Bernoulli **20**(2), 2014, 545–585
DOI: 10.3150/12-BEJ497

Markovian stochastic approximation with expanding projections

CHRISTOPHE ANDRIEU[1] and MATTI VIHOLA[2]

[1]*School of Mathematics, University of Bristol, BS8 1TW, United Kingdom.*
E-mail: C.Andrieu@bristol.ac.uk
[2]*Department of Mathematics and Statistics, University of Jyväskylä, P.O. Box 35, FI-40014, Finland.*
E-mail: matti.vihola@iki.fi

Stochastic approximation is a framework unifying many random iterative algorithms occurring in a diverse range of applications. The stability of the process is often difficult to verify in practical applications and the process may even be unstable without additional stabilisation techniques. We study a stochastic approximation procedure with expanding projections similar to Andradóttir [*Oper. Res.* **43** (1995) 1037–1048]. We focus on Markovian noise and show the stability and convergence under general conditions. Our framework also incorporates the possibility to use a random step size sequence, which allows us to consider settings with a non-smooth family of Markov kernels. We apply the theory to stochastic approximation expectation maximisation with particle independent Metropolis–Hastings sampling.

Keywords: expectation maximisation; independent Metropolis–Hastings; particle Markov chain Monte Carlo; stability; stochastic approximation

1. Introduction

Stochastic approximation (SA) is concerned with finding the zeros of a function defined on the space $\Theta \subset \mathbb{R}^d$ as

$$h(\theta) := \int_{\mathsf{X}} H(\theta, x)\pi_\theta(\mathrm{d}x), \tag{1.1}$$

where $\{\pi_\theta\}_{\theta\in\Theta}$ is a family of probability distributions on a generic measurable space $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ and $H : \Theta \times \mathsf{X} \to \Theta$ is a measurable function. In numerous situations h behaves like a gradient, suggesting that a recursion of the type $\theta_{i+1} = \theta_i + \gamma_{i+1}h(\theta_i)$ where $(\gamma_i)_{i\geq 1}$ is a sequence of nonnegative step sizes decaying to zero, can be used to find the aforementioned roots.

Often in applications, the integral (1.1) needs to be approximated numerically. We focus here on methods relying on Monte Carlo simulation where sampling exactly from π_θ for any $\theta \in \Theta$ is not possible directly and instead Markov chain Monte Carlo methods are used. Let $\{P_\theta\}_{\theta\in\Theta}$ be a family of Markov transition probabilities with stationary distributions $\{\pi_\theta\}_{\theta\in\Theta}$, respectively. Then, the standard SA recursion with Markovian dynamic is as follows

$$\begin{aligned} X_{i+1}|\theta_0, X_0, \ldots, \theta_i, X_i &\sim P_{\theta_i}(X_i, \cdot), \\ \theta_{i+1} &= \theta_i + \gamma_{i+1}H(\theta_i, X_{i+1}). \end{aligned}$$

Stability of this process is far from obvious and a significant effort has been dedicated to its study (e.g., [7], Section 7.3). Problems occur in particular when ergodicity, a term to be made more precise later, of P_θ vanishes as θ approaches a set of critical values denoted $\partial\Theta$ hereafter. Younes [30], Section 6.3, gives an example of a situation where the Robbins–Monro algorithm fails for this reason.

Cures include projection on a fixed set $\mathcal{R}_0 \subset \Theta$, that is, given a projection mapping $\Pi_{\mathcal{R}_0} : \Theta \setminus \mathcal{R}_0 \to \mathcal{R}_0$, one can define [20,21]

$$\begin{aligned}\theta^*_{i+1} &= \theta_i + \gamma_{i+1} H(\theta_i, X_{i+1}),\\ \theta_{i+1} &= \theta^*_{i+1}\mathbb{I}\{\theta^*_{i+1} \in \mathcal{R}_0\} + \Pi_{\mathcal{R}_0}(\theta^*_{i+1})\mathbb{I}\{\theta^*_{i+1} \notin \mathcal{R}_0\}.\end{aligned}$$

Projection on a fixed set $\mathcal{R}_0$ might not be satisfactory when for example the location of the zeros of $h(\theta)$ is not known a priori. It is also possible that the projection induces spurious attractors on the boundary of $\mathcal{R}_0$.

Adaptive projections overcome these difficulties by considering an increasing sequence of projection sets $\{\mathcal{R}_i\}_{i\geq 0}$ which forms a covering of Θ. The process is defined through [4,11–13, 28]

$$\begin{aligned}\theta^*_{i+1} &= \theta_i + \gamma_{i+1} H(\theta_i, X_{i+1}),\\ \theta_{i+1} &= \theta^*_{i+1}\mathbb{I}\{\theta^*_{i+1} \in \mathcal{R}_{r_i}\} + \Pi_{\mathcal{R}_0}(\theta^*_{i+1})\mathbb{I}\{\theta^*_{i+1} \notin \mathcal{R}_{r_i}\},\\ r_{i+1} &= r_i + \mathbb{I}\{\theta^*_{i+1} \notin \mathcal{R}_{r_i}\},\end{aligned}$$

where r_i is the indicator of the current reprojection set and $r_0 \equiv 0$. Adaptive projections can be shown to lead to stable recursions under rather general conditions. In the case of a Markovian noise, one usually modifies also X_{i+1} so that [4]

$$\begin{aligned}X_{i+1}|\theta_0, X_0, \ldots, \theta_i, X_i \;&\sim\; P_{\theta_i}(X^*_i, \cdot) \qquad \text{with}\\ X^*_i \;&:= \mathbb{I}\{\theta^*_i \in \mathcal{R}_{i-1}\}X_i + \mathbb{I}\{\theta^*_i \notin \mathcal{R}_{i-1}\}\hat{\Pi}_{\mathsf{K}_0}(X_i),\end{aligned}$$

where $\hat{\Pi}_{\mathsf{K}_0} : \mathsf{X} \to \mathsf{K}_0$ maps X_i to a suitable (usually compact) set $\mathsf{K}_0 \subset \mathsf{X}$. This corresponds effectively to 'restarting' the process, with a smaller step size sequence and a bigger feasible set $\mathcal{R}_{r_i+1}$. One can show that the projections occur finitely often under fairly general conditions, whence the process is eventually stable [4]. In practice, this algorithm may be wasteful if $\{\mathcal{R}_i\}_{i\geq 0}$ or K_0 are ill-defined, and the projections occur frequently.

We focus here on the study of a different stabilising approach where projection occurs on an expanding (with time) sequence of projection sets $\{\mathcal{R}_i\}$. Our approach is similar to Andradóttir's [1]; see also [26,27], but we consider a more general framework with two major differences. First, we focus on a Markovian noise setting, and second, we allow the step size sequence, now denoted $(\Gamma_i)_{i\geq 1}$, to be random.[1] Our analysis is inspired by earlier related work in adaptive Markov chain Monte Carlo [25]. The generic algorithm can be given as follows.

[1]The recent work of Sharia [27] includes random step sizes as well, but our assumptions on Γ_i are completely different.

Algorithm 1.1. *Let $\{\mathcal{R}_i\}_{i\geq 0}$ be subsets of Θ and let the weights $(\Gamma_i)_{i\geq 1}$ be nonnegative random variables. The stochastic approximation process $(\theta_i, X_i)_{i\geq 0}$ with expanding projection sets $\{\mathcal{R}_i\}_{i\geq 0}$ is defined for any starting point $(\theta_0, X_0) \equiv (\theta, x) \in \mathcal{R}_0 \times \mathsf{X}$ and recursively for $i \geq 0$ as follows*

$$\begin{aligned} X_{i+1}|\mathcal{F}_i &\sim P_{\theta_i}(X_i, \cdot), \\ \theta^*_{i+1} &= \theta_i + \Gamma_{i+1} H(\theta_i, X_{i+1}), \\ \theta_{i+1} &= \theta^*_{i+1}\mathbb{I}\{\theta^*_{i+1} \in \mathcal{R}_{i+1}\} + \theta^{\mathrm{proj}}_{i+1}\mathbb{I}\{\theta^*_{i+1} \notin \mathcal{R}_{i+1}\}, \end{aligned}$$

where $\mathcal{F}_i$ stands for the σ-algebra generated by $\theta_0, X_0, \theta_1, X_1, \Gamma_1, \ldots, \theta_i, X_i, \Gamma_i$, and where $\theta^{\mathrm{proj}}_{i+1}$ is a $\sigma(\mathcal{F}_i, X_{i+1}, \theta^*_{i+1})$-measurable random variable taking values in $\mathcal{R}_{i+1}$.

Most common practical projection mechanisms include $\theta^{\mathrm{proj}}_{i+1} := \theta_i$ 'rejecting' an update outside the current feasible set, and $\theta^{\mathrm{proj}}_{i+1} := \Pi_{\mathcal{R}_{i+1}}(\theta^*_{i+1})$, where $\Pi_{\mathcal{R}_{i+1}} : \Theta \setminus \mathcal{R}_{i+1} \to \mathcal{R}_{i+1}$ is a measurable mapping.

In words, the expanding projections approach only ensures that θ_i is in a feasible set $\mathcal{R}_i$ but does not involve potentially harmful 'restarts' as is the case with the adaptive reprojection strategy. Note particularly that unlike with the adaptive reprojections strategy, we need not project X_{i+1} at all. We believe that these advantages can provide significantly better results in certain settings, but this is at the expense of requiring more when proving the stability and the convergence of the process. In short, we must be able to control certain quantitative criteria within each feasible set $\mathcal{R}_i$. The random step size sequence allows one to consider situations where the family of Markov kernels $\{P_\theta\}_{\theta\in\Theta}$ is not necessarily smooth in a manner that is usually considered in the stochastic approximation literature (e.g., [8]).

Other stabilisation techniques in the literature related to our approach include the state-dependent averaging framework of Younes [30] and a state-dependent step size sequence of Kamal [19]. Particularly the former shares similarities with the present work, as it also relies on quantifying the ergodicity rates of Markov kernels explicitly. Our stabilisation approach differs, however, crucially from these methods, adding only the projections to the basic Robbins–Monro algorithm. We remark also that our present approach may be used in some situations to prove the stability and convergence of an *unmodified* Robbins–Monro stochastic approximation. This is possible, loosely speaking, if one can show that projections do not occur at all with a positive probability; see [25] for an example of such a situation. We point out also the work [6] suggesting a generic method to establish the stability of unmodified Markovian Robbins–Monro stochastic approximation at the expense of more stringent assumptions.

Our main results show that the SA process $(\theta_i)_{i\geq 0}$ produced by our expanding projections algorithm 'stays away from $\partial\Theta$' almost surely for any starting point $(\theta, x) \in \mathcal{R}_0 \times \mathsf{X}$ under conditions on $H(\cdot,\cdot)$, $\{P_\theta\}_{\theta\in\Theta}$, $(\mathcal{R}_i)_{i\geq 0}$ and $(\Gamma_i)_{i\geq 1}$. Figure 1 summarises the inter-dependency between our various main conditions and results and in order to help the reader we provide a nomenclature of some of the constants involved in Appendix D.

Section 2 contains two fundamental results, Theorems 2.5 and 2.8, which both establish stability of Algorithm 1.1 under abstract noise conditions and the existence of a Lyapunov function

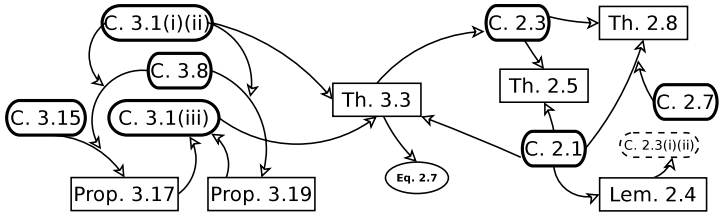

Figure 1. Road map of the main results and assumptions.

satisfying two distinct sets of assumptions which, roughly speaking, allow us to tackle instability at infinity or at a finite point. Section 3 focuses on establishing the required noise conditions with verifiable assumptions on the Markov kernels. First, Theorem 3.3 establishes the aforementioned noise conditions under Condition 3.1, which essentially involves a trade-off between the sequences $(\Gamma_i)_{i\geq 0}$ and $(\xi_i)_{i\geq 0}$ and properties of the solution of the Poisson equation related to $\{P_\theta\}_{\theta\in\Theta}$ and $H(\cdot,\cdot)$. Second, essentially assuming geometric ergodicity, Propositions 3.17 and 3.19 establish the required conditions in the scenarios where $\{P_\theta\}_{\theta\in\Theta}$ depends smoothly on θ and where it does not respectively—the latter case requires the introduction of random step-sizes $(\Gamma_i)_{i\geq 0}$ (see also the comments in the introduction of Section 3.3).

We complement our stability results in Section 4 with a discussion on how one can use existing results in the literature to obtain convergence of $(\theta_i)_{i\geq 0}$ to a zero of h. Finally, we apply our theory to a new stochastic approximation expectation maximisation algorithm involving particle independent Metropolis–Hastings sampling in Section 5.

2. General stability results

We denote throughout the article the probability distribution associated to the process $(\theta_i, X_i)_{i\geq 0}$ defined in Algorithm 1.1 and starting at $(\theta_0, X_0) \equiv (\theta, x) \in \Theta \times \mathsf{X}$ as $\mathbb{P}_{\theta,x}(\cdot)$ and the associated expectation as $\mathbb{E}_{\theta,x}[\cdot]$. For any subset $A \subset E$ of some space E, we denote A^c its complement in E. We also denote $\langle\cdot,\cdot\rangle$ the standard inner product and $|\cdot|$ the associated norm on $\Theta \subset \mathbb{R}^d$. We also use the notation $a \vee b := \max\{a, b\}$ and $a \wedge b := \min\{a, b\}$.

The approach we develop relies on the existence of a Lyapunov function $w : \Theta \to [0, \infty)$ for the recursion on θ and the subsequent proof that $\{w(\theta_i)\}$ is $\mathbb{P}_{\theta,x}$-a.s. under some adequate level. For any $M > 0$, we define the level sets $\mathcal{W}_M := \{\theta \in \Theta\colon\ w(\theta) \leq M\}$. Our general stability results are inspired by a proof due to Benveniste, Metivier and Priouret [8], Theorem 17, page 239, but differ in many respects as we shall see.

We consider two different settings concerning the way w behaves on the boundary $\partial\Theta$ of Θ. Section 2.1 assumes that $\lim_{\theta\to\partial\Theta} w(\theta) = \infty$, which is well suited for example to the case $\Theta = \mathbb{R}$ and $\partial\Theta = \{-\infty, \infty\}$. Section 2.2 considers the case where w may not be unbounded, which requires stronger assumptions on the behaviour of w. This setting subsumes for example the case where $\Theta \subset \mathbb{R}$ and $\partial\Theta$ contains some points on the real line. Both of the scenarios share the following set of assumptions.

Condition 2.1. *There exists a twice continuously differentiable function* $w : \Theta \to [0, \infty)$ *such that*

(i) *the Hessian matrix* $\mathrm{Hess}_w : \Theta \to \mathbb{R}^{d \times d}$ *of* w *is bounded so that*

$$C_w := \sup_{\theta \in \Theta} \sup_{|\theta_0|=1} \left| \mathrm{Hess}_w(\theta)\theta_0 \right| < \infty,$$

(ii) *the projection sets are increasing subsets of* Θ, *that is,* $\mathcal{R}_i \subset \mathcal{R}_{i+1}$ *for all* $i \geq 0$, *and* $\hat{\Theta} := \bigcup_{i=0}^{\infty} \mathcal{R}_i \subset \Theta$,

(iii) *there exists a constant* $M_0 > 0$ *such that for any* $\theta \in \mathcal{W}_{M_0}^c \cap \hat{\Theta}$

$$\left\langle \nabla w(\theta), h(\theta) \right\rangle \leq 0,$$

(iv) *the family of random variables* $\{\theta_i^{\mathrm{proj}}\}_{i \geq 1}$ *satisfies for all* $i \geq 1$ *whenever* $\theta_i^* \notin \mathcal{R}_i$

$$\theta_i^{\mathrm{proj}} \in \mathcal{R}_i \quad \textit{and} \quad w\left(\theta_i^{\mathrm{proj}}\right) \leq w\left(\theta_i^*\right) \qquad \mathbb{P}_{\theta,x}\text{-a.s.},$$

(v) *there exists constants* $\alpha_w, c \in [0, \infty)$ *and a non-decreasing sequence of constants* $\xi_i \in [1, \infty)$ *satisfying* $\sup_{\theta \in \mathcal{R}_i} |\nabla w(\theta)| \leq c\xi_i^{\alpha_w}$ *for all* $i \geq 0$.

Remark 2.2. Condition 2.1

(i) Can often be established by introducing a Lyapunov function defined through $w := \psi \circ \tilde{w}$, where $\psi : [0, \infty) \to [0, \infty)$ is a suitable concave function modifying the values of another Lyapunov function $\tilde{w}$ which satisfies the drift condition (iii) but does not have finite second derivatives; see [8], Remark on page 239.

(ii) Is often satisfied with $\hat{\Theta} = \Theta$, but accomodates also projections sets which do not cover Θ, but only certain admissible values $\hat{\Theta} \subsetneq \Theta$. As an extreme case, this allows to use the present framework to check that a fixed projection does not induce spurious attractors on the boundary of $\hat{\Theta}$. Notice also that the function $H(\theta, x)$ and the corresponding mean field $h(\theta)$ need only be defined for values $\theta \in \hat{\Theta}$.

(iii) Will be replaced with a stricter drift in Theorem 2.8, where w is not required to diverge on the boundary $\partial\hat{\Theta}$.

(iv) Is satisfied trivially by the choices $\theta_i^{\mathrm{proj}} := \theta_{i-1}$ and $\theta_i^{\mathrm{proj}} := \Pi_{\mathcal{R}_i}(\theta_i^*)$, if the projection sets are defined as the level sets of the Lyapunov function, that is $\mathcal{R}_i := \mathcal{W}_{M_i}$ for some $M_i > 0$. In the Markovian case, the projections are assumed to satisfy an additional continuity condition; see Theorem 3.3.

(v) Involves in practice a sequence that grows at most at a rate $\xi_i := i \vee 1$, with some power $\alpha_w \in [0, 1)$. The sequence ξ_i plays a central role also in controlling the ergodicity rate of the Markov chain in $\mathcal{R}_i$; see Remark 3.2.

Hereafter, we denote the 'centred' version of H as $\bar{H}(\theta, x) := H(\theta, x) - h(\theta)$. For the stability results, we shall introduce the following general condition on the noise sequence. In general terms, it is related to the rate at which $\{\theta_i\}$ may approach $\partial\hat{\Theta}$ in relation to the growth of $|H(\theta, x)|$ and the loss of ergodicity of $\{P_\theta\}$. Establishing practical and realistic conditions under which this assumption holds will be the topic of Section 3.

Condition 2.3. *For any $(\theta, x) \in \mathcal{R}_0 \times \mathsf{X}$ it holds that*

(i) $\mathbb{P}_{\theta,x}\Big(\lim_{i\to\infty} \Gamma_{i+1}\big|\nabla w(\theta_i)\big| \cdot \big|H(\theta_i, X_{i+1})\big| = 0\Big) = 1,$

(ii) $\mathbb{E}_{\theta,x}\left[\sum_{i=0}^{\infty} \Gamma_{i+1}^2 \big|H(\theta_i, X_{i+1})\big|^2\right] < \infty,$

(iii) $\mathbb{E}_{\theta,x}\left[\sup_{k\geq 0}\left|\sum_{i=0}^{k} \Gamma_{i+1}\big\langle \nabla w(\theta_i), \bar{H}(\theta_i, X_{i+1})\big\rangle\right|\right] < \infty.$

In what follows, we shall focus on a single condition implying Condition 2.3(i) and (ii). It is slightly more stringent, but more convenient to check in practice.

Lemma 2.4. *Suppose Condition 2.1 holds and*

$$\mathbb{E}_{\theta,x}\left[\sum_{i=0}^{\infty} \Gamma_{i+1}^2 \xi_i^{2\alpha_w} \big|H(\theta_i, X_{i+1})\big|^2\right] < \infty. \tag{2.1}$$

Then, Condition 2.3(i) and (ii) hold.

Proof. Note first that Condition 2.3(ii) holds trivially, because $\xi_i^{2\alpha_w} \geq 1$. For Condition 2.3(i), consider

$$\mathbb{E}_{\theta,x}\left[\sum_{i=0}^{\infty} \big(\Gamma_{i+1}\big|\nabla w(\theta_i)\big| \cdot \big|H(\theta_i, X_{i+1})\big|\big)^2\right] \leq c^2 \mathbb{E}_{\theta,x}\left[\sum_{i=0}^{\infty} \Gamma_{i+1}^2 \xi_i^{2\alpha_w} \big|H(\theta_i, X_{i+1})\big|^2\right]. \qquad \square$$

2.1. Unbounded Lyapunov function

When $\lim_{\theta\to\partial\hat{\Theta}} w(\theta) = \infty$, it is enough to show that the sequence $w(\theta_i)$ is bounded in order to ensure the stability of θ_i.

Theorem 2.5. *Assume Conditions 2.1 and 2.3 hold. Then, for any $(\theta, x) \in \mathcal{R}_0 \times \mathsf{X}$*

$$\mathbb{P}_{\theta,x}\Big(\limsup_{i\to\infty} w(\theta_i) < \infty\Big) = 1.$$

Proof. To show the $\mathbb{P}_{\theta,x}$-a.s. boundedness of $\{w(\theta_i)\}$ we fix $(\theta, x) \in \mathcal{R}_0 \times \mathsf{X}$ and introduce the following quantities. Let $M_0 < M_1 < \cdots < M_n \to \infty$ be an increasing sequence tending to infinity and consider the level sets $\mathcal{W}_{M_i} \subset \Theta$. We assume that M_0 is chosen large enough so that $\theta_0 = \theta \in \mathcal{W}_{M_0}$. For any $n \geq 0$, we define the first exit time of θ_i from the level set $\mathcal{W}_{M_n}$ as

$$\sigma_n := \inf\{i \geq 0\colon \theta_i \notin \mathcal{W}_{M_n}\},$$

with the usual convention that $\inf\{\varnothing\} = \infty$. For any $n \geq 0$, we define the time following the last exit of θ_i from $\mathcal{W}_{M_0}$ before σ_n as

$$\tau_n := 1 + \sup\{i \leq \sigma_n \colon \theta_i \in \mathcal{W}_{M_0}\},$$

which is finite at least whenever σ_n is finite by our assumption that $\theta_0 \in \mathcal{W}_{M_0}$. With these definitions, the claim holds once we show that $\lim_{n\to\infty} \mathbb{P}_{\theta,x}(\sigma_n < \infty) = 0$.

To begin with, define for $n \geq 1$ the following sets characterising the jumps out of $\mathcal{W}_{M_0}$

$$D_n := \left\{ \mathbb{I}\{\tau_n < \infty\}\big[w(\theta_{\tau_n}) - w(\theta_{\tau_n - 1})\big] \leq \frac{M_n - M_0}{2} \right\}.$$

We first show that $\lim_{n\to\infty} \mathbb{P}_{\theta,x}(D_n) = 1$. Clearly

$$\tilde{D}_n := \left\{ \sup_{i\geq 0}\big[w(\theta_{i+1}) - w(\theta_i)\big] \leq \frac{M_n - M_0}{2} \right\} \subset D_n \tag{2.2}$$

and since $M_n \to \infty$, one has $\{\sup_{i\geq 0}[w(\theta_{i+1}) - w(\theta_i)] < \infty\} = \bigcup_{n=1}^{\infty} \tilde{D}_n$. Lemma 2.6 shows that $1 = \mathbb{P}_{\theta,x}(\bigcup_{n=1}^{\infty} \tilde{D}_n) = \lim_{n\to\infty} \mathbb{P}_{\theta,x}(\tilde{D}_n) \leq \lim_{n\to\infty} \mathbb{P}_{\theta,x}(D_n)$ because $\tilde{D}_n$ is an increasing sequence and by (2.2), respectively.

Now, it remains to focus on proving that

$$\lim_{n\to\infty} \mathbb{P}_{\theta,x}\big(D_n \cap \{\sigma_n < \infty\}\big) = 0.$$

In order to achieve this observe first that $w(\theta_{\sigma_n}) - w(\theta_{\tau_n - 1}) \geq M_n - M_0$ on $\{\sigma_n < \infty\}$, implying that on $D_n \cap \{\sigma_n < \infty\}$,

$$w(\theta_{\sigma_n}) - w(\theta_{\tau_n}) = w(\theta_{\sigma_n}) - w(\theta_{\tau_n - 1}) - \big[w(\theta_{\tau_n}) - w(\theta_{\tau_n - 1})\big] \geq \frac{M_n - M_0}{2}.$$

This allows us to deduce the following bound

$$\begin{aligned}
\mathbb{P}_{\theta,x}\big(D_n \cap \{\sigma_n < \infty\}\big) &= \mathbb{E}_{\theta,x}\big[\mathbb{I}\big\{D_n \cap \{\sigma_n < \infty\}\big\}\big] \\
&\leq \mathbb{E}_{\theta,x}\left[\mathbb{I}\big\{D_n \cap \{\sigma_n < \infty\}\big\} \frac{w(\theta_{\sigma_n}) - w(\theta_{\tau_n})}{(1/2)(M_n - M_0)}\right] \\
&\leq \frac{2}{M_n - M_0} \mathbb{E}_{\theta,x}\big[\mathbb{I}\{\sigma_n < \infty\}\big[w(\theta_{\sigma_n}) - w(\theta_{\tau_n})\big]\big].
\end{aligned}$$

Since $M_n \to \infty$, the proof will be finished once we show that

$$\sup_{n\geq 0} \mathbb{E}_{\theta,x}\big[\mathbb{I}\{\sigma_n < \infty\}\big[w(\theta_{\sigma_n}) - w(\theta_{\tau_n})\big]\big] < \infty. \tag{2.3}$$

Thanks to Condition 2.1(iv), we have for any $i \geq 0$ that $w(\theta_{i+1}) \leq w(\theta^*_{i+1})$ and consequently

$$\begin{aligned}
w(\theta_{i+1}) - w(\theta_i) &\leq \Gamma_{i+1}\big\langle \nabla w(\theta_i), h(\theta_i)\big\rangle \\
&\quad + \Gamma_{i+1}\big\langle \nabla w(\theta_i), \bar{H}(\theta_i, X_{i+1})\big\rangle + \Gamma_{i+1}^2 \frac{C_w}{2} \big|H(\theta_i, X_{i+1})\big|^2.
\end{aligned}$$

So in particular, since $\langle \nabla w(\theta_i), h(\theta_i)\rangle \le 0$ whenever $\theta_i \in \mathcal{W}^c_{M_0}$,

$$\begin{aligned}
&\mathbb{I}\{\sigma_n < \infty\}\big[w(\theta_{\sigma_n}) - w(\theta_{\tau_n})\big] \\
&\quad = \mathbb{I}\{\sigma_n < \infty\} \sum_{i=\tau_n}^{\sigma_n - 1} \big[w(\theta_{i+1}) - w(\theta_i)\big] \\
&\quad \le \mathbb{I}\{\sigma_n < \infty\} \left(\sum_{i=\tau_n}^{\sigma_n - 1} \Gamma_{i+1} \big\langle \nabla w(\theta_i), \bar{H}(\theta_i, X_{i+1}) \big\rangle + \Gamma_{i+1}^2 \frac{C_w}{2} \big| H(\theta_i, X_{i+1}) \big|^2 \right).
\end{aligned}$$

Recall the following estimate for partial sums

$$\left| \sum_{i=j}^{k} a_i \right| = \left| \sum_{i=0}^{k} a_i - \sum_{i=0}^{j-1} a_i \right| \le \left| \sum_{i=0}^{k} a_i \right| + \left| \sum_{i=0}^{j-1} a_i \right| \le 2 \sup_{k \ge 0} \left| \sum_{i=0}^{k} a_i \right|, \tag{2.4}$$

implying in our case that

$$\begin{aligned}
&\frac{1}{2}\mathbb{I}\{\sigma_n < \infty\}\big[w(\theta_{\sigma_n}) - w(\theta_{\tau_n})\big] \\
&\quad \le \mathbb{I}\{\sigma_n < \infty\} \left(\sup_{k \ge 0} \left| \sum_{i=0}^{k} \Gamma_{i+1} \big\langle \nabla w(\theta_i), \bar{H}(\theta_i, X_{i+1}) \big\rangle \right| + \sum_{i=0}^{\infty} \Gamma_{i+1}^2 \frac{C_w}{2} \big| H(\theta_i, X_{i+1}) \big|^2 \right).
\end{aligned}$$

Now, Condition 2.3(ii) and (iii) imply (2.3) allowing us to conclude. □

Lemma 2.6. *Under Condition 2.3 we have, $\mathbb{P}_{\theta,x}$-almost surely*

$$\limsup_{i \to \infty} \big[w(\theta_{i+1}) - w(\theta_i)\big] \le 0, \tag{2.5}$$

$$\sup_{i \ge 0} \big[w(\theta_{i+1}) - w(\theta_i)\big] < \infty. \tag{2.6}$$

Proof. We first prove that $\lim_{i\to\infty} |w(\theta^*_{i+1}) - w(\theta_i)| = 0$, $\mathbb{P}_{\theta,x}$-a.s. By a Taylor expansion, we get

$$\big|w\big(\theta^*_{i+1}\big) - w(\theta_i)\big| \le \big|\nabla w(\theta_i)\big| \cdot \big|\Gamma_{i+1} H(\theta_i, X_{i+1})\big| + \Gamma_{i+1}^2 C_w \big|H(\theta_i, X_{i+1})\big|^2.$$

The terms on the right converge to zero $\mathbb{P}_{\theta,x}$-a.s. by Condition 2.3(i) and (ii), respectively. Now, (2.5) follows since by Condition 2.1(iv) $w(\theta_{i+1}) - w(\theta_i) \le w(\theta^*_{i+1}) - w(\theta_i)$. We conclude by noting that (2.6) follows directly from (2.5). □

2.2. Bounded Lyapunov function

In the previous section, the Lyapunov function satisfied $\lim_{\theta \to \partial\hat{\Theta}} w(\theta) = \infty$. If this is not the case, we need to replace Condition 2.1(iii) with a more stringent condition quantifying the drift

outside $\mathcal{W}_{M_0}$, while not requiring $\lim_{\theta\to\partial\hat{\Theta}} w(\theta)=\infty$.

Condition 2.7. *The Lyapunov function and the step size sequence satisfy*

$$\delta_i := \inf_{\theta\in\mathcal{R}_i\setminus\mathcal{W}_{M_0}} -\big\langle\nabla w(\theta), h(\theta)\big\rangle > 0 \quad \textit{and} \quad \sum_{i=1}^{\infty}\Gamma_i\delta_i=\infty \qquad \mathbb{P}_{\theta,x}\textit{-almost surely.}$$

Theorem 2.8. *Assume Conditions 2.1, 2.3 and 2.7 hold, and in addition that the following condition on the noise holds*

$$\lim_{m\to\infty}\sup_{k>m}\left|\sum_{i=m}^{k}\Gamma_{i+1}\big\langle\nabla w(\theta_i),\bar{H}(\theta_i,X_{i+1})\big\rangle\right|=0. \tag{2.7}$$

Then for any $M>M_0$, the tails of the trajectories of $\{\theta_i\}$ are eventually contained within $\mathcal{W}_M$ $\mathbb{P}_{\theta,x}$-a.s., that is,

$$\mathbb{P}_{\theta,x}\Bigg(\bigcup_{m\geq 0}\bigcap_{n\geq m}\{\theta_n\in\mathcal{W}_M\}\Bigg)=1.$$

Proof. We first show that θ_n must visit $\mathcal{W}_{M_0}$ infinitely often $\mathbb{P}_{\theta,x}$-a.s., in other words

$$\mathbb{P}_{\theta,x}\Bigg(\bigcup_{m\geq 1}\bigcap_{n\geq m}\{\theta_n\notin\mathcal{W}_{M_0}\}\Bigg)=0. \tag{2.8}$$

For any $m\geq 0$, we define the hitting times $\kappa_m:=\inf\{i>m\colon\ \theta_i\in\mathcal{W}_{M_0}\}$ and notice that

$$\bigcup_{m\geq 1}\bigcap_{n\geq m}\{\theta_n\notin\mathcal{W}_{M_0}\}=\bigcup_{m\geq 1}\{\theta_m\notin\mathcal{W}_{M_0}\}\cap\{\kappa_m=\infty\}.$$

Recall that for any $i\geq 0$

$$\begin{aligned} w(\theta_{i+1})-w(\theta_i)&\leq\Gamma_{i+1}\big\langle\nabla w(\theta_i),h(\theta_i)\big\rangle\\ &\quad+\Gamma_{i+1}\big\langle\nabla w(\theta_i),\bar{H}(\theta_i,X_{i+1})\big\rangle+\Gamma_{i+1}^2\frac{C_w}{2}\big|H(\theta_i,X_{i+1})\big|^2. \end{aligned}$$

So in particular, and thanks to Condition 2.7, for $n>m$

$$\begin{aligned} &\mathbb{I}\{\theta_m\notin\mathcal{W}_{M_0}\}\big[w(\theta_{n\wedge\kappa_m})-w(\theta_m)\big]\\ &\quad=\mathbb{I}\{\theta_m\notin\mathcal{W}_{M_0}\}\sum_{i=m}^{(n\wedge\kappa_m)-1}\mathbb{I}\{\theta_i\notin\mathcal{W}_{M_0}\}\big[w(\theta_{i+1})-w(\theta_i)\big]\\ &\quad\leq\mathbb{I}\{\theta_m\notin\mathcal{W}_{M_0}\}\sum_{i=m}^{(n\wedge\kappa_m)-1}\Gamma_{i+1}\bigg[-\delta_i+\big\langle\nabla w(\theta_i),\bar{H}(\theta_i,X_{i+1})\big\rangle+\Gamma_{i+1}\frac{C_w}{2}\big|H(\theta_i,X_{i+1})\big|^2\bigg]. \end{aligned}$$

From this, we obtain the following inequality holding $\mathbb{P}_{\theta,x}$-a.s. on $\{\theta_m \notin \mathcal{W}_{M_0}\}$ for any $n > m$

$$\begin{aligned}\mathbb{E}_{\theta,x}\left[\mathbb{I}\{\kappa_m = \infty\}\sum_{i=m}^{n-1}\Gamma_{i+1}\delta_i \Big| \mathcal{F}_m\right] - w(\theta_m) \\ \leq \mathbb{E}_{\theta,x}\left[\mathbb{I}\{\kappa_m = \infty\}\sum_{i=m}^{n-1}\Gamma_{i+1}\left\langle \nabla w(\theta_i), \bar{H}(\theta_i, X_{i+1})\right\rangle \right. \\ \left. + \Gamma_{i+1}^2\frac{C_w}{2}\left|H(\theta_i, X_{i+1})\right|^2 \Big| \mathcal{F}_m\right].\end{aligned} \tag{2.9}$$

Using this inequality, we shall see that for any $m > 0$

$$\mathbb{P}_{\theta,x}\left(\{\theta_m \notin \mathcal{W}_{M_0}\} \cap \{\kappa_m = \infty\}\right) = 0. \tag{2.10}$$

Suppose the contrary, $\mathbb{P}_{\theta,x}(\{\theta_m \notin \mathcal{W}_{M_0}\} \cap \{\kappa_m = \infty\}) > 0$. Then, because of Condition 2.7, we observe that the conditional expectation on the left hand side of (2.9) necessarily tends to infinity almost surely as $n \to \infty$. Denote then the conditional expectation on the right hand side of (2.9) by $E_{\theta,x}^{(m,n)}$. As in the proof of Theorem 2.5, we have the following upper bound

$$\mathbb{E}_{\theta,x}\left[E_{\theta,x}^{(m,n)}\right] \leq \mathbb{E}_{\theta,x}\left[\sup_{k\geq 0}\left|\sum_{i=0}^{k}\Gamma_{i+1}\left\langle \nabla w(\theta_i), \bar{H}(\theta_i, X_{i+1})\right\rangle\right| + \sum_{i=0}^{\infty}\Gamma_{i+1}^2\frac{C_w}{2}\left|H(\theta_i, X_{i+1})\right|^2\right],$$

which is finite by Condition 2.3 and independent of m and n. By letting $n \to \infty$ we end up with a contradiction, unless (2.10) holds. Consequently, the event

$$\bigcup_{m\geq 1}\{\theta_m \notin \mathcal{W}_{M_0}\} \cap \{\kappa_m = \infty\}$$

has null probability and we obtain (2.8).

We now show that for any fixed $M > M_0$

$$\mathbb{P}_{\theta,x}\left(\bigcup_{m\geq 0}\bigcap_{n\geq m}\{\theta_n \in \mathcal{W}_M\}\right) = 1.$$

We are going to apply Lemma 2.9 below with $\delta = M - M_0 > 0$ to the events

$$A_m = \{\theta_m \in \mathcal{W}_{M_0}\} \cap \bigcup_{k>m}\{\theta_k \notin \mathcal{W}_M\},$$

and denote

$$B_m := \{\theta_m \in \mathcal{W}_{M_0}\} \setminus A_m = \{\theta_m \in \mathcal{W}_{M_0}\} \cap \bigcap_{k>m}\{\theta_k \in \mathcal{W}_M\}.$$

We may write

$$\bigcap_{n\geq 1}\bigcup_{m\geq n}\{\theta_m \in \mathcal{W}_{M_0}\} = \bigcap_{n\geq 1}\bigcup_{m\geq n} A_m \cup B_m$$
$$= \bigcap_{n\geq 1}\biggl[\biggl(\bigcup_{m\geq n} A_m\biggr)\cup\biggl(\bigcup_{m\geq n} B_m\biggr)\biggr].$$

Now, since $\bigcup_{m\geq n} A_m$ and $\bigcup_{m\geq n} B_m$ are both decreasing events with respect to $n\to\infty$, we have

$$1 = \lim_{n\to\infty}\mathbb{P}_{\theta,x}\biggl(\bigcup_{m\geq n}\{\theta_m\in\mathcal{W}_{M_0}\}\biggr)$$
$$= \lim_{n\to\infty}\biggl[\mathbb{P}_{\theta,x}\biggl(\bigcup_{m\geq n} A_m\biggr)+\mathbb{P}_{\theta,x}\biggl(\bigcup_{m\geq n} B_m\biggr)-\mathbb{P}_{\theta,x}\biggl(\bigcup_{m\geq n} A_m\cap\bigcup_{m\geq n} B_m\biggr)\biggr].$$

By Lemma 2.9, $\lim_{n\to\infty}\mathbb{P}_{\theta,x}(\bigcup_{m\geq n} A_m)=0$, so we end up with $\lim_{n\to\infty}\mathbb{P}_{\theta,x}(\bigcup_{m\geq n} B_m)=1$, implying the claim. □

Lemma 2.9. *Assume the conditions of Theorem 2.8, let $\delta>0$ and denote*

$$A_m := \{\theta_m\in\mathcal{W}_{M_0}\}\cap\bigcup_{k>m}\{\theta_k\notin\mathcal{W}_{M_0+\delta}\}.$$

Then, $\lim_{n\to\infty}\mathbb{P}_{\theta,x}(\bigcup_{m\geq n} A_m)=0$.

Proof. Define the random times $\sigma_m := \inf\{i>m\colon \theta_i\notin\mathcal{W}_{M_0+\delta}\}$ and $\tau_m := \sup\{i\in[m,\sigma_m)\colon \theta_i\in\mathcal{W}_{M_0}\}+1$, both finite on A_m. Recall that on $\{\theta_i\in\mathcal{W}_{M_0}^c\}$ we have

$$w(\theta_{i+1})-w(\theta_i)\leq\Gamma_{i+1}\bigl\langle\nabla w(\theta_i),\bar{H}(\theta_i,X_{i+1})\bigr\rangle+\Gamma_{i+1}^2\frac{C_w}{2}\bigl|H(\theta_i,X_{i+1})\bigr|^2,$$

so on A_m we may bound

$$\begin{aligned} w(\theta_{\sigma_m})-w(\theta_{\tau_m}) &\leq \sum_{i=\tau_m}^{\sigma_m-1}\Gamma_{i+1}\bigl\langle\nabla w(\theta_i),\bar{H}(\theta_i,X_{i+1})\bigr\rangle+\Gamma_{i+1}^2\frac{C_w}{2}\bigl|H(\theta_i,X_{i+1})\bigr|^2 \\ &\leq 2\sup_{k>m}\biggl|\sum_{i=m}^{k}\Gamma_{i+1}\bigl\langle\nabla w(\theta_i),\bar{H}(\theta_i,X_{i+1})\bigr\rangle\biggr|+\sum_{i=m}^{\infty}\Gamma_{i+1}^2\frac{C_w}{2}\bigl|H(\theta_i,X_{i+1})\bigr|^2 \\ &=: C_m \end{aligned}$$

by a similar argument as in (2.4). On A_m one clearly has $w(\theta_{\sigma_m})-w(\theta_{\tau_m-1})>\delta$, implying that $C_m+w(\theta_{\tau_m})-w(\theta_{\tau_m-1})>\delta$. We deduce that

$$\tilde{A}_m := \Bigl\{C_m+\sup_{i\geq m}\bigl[w(\theta_{i+1})-w(\theta_i)\bigr]>\delta\Bigr\}\supset A_m.$$

The sets $\tilde{A}_m$ are clearly decreasing with respect to m and $\lim_{m\to\infty} \mathbb{P}_{\theta,x}(\tilde{A}_m) = 0$ by Lemma 2.6 and because Condition 2.3(ii) and (2.7) imply $\lim_{m\to\infty} C_m = 0$. This concludes the proof, because $\bigcup_{m\geq n} A_m \subset \bigcup_{m\geq n} \tilde{A}_m = \tilde{A}_n$. □

3. Verifying noise conditions

The aim of this section is to provide verifiable conditions which will imply the conditions of the stability theorems in Section 2. We proceed progressively and start by a general result in Theorem 3.3 which ensures both Condition 2.3 and that in (2.7) hold given a set of abstract conditions involving some expectations as well as properties of the solutions of the Poisson equation.

Condition 3.1, required in Theorem 3.3, shall be verified in detail below for a family of geometrically ergodic Markov kernels. In Section 3.1, we first gather general known results related to Condition 3.1(ii) and (iii). In Section 3.2, we consider the case where the mapping $\theta \to P_\theta$ is Hölder continuous, which allows us to establish Condition 3.1(iv). In Section 3.3, we consider the case where the aforementioned Hölder continuity may not hold, and a continuity is enforced by using a random step size sequence, allowing us to recover Condition 3.1(iv) in such situations.

Condition 3.1. *Condition 2.1 holds with constants $(\xi_i)_{i\geq 0}$ and $\alpha_w \in (0,\infty)$. For all $\theta \in \hat{\Theta}$, the solution $g_\theta : \mathsf{X} \to \Theta$ to the Poisson equation $g_\theta(x) - P_\theta g_\theta(x) \equiv \bar{H}(\theta, x)$ exists and for all $i \geq 0$ the step size Γ_{i+1} is independent of $\mathcal{F}_i$ and X_{i+1}. Moreover, there exist a measurable function $V : \mathsf{X} \to [1,\infty)$ and constants $c < \infty$, $\beta_H, \beta_g \in [0, 1/2]$ and $\alpha_g, \alpha_H, \alpha_V \in [0,\infty)$ such that for all $(\theta, x) \in \mathcal{R}_0 \times \mathsf{X}$*

(i) $\sup_{\theta\in\mathcal{R}_i} \big|H(\theta,x)\big| \leq c\xi_i^{\alpha_H} V^{\beta_H}(x)$,

(ii) $\mathbb{E}_{\theta,x}\big[V(X_i)\big] \leq c\xi_i^{\alpha_V} V(x)$,

(iii) $\sup_{\theta\in\mathcal{R}_i} \big[\big|g_\theta(x)\big| + \big|P_\theta g_\theta(x)\big|\big] \leq c\xi_i^{\alpha_g} V^{\beta_g}(x)$,

(iv) $\sum_{i=1}^{\infty} \mathbb{E}[\Gamma_{i+1}]\xi_i^{\alpha_w} \mathbb{E}_{\theta,x}\big[\big|P_{\theta_i} g_{\theta_i}(X_i) - P_{\theta_{i-1}} g_{\theta_{i-1}}(X_i)\big|\big] < \infty$,

(v) $\sum_{i=1}^{\infty} \mathbb{E}\big[\Gamma_i^2\big]\xi_i^{2\alpha_w + 2((\alpha_H+\beta_H\alpha_V)\vee(\alpha_g+\beta_g\alpha_V))} < \infty$,

(vi) $\sum_{i=1}^{\infty} \mathbb{E}[\Gamma_{i+1}\Gamma_i]\xi_i^{\alpha_H+\alpha_g+(\beta_H+\beta_g)\alpha_V} < \infty$,

(vii) $\sum_{i=1}^{\infty} \big|\mathbb{E}[\Gamma_{i+1} - \Gamma_i]\big|\xi_i^{\alpha_w+\alpha_g+\beta_g\alpha_V} < \infty$,

where we write $\mathbb{E} := \mathbb{E}_{\theta,x}$ whenever the expectation does not depend on θ and x.

Remark 3.2. These assumptions call for various comments of practical relevance to the actual implementation of the algorithm with expanding projections. Once $H(\cdot,\cdot)$ and $\{P_\theta\}_{\theta\in\Theta}$ are chosen the user is left with the choice of $(\xi_i)_{i\geq 0}$ and $(\Gamma_i)_{i\geq 0}$, which must in particular satisfy the summability conditions above. For the purpose of efficiency we would like $(\xi_i)_{i\geq 0}$ to grow as fast as possible, as we may otherwise slow convergence down. A common choice for the step-size sequence is $\Gamma_i = ci^{-\eta}$ for some constants $c \in (0,\infty)$ and $\eta \in (1/2, 1]$ – this implies a required condition to establish convergence. The sequence $(\xi_i)_{i\geq 0}$ is determined by the user through the choice of the sequence of reprojection sets $(\mathcal{R}_i)_{i\geq 0}$ and we point out that the constants α_H, α_V and α_g typically depend on that choice (whereas β_H and β_g typically do not). We show how these constants can be obtained from the properties of $\{P_\theta\}_{\theta\in\Theta}$ in Sections 3.1–3.3. Now if $(\xi_i)_{i\geq 0}$ is increasing at a rate slower than any power sequence, for example of the order $\log i$ or $i^{(\log i)^{-p}}$ for some $p \in (0,1)$, then it is easy to see that the summability conditions (v)–(vii) are always satisfied. In the situation where $\xi_i = i^p$ for $p \in (0,1]$, then the conditions (v)–(vii) require stricter assumptions on η and the constants $\alpha_H, \alpha_V, \alpha_g, \beta_H$ and β_g which may not be satisfiable. We however point out a possible sub-optimality of the results stated above. Indeed, in order to simplify presentation we have decided to quantify the growth of the various quantities involved in the algorithm in terms of powers of $(\xi_i)_{i\geq 0}$ only, whereas other scales may be possible, such as $\log(\xi_i)$, in which case some of the constants α_H, α_V or α_g may be taken arbitrarily small in the statement above. It is also possible to revisit our proofs with such more precise estimates and obtain a set of weaker assumptions.

In practice, the conditions (iii) and (iv) add more requirements which are inter-related with (v)–(vii); Propositions 3.17 and 3.19 summarise the conditions when $\theta \mapsto P_\theta$ admits a Hölder-continuity, and when a random step size sequence is used to satisfy (iv), respectively. Appendix D contains a summary of the related constants.

Theorem 3.3. *Suppose Conditions 2.1 and 3.1 hold and for all $i \geq 0$ the projections satisfy $|\theta_{i+1} - \theta_i| \leq |\theta^*_{i+1} - \theta_i|$. Then, for all $(\theta, x) \in \mathcal{R}_0 \times \mathsf{X}$,*

$$\mathbb{E}_{\theta,x}\left[\sum_{i=0}^{\infty} \Gamma_{i+1}^2 \xi_i^{2\alpha_w} \left|H(\theta_i, X_{i+1})\right|^2\right] < \infty, \tag{3.1}$$

$$\lim_{m\to\infty} \mathbb{E}_{\theta,x}\left[\sup_{n\geq m}\left|\sum_{i=m}^{n} \Gamma_{i+1}\left\langle \nabla w(\theta_i), \bar{H}(\theta_i, X_{i+1})\right\rangle\right|\right] = 0. \tag{3.2}$$

Proof. Throughout the proof, C denotes a constant which may have a different value upon each appearance. For (3.1), we may use Condition 3.1(i) and (ii) with Jensen's inequality to obtain

$$\begin{aligned}\mathbb{E}_{\theta,x}\left[\sum_{i=0}^{\infty} \Gamma_{i+1}^2 \xi_i^{2\alpha_w} \left|H(\theta_i, X_{i+1})\right|^2\right] &\leq C\sum_{i=0}^{\infty} \mathbb{E}\left[\Gamma_{i+1}^2\right] \xi_i^{2\alpha_w+2\alpha_H} \mathbb{E}_{\theta,x}\left[V^{2\beta_H}(X_{i+1})\right] \\ &\leq C V^{2\beta_H}(x) \sum_{i=0}^{\infty} \mathbb{E}\left[\Gamma_{i+1}^2\right] \xi_i^{2\alpha_w+2\alpha_H+2\beta_H\alpha_V},\end{aligned}$$

where the sum converges by Condition 3.1(v).

Consider then (3.2), and denote the partial sums for $n \geq m \geq 1$ as

$$A_{m,n} := \sum_{i=m}^{n} \Gamma_{i+1} \langle \nabla w(\theta_i), \bar{H}(\theta_i, X_{i+1}) \rangle.$$

Since $\bar{H}(\theta_i, X_{i+1}) = g_{\theta_i}(X_{i+1}) - P_{\theta_i} g_{\theta_i}(X_{i+1})$, we may write

$$\begin{aligned}
&\Gamma_{i+1} \langle \nabla w(\theta_i), \bar{H}(\theta_i, X_{i+1}) \rangle \\
&\quad = \Gamma_{i+1} \langle \nabla w(\theta_i), g_{\theta_i}(X_{i+1}) - P_{\theta_i} g_{\theta_i}(X_i) \rangle \\
&\qquad + \Gamma_{i+1} \langle \nabla w(\theta_i), P_{\theta_i} g_{\theta_i}(X_i) - P_{\theta_{i-1}} g_{\theta_{i-1}}(X_i) \rangle \\
&\qquad + \Gamma_{i+1} \langle \nabla w(\theta_i), P_{\theta_{i-1}} g_{\theta_{i-1}}(X_i) - P_{\theta_i} g_{\theta_i}(X_{i+1}) \rangle,
\end{aligned}$$

where the last term can be written as

$$\begin{aligned}
&\Gamma_{i+1} \langle \nabla w(\theta_i), P_{\theta_{i-1}} g_{\theta_{i-1}}(X_i) - P_{\theta_i} g_{\theta_i}(X_{i+1}) \rangle \\
&\quad = \Gamma_{i+1} \langle \nabla w(\theta_i) - \nabla w(\theta_{i-1}), P_{\theta_{i-1}} g_{\theta_{i-1}}(X_i) \rangle \\
&\qquad + \Gamma_i \langle \nabla w(\theta_{i-1}), P_{\theta_{i-1}} g_{\theta_{i-1}}(X_i) \rangle - \Gamma_{i+1} \langle \nabla w(\theta_i), P_{\theta_i} g_{\theta_i}(X_{i+1}) \rangle \\
&\qquad + (\Gamma_{i+1} - \Gamma_i) \langle \nabla w(\theta_{i-1}), P_{\theta_{i-1}} g_{\theta_{i-1}}(X_i) \rangle.
\end{aligned}$$

When summing up, the middle term on the right is telescoping, so in total we may write $A_{m,n} = \sum_{k=1}^{5} R^k_{m,n}$ where

$$\begin{aligned}
R^1_{m,n} &:= \sum_{i=m}^{n} \Gamma_{i+1} \langle \nabla w(\theta_i), g_{\theta_i}(X_{i+1}) - P_{\theta_i} g_{\theta_i}(X_i) \rangle, \\
R^2_{m,n} &:= \sum_{i=m}^{n} \Gamma_{i+1} \langle \nabla w(\theta_i), P_{\theta_i} g_{\theta_i}(X_i) - P_{\theta_{i-1}} g_{\theta_{i-1}}(X_i) \rangle, \\
R^3_{m,n} &:= \sum_{i=m}^{n} \Gamma_{i+1} \langle \nabla w(\theta_i) - \nabla w(\theta_{i-1}), P_{\theta_{i-1}} g_{\theta_{i-1}}(X_i) \rangle, \\
R^4_{m,n} &:= \Gamma_m \langle \nabla w(\theta_{m-1}), P_{\theta_{m-1}} g_{\theta_{m-1}}(X_m) \rangle - \Gamma_{n+1} \langle \nabla w(\theta_n), P_{\theta_n} g_{\theta_n}(X_{n+1}) \rangle, \\
R^5_{m,n} &:= \sum_{i=m}^{n} (\Gamma_{i+1} - \Gamma_i) \langle \nabla w(\theta_{i-1}), P_{\theta_{i-1}} g_{\theta_{i-1}}(X_i) \rangle.
\end{aligned}$$

We shall show that (3.2) holds for each of these five terms in turn, which is sufficient to yield the claim.

Notice that $\{R^1_{m,i}\}_{i=m}^n$ is a martingale with respect to the filtration $\{\mathcal{F}_i\}_{i=m}^n$, whence

$$\begin{aligned}\mathbb{E}_{\theta,x}\big[\big|R^1_{m,n}\big|^2\big] &= \sum_{i=m}^n \mathbb{E}_{\theta,x}\big[\Gamma^2_{i+1}\big|\big\langle \nabla w(\theta_i), g_{\theta_i}(X_{i+1}) - P_{\theta_i} g_{\theta_i}(X_i)\big\rangle\big|^2\big] \\ &\le C\sum_{i=m}^n \xi_i^{2\alpha_w}\mathbb{E}\big[\Gamma^2_{i+1}\big]\mathbb{E}_{\theta,x}\big[\big|g_{\theta_i}(X_{i+1})\big|^2 + \big|P_{\theta_i} g_{\theta_i}(X_i)\big|^2\big] \\ &\le C\sum_{i=m}^n \xi_i^{2\alpha_w+2\alpha_g}\mathbb{E}\big[\Gamma^2_{i+1}\big]\mathbb{E}_{\theta,x}\big[V^{2\beta_g}(X_{i+1}) + V^{2\beta_g}(X_i)\big] \\ &\le CV^{2\beta_g}(x)\sum_{i=m}^n \xi_{i+1}^{2\alpha_w+2\alpha_g+2\beta_g\alpha_V}\mathbb{E}\big[\Gamma^2_{i+1}\big],\end{aligned}$$

by the fact that Γ_{i+1} is independent of $\mathcal{F}_i$ and X_{i+1}, Condition 2.1(v), Condition 3.1(ii) and (iii). Now, Jensen's and Doob's inequality imply

$$\Big(\mathbb{E}_{\theta,x}\Big[\sup_{n\ge m}\big|R^1_{m,n}\big|\Big]\Big)^2 \le \mathbb{E}_{\theta,x}\Big[\sup_{n\ge m}\big|R^1_{m,n}\big|^2\Big] \le CV^{2\beta_g}(x)\sum_{i=m}^\infty \xi_{i+1}^{2\alpha_w+2\alpha_g+2\beta_g\alpha_V}\mathbb{E}\big[\Gamma^2_{i+1}\big].$$

This yields $\lim_{m\to\infty}\mathbb{E}_{\theta,x}[\sup_{n\ge m}|R^1_{m,n}|] = 0$, because the term on the right tends to zero as $m\to\infty$ by Condition 3.1(v).

For the second term $R^2_{m,n}$, we may simply write

$$\begin{aligned}\mathbb{E}_{\theta,x}\Big[\sup_{n\ge m}\big|R^2_{m,n}\big|\Big] &\le \mathbb{E}_{\theta,x}\Big[\sum_{i=m}^\infty\big|\Gamma_{i+1}\big\langle\nabla w(\theta_i), P_{\theta_i} g_{\theta_i}(X_i) - P_{\theta_{i-1}} g_{\theta_{i-1}}(X_i)\big\rangle\big|\Big] \\ &\le C\sum_{i=m}^\infty \xi_i^{\alpha_w}\mathbb{E}[\Gamma_{i+1}]\mathbb{E}_{\theta,x}\big[\big|P_{\theta_i} g_{\theta_i}(X_i) - P_{\theta_{i-1}} g_{\theta_{i-1}}(X_i)\big|\big],\end{aligned}$$

which converges to zero as $m\to\infty$ by Condition 3.1(iv).

Now we inspect $R^3_{m,n}$. First, since the Hessian is bounded as in Condition 2.1(i), we have

$$\begin{aligned}\big|\nabla w(\theta_i) - \nabla w(\theta_{i-1})\big| &\le C_w|\theta_i - \theta_{i-1}| \le C_w\big|\theta^*_i - \theta_{i-1}\big| = C_w\Gamma_i\big|H(\theta_{i-1}, X_i)\big| \\ &\le C_w\xi_i^{\alpha_H}\Gamma_i V^{\beta_H}(X_i),\end{aligned}$$

and consequently

$$\begin{aligned}\mathbb{E}_{\theta,x}\Big[\sup_{n\ge m}\big|R^3_{m,n}\big|\Big] &\le C\sum_{i=m}^\infty \mathbb{E}[\Gamma_{i+1}\Gamma_i]\xi_i^{\alpha_g+\alpha_H}\mathbb{E}_{\theta,x}\big[V^{\beta_g+\beta_H}(X_i)\big] \\ &\le CV^{\beta_g+\beta_H}(x)\sum_{i=m}^\infty \mathbb{E}[\Gamma_{i+1}\Gamma_i]\xi_i^{\alpha_g+\alpha_H+(\beta_g+\beta_H)\alpha_V},\end{aligned}$$

by Condition 3.1(i), (ii) and (iii). The claim follows for $R^3_{m,n}$ by Condition 3.1(vi).

Let us then focus on $R^4_{m,n}$. We have for any $i \geq m$

$$\left|\Gamma_i \left\langle \nabla w(\theta_{i-1}), P_{\theta_{i-1}} g_{\theta_{i-1}}(X_i) \right\rangle\right| \leq C\Gamma_i \xi_i^{\alpha_w+\alpha_g} V^{\beta_g}(X_i).$$

Now we have

$$\begin{aligned}\mathbb{E}_{\theta,x}\Big[\sup_{n\geq m}\big|R^4_{m,n}\big|^2\Big] &\leq C\sum_{i=m}^{\infty} \xi_i^{2\alpha_w+2\alpha_g}\mathbb{E}\big[\Gamma_i^2\big]\mathbb{E}_{\theta,x}\big[V^{2\beta_g}(X_i)\big] \\ &\leq CV^{2\beta_g}(x)\sum_{i=m}^{\infty} \xi_i^{2\alpha_w+2\alpha_g+2\beta_g\alpha_V}\mathbb{E}\big[\Gamma_i^2\big],\end{aligned}$$

so (3.2) holds for $R^4_{m,n}$ by Condition 3.1(v).

We shall apply Lemma 3.4 below for the last term $R^5_{m,n}$, with $Z_i := \Gamma_i$ and

$$B_{i-1} := \left\langle \nabla w(\theta_{i-1}), P_{\theta_{i-1}} g_{\theta_{i-1}}(X_i) \right\rangle \qquad \text{with } |B_{i-1}| \leq C\xi_{i-1}^{\alpha_w+\alpha_g} V^{\beta_g}(X_i).$$

By the independence of Γ_{i+1} and Γ_i, and because $\xi_{i+1} \geq \xi_i \geq \xi_{i-1}$, we easily establish the required bounds

$$\begin{aligned}&\sum_{i=1}^{\infty} \mathrm{Var}(\Gamma_{i+1}-\Gamma_i)\mathbb{E}_{\theta,x}\big[B_{i-1}^2\big] \leq CV^{2\beta_g}(x)\sum_{i=1}^{\infty}\mathbb{E}\big[\Gamma_i^2\big]\xi_i^{2\alpha_w+2\alpha_g+2\beta_g\alpha_V} < \infty, \\ &\sum_{i=1}^{\infty}\big|\mathbb{E}[\Gamma_{i+1}-\Gamma_i]\big|\mathbb{E}[|B_{i-1}|] \leq CV^{\beta_g}(x)\sum_{i=1}^{\infty}\big|\mathbb{E}[\Gamma_{i+1}-\Gamma_i]\big|\xi_i^{\alpha_w+\alpha_g+\beta_g\alpha_V} < \infty,\end{aligned}$$

by Condition 3.1(v) and (vii), respectively. □

Lemma 3.4. *Let $\{\mathcal{G}_i\}_{i\geq 0}$ be a filtration and for all $i \geq 0$ let B_i and Z_i be $\mathcal{G}_i$-adapted random variables so that Z_i is independent of $\mathcal{G}_{i-1}$ and*

$$\sum_{i=1}^{\infty}\mathrm{Var}(Z_{i+1}-Z_i)\mathbb{E}\big[B_{i-1}^2\big] < \infty \quad \textit{and} \quad \sum_{i=1}^{\infty}\big|\mathbb{E}[Z_{i+1}-Z_i]\big|\mathbb{E}[|B_{i-1}|] < \infty.$$

Then,

$$\lim_{m\to\infty}\mathbb{E}\left[\sup_{n\geq m}\left|\sum_{i=m}^{n}(Z_{i+1}-Z_i)B_{i-1}\right|\right] = 0.$$

Proof. Suppose for now that m is even and n odd and denote $m = 2\bar{m}$ and $n = 2\bar{n} + 1$. Write the sum

$$\sum_{i=m}^{n} (Z_{i+1} - Z_i) B_{i-1} = \sum_{j=\bar{m}}^{\bar{n}} (Z_{2j+1} - Z_{2j}) B_{2j-1} + \sum_{k=\bar{m}}^{\bar{n}} (Z_{2k+2} - Z_{2k+1}) B_{2k}. \tag{3.3}$$

We shall first show that the claim holds for the first term on the right. Denote $\bar{\mathcal{G}}_j = \mathcal{G}_{2j+1}$, $\bar{Z}_j = Z_{2j+1} - Z_{2j}$ and $\bar{B}_{j-1} = B_{2j-1}$. Observe that $\mathbb{E}[\bar{Z}_j | \bar{\mathcal{G}}_{j-1}] = \mathbb{E}[\bar{Z}_j]$ and write

$$\sum_{j=\bar{m}}^{\bar{n}} (Z_{2j+1} - Z_{2j}) B_{2j-1} = \sum_{j=\bar{m}}^{\bar{n}} \big(\bar{Z}_j - \mathbb{E}[\bar{Z}_j]\big) \bar{B}_{j-1} + \sum_{j=\bar{m}}^{\bar{n}} \mathbb{E}[\bar{Z}_j] \bar{B}_{j-1}.$$

Now, the first term on the right-hand side is a martingale with respect to $\bar{\mathcal{G}}_j$, and so by Doob's inequality and by assumption

$$\mathbb{E}\left[\sup_{\bar{n} \geq \bar{m}} \left(\sum_{j=\bar{m}}^{\bar{n}} \big(\bar{Z}_j - \mathbb{E}[\bar{Z}_j]\big) \bar{B}_{j-1} \right)^2 \right] \leq 4 \sum_{j=\bar{m}}^{\infty} \mathrm{Var}(\bar{Z}_j) \mathbb{E}\big[\bar{B}_{j-1}^2\big] \xrightarrow{\bar{m} \to \infty} 0.$$

For the second term, by assumption

$$\mathbb{E}\left[\sup_{\bar{n} \geq \bar{m}} \left| \sum_{j=\bar{m}}^{\bar{n}} \mathbb{E}[\bar{Z}_j] \bar{B}_{j-1} \right| \right] \leq \sum_{j=\bar{m}}^{\infty} \big|\mathbb{E}[\bar{Z}_j]\big| \mathbb{E}\big[|\bar{B}_{j-1}|\big] \xrightarrow{\bar{m} \to \infty} 0.$$

The same arguments apply also for the second term on the right-hand side of (3.3), and for any integers $m \geq n \geq 1$, by a change of the indices. □

3.1. Geometrically ergodic Markov kernels

In this section, we focus on the scenario where for any $\theta \in \Theta$ the kernel P_θ is geometrically ergodic. This condition is satisfied by numerous Markov chains of practical interest, see for example, [17,18,22] and references therein. This section gathers together standard results about the regularity of the solutions to the Poisson equation (see, e.g., [3,4]).

Throughout this section, suppose $V : \mathsf{X} \to [1, \infty)$ is a fixed measurable function. We shall denote the V-norm of a measurable function $f : \mathsf{X} \to \mathbb{R}^d$ by $\|f\|_V := \sup_x |f(x)|/V(x)$. We also assume that for each $\theta \in \hat{\Theta}$, the Markov kernel P_θ admits a unique invariant probability measure π_θ.

Condition 3.5. *For any $r \in (0, 1]$ and any $\theta \in \hat{\Theta}$, there exist constants $M_{\theta,r} \in [0, \infty)$ and $\rho_{\theta,r} \in (0, 1)$, such that for any function $\|f\|_{V^r} < \infty$*

$$\big| P_\theta^k(x, f) - \pi_\theta(f) \big| \leq V^r(x) \|f\|_{V^r} M_{\theta,r} \rho_{\theta,r}^k$$

for all $k \geq 0$ and all $x \in \mathsf{X}$.

Having Condition 3.5 one can bound the V^r-norm of the solutions of the Poisson equation, making the dependence on θ explicit. This result is a restatement of [3], Proposition 3, in quantitative form; we provide it here for the reader's convenience.

Proposition 3.6. *Assume Condition 3.5 holds. Then, for any function $\|f\|_{V^r} < \infty$, the functions $g_\theta : \mathsf{X} \to \mathbb{R}^d$ defined for all $\theta \in \hat{\Theta}$ by*

$$g_\theta(x) := \sum_{k=0}^{\infty} \big[P_\theta^k f(x) - \pi_\theta(f) \big]$$

exist, solve the Poisson equation $g_\theta(x) - P_\theta g_\theta(x) \equiv f(x) - \pi_\theta(f)$, and satisfy the bound

$$\|g_\theta\|_{V^r} \vee \|P_\theta g_\theta\|_{V^r} \le M_{\theta,r}(1-\rho_{\theta,r})^{-1} \|f\|_{V^r}. \tag{3.4}$$

Proof. It is evident that g_θ solves the Poisson equation whenever the sum converges. By the definition of g_θ and Condition 3.5, we have

$$\|g_\theta\|_{V^r} \le \sum_{k=0}^{\infty} \big\| P_\theta^k f - \pi_\theta(f) \big\|_{V^r} \le M_{\theta,r} \|f\|_{V^r} \sum_{k=0}^{\infty} \rho_{\theta,r}^k = M_{\theta,r}(1-\rho_{\theta,r})^{-1} \|f\|_{V^r}.$$

The same bound applies clearly also for $P_\theta g_\theta$, establishing (3.4). □

We also need the following simple lemma in order to establish Condition 3.1(ii).

Lemