$,

\[ \int_{\mathcal{X}} P(x, dy) |y|^{s+1} \leq C (1 + |x|^{s+1}) \quad \forall x \in \mathcal{X} \tag{C.4} \]

\[ |P^n g(x) - P^n g(y)| \leq C \rho^n (|g_x| \vee |g_{y+1}|)(1 + |x|^{s+1} + |y|^{s+1}) \quad \forall (x, y) \in \mathcal{X} \times \mathcal{X}, \forall n \geq 0. \tag{C.5} \]

Assume in addition that for any Lipschitz function $\phi$ having a compact support, $x \mapsto P\phi(x)$ is continuous. Then, $P$ has a single invariant probability measure $\pi$, and there exist a constants $C$ (that depend only on
Let $C, \rho$ and $s$ be such that for any $g \in \text{Li}(s)$,
\begin{equation}
\int_X \pi(dy)(1 + |y|)^{s+1} \leq C \tag{C.6}
\end{equation}
\begin{equation}
\sup_{j \geq 1} \left\| \int_X P^j(x, dy)(1 + |y|)^{s+1} \right\|_{s+1} \leq C, \tag{C.7}
\end{equation}
\begin{equation}
\|P^ng - \pi(g)\|_{s+1} \leq C \rho^n (|g|_s \vee \|g\|_{s+1}), \tag{C.8}
\end{equation}
and $u = \sum_{n \geq 0} (P^n g - \pi(g))$ is a solution of the Poisson equation $(I - P)u = g - \pi(g)$.

**Proof.** The only difficulty here is to prove the existence of the invariant probability measure $\pi$ (see [4], Proposition 2). The other statements are then a direct consequence of Lemma C.1, since $V_{s+1} \in \text{Li}(s)$. \qed

**Proposition C.3** ([4], Proposition 3, pp. 255). *Assume that there exist integers $m \geq 1, s \geq 0$ and constants $C$ and $\rho < 1$ and $x_0 \in X$, such that for any $g \in \text{Li}(q)$,
\begin{equation}
\int_X P(x_0, dy)|y|^{s+1} \leq C, \tag{C.9}
\end{equation}
\begin{equation}
[P^g]_s \leq C|g|_s \text{ and } [P^n g]_s \leq \rho^n |g|_s. \tag{C.10}
\end{equation}
Then, for any Lipschitz function $\phi$ having a bounded support $x \mapsto P\phi(x)$ is continuous and there exist constants $C$ and $\rho < 1$ such that (C.4) and (C.5) hold.*

**Proof.** Let $\phi$ be a Lipschitz function having a bounded support; $\phi$ belongs to $\text{Li}(s)$ and (C.10) shows that $P\phi$ is locally Lipschitz and hence continuous. Note that $V_{s+1} \in \text{Li}(s)$. We have
\begin{equation*}
P V_{s+1}(x) = PV_{s+1}(x_0) + PV_{s+1}(x) - PV_{s+1}(x_0),
\end{equation*}
\begin{equation*}
\leq 1 + C + C [V_{s+1}]_s|x - x_0|(1 + |x|^s + |x_0|^s) \leq CV_{s+1}(x),
\end{equation*}
showing (C.4). We now turn to the proof of (C.5). Since for any $g \in \text{Li}(s)$, $[P^g]_s \leq C|g|_s$, then for any $n \geq 1$, $[P^n g]_s \leq C^n |g|_s$. Hence,
\begin{equation*}
\sup_{k \in \{1, \ldots, m-1\}} [P^k g]_s \leq C^{m-1}|g|_s.
\end{equation*}
Similarly, since $[P^n g]_s \leq \rho^n |g|_s$, then for any integer $a \geq 1$, $[P^{(m^n)} g]_s \leq \rho^n |g|_s$. Write $n = am + b$ with $b \in \{0, \ldots, m-1\}$ so that,
\begin{equation*}
[P^n g]_s = [P^{(m^n + b)} g]_s \leq C^{m-1}[P^{(m^n)} g]_s \leq C^{m-1}\rho^n |g|_s.
\end{equation*}
Setting $\rho = \rho^{1/m}$, there exists $C$ such that the latter equation may be rewritten as
\begin{equation*}
[P^n g]_s \leq C\rho^n |g|_s,
\end{equation*}
which proves the first assertion in Eq. (C.5). \qed

We now state and prove an original lemma, useful for the proof of Proposition 5.2.

**Lemma C.4.** *Assume (BMP). Then, for any compact set $K \subset \Theta$, there exist constants $C$ and $\rho < 1$, such that for any $n \in \mathbb{N}$ and $g \in \text{Li}(q)$, then
\begin{equation}
\sup_{\theta \in K} |P^n g|_\theta \leq C\rho^n, \tag{C.11}
\end{equation}
\begin{equation}
\sup_{\theta \in K} \sup_{j \geq 1} \left\| P^j V_{q+1} \right\|_{q+1} \leq C, \tag{C.12}
\end{equation}
\begin{equation}
\sup_{(\theta, \theta') \in K \times K} \left\| \theta - \theta' \right\|^{-\beta} \left\| P^n g - P^{n\beta} g \right\|_{q+1} \leq C |g|_q. \tag{C.13}
\end{equation}
Proof. We first prove that the assumptions of Proposition C.3 are satisfied. From (BMP1),

\[ \int_X P_0(x, dy) |y|^{q+1} \leq \left( \int_X P_0(x, dy) |y|^{q+1} \right)^{\frac{1}{q+1}} \leq C(1 + |x|^{q+1})^{\frac{1}{q+1}} \leq C(1 + |x|^{q+1}) , \]

for some constant \( C \), so that (C.9) holds. From (BMP3), Eq. (5.16), (C.10) is satisfied and therefore (C.11) and (C.12) follow from Proposition C.3 with \( s = q \). We now prove the last statement (C.13). For any \( (\theta, \theta') \in \Theta \times \Theta \) and \( n \in \mathbb{N} \) we have,

\[ P^n_\theta - P^n_{\theta'} = \sum_{j=0}^{n-1} P^j_\theta (P_\theta - P_{\theta'}) P^{n-j-1}_{\theta'} . \]

For \( g \in L_i(q) \) and \( x \in X \), we have by applying (5.18) in (BMP4),

\[ |P^n_\theta g(x) - P^n_{\theta'} g(x)| \leq \sum_{j=0}^{n-1} P^j_\theta (P_\theta - P_{\theta'}) P^{n-j-1}_{\theta'} |g(x)| , \]

which together with Eqs. (C.11) and (C.12) allows us to conclude.  \( \boxdot \)

Proof of Proposition 5.2. The existence, for any \( \theta \in \Theta \), of the stationary distribution \( \pi_\theta \), the function \( h_\theta \), the solution of the Poisson equation and the bound for this solution follows from the application of Propositions C.2 and C.3.

Proof. [Proof of (5.20)]. Let \( x_0 \in X \) and \( k \in \mathbb{N} \). Notice that \( h(\theta) - h(\theta') = A(\theta, \theta') + B(\theta, \theta') + C(\theta, \theta') \) where \( A(\theta, \theta') \), \( B(\theta, \theta') \) and \( C(\theta, \theta') \) are defined in (B.7), (B.8) and (B.9), respectively. Under (BMP2), \( \sup_{\theta \in \mathcal{K}} (|H_\theta|_2 \vee |H_\theta|_{q+1}) < \infty \). Lemma C.4 and Proposition C.2 shows that there exists constants \( C \) and \( \rho < 1 \) such that,

\begin{align}
\sup_{(\theta, \theta') \in \mathcal{K}} |P^n_\theta H_\theta|_q &\leq C[H_\theta]_q \rho^n \quad \forall n \geq 0 \quad \text{(C.14)} \\
\sup_{(\theta, \theta') \in \mathcal{K}} \|P^n_\theta H_\theta - \pi_\theta (H_\theta)\|_{q+1} &\leq C \rho^n \quad \text{(C.15)} \\
\sup_{\theta \in \mathcal{K}} \sup_{j \geq 1} \left\| \int_X P^j_\theta (x, dy) (1 + |y|^{q+1}) \right\|_{q+1} &\leq C \quad \text{(C.16)} \\
\sup_{(\theta, \theta', \theta') \in \mathcal{K} \times \mathcal{K}} |\theta - \theta'|^{-\beta} \|P^n_\theta H_{\theta'} - P^n_{\theta'} H_{\theta'}\|_{q+1} &\leq C \rho^n \quad \forall n \geq 0. \quad \text{(C.17)}
\end{align}

(C.15) and (C.17) imply that, for all \( k \geq 0 \)

\begin{align}
|A(\theta, \theta')| &\leq C \rho^k (1 + |x_0|^{q+1}) , \quad \text{(C.18)} \\
|B(\theta, \theta')| &\leq C |\theta - \theta'|^{\beta} (1 + |x_0|^{q+1}) . \quad \text{(C.19)}
\end{align}
Under (BMP2), there exists a constant $C$ such that, for all $(\theta, \theta') \in \mathcal{K} \times \mathcal{K}$ and all $x \in X$, $|H_\theta(x) - H_\theta'(x)| \leq C(1 + |x|^{q+1})|\theta - \theta'|^\beta$. Hence, using (C.16), we have

$$|C(\theta, \theta')| \leq \int_X P^\rho_\theta(x_0, dy) |H_\theta(y) - H_\theta'(y)|,$$

$$\leq C|\theta - \theta'|^\beta \int P^\rho_\theta(x_0, dy)(1 + |y|^{q+1}) \leq C|\theta - \theta'|^\beta (1 + |x_0|^{q+1}). \quad (C.20)$$

Gathering the results above, there exists a constant $C$ such that, for any $(\theta, \theta') \in \mathcal{K} \times \mathcal{K}$,

$$|h(\theta) - h(\theta')| \leq C(1 + |x_0|^{q+1}) \left( \rho^k + |\theta - \theta'|^\beta \right). \quad (C.21)$$

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

Proof. [Proof of (5.21)] There exists a constant $C$ and $\rho < 1$ such that, for all $(\theta, \theta') \in \mathcal{K} \times \mathcal{K}$ and all $x \in X$,

$$|P^\rho_\theta h_\theta(x) - h(\theta)) - (P^\rho_\theta h_\theta'(x) - h(\theta'))|$$

$$\leq |P^\rho_\theta h_\theta(x) - P^\rho_\theta h_\theta'(x)| + |P^\rho_\theta h_\theta(x) - P^\rho_\theta h_\theta'(x)| + |h(\theta) - h(\theta')| \leq C(1 + |x|^{q+1})|\theta - \theta'|^\beta,$$

$$|P^\rho_\theta h_\theta(x) - h(\theta)| \leq C\rho^k (1 + |x|^{q+1}).$$

The proof is concluded as above. ◻

Appendix D. Proof of technical lemmas.

Proof. [Lemma 3.1] We first prove 1. Let

$$\alpha_{\mathcal{K}} = \inf \left\{ -\langle \nabla w(a), h(a) \rangle, a \in \mathcal{K} \right\} > 0. \quad (D.1)$$

Since $\mathcal{K}$ is compact, $\Theta$ open and $\sup_{a \in \mathcal{K}} |h(a)| < \infty$, for any $0 < \delta_{\mathcal{K}} < \alpha_{\mathcal{K}}$, we can find $\lambda_{\mathcal{K}} > 0$, $\beta_{\mathcal{K}} > 0$ such that, for all $c \in X$, $\lambda \leq \lambda_{\mathcal{K}}$, $|c| \leq \beta_{\mathcal{K}}$, $0 \leq t \leq 1$,

$$a + \lambda t h(a) + \lambda t c \in \Theta$$

$$\left| \left\langle \nabla w(a + \lambda t h(a) + \lambda t c), h(a) + c \right\rangle - \left\langle \nabla w(a), h(a) \right\rangle \right| \leq \alpha_{\mathcal{K}} - \delta_{\mathcal{K}}.$$

Then we have

$$w(a + \lambda t h(a) + \lambda t c) - w(a) = \lambda \int_0^1 \left\langle \nabla w(a + t \lambda h(a) + t \lambda c), h(a) + c \right\rangle \, dt$$

$$= \lambda \left( \left\langle \nabla w(a), h(a) \right\rangle + \lambda \int_0^1 \left( \left\langle \nabla w(a + t \lambda h(a) + t \lambda c), h(a) + c \right\rangle - \left\langle \nabla w(a), h(a) \right\rangle \right) \, dt \right)$$

$$= -\delta_{\alpha} + \lambda (\alpha - \delta_{\mathcal{K}}) = -\lambda \delta_{\mathcal{K}}.$$
REFERENCES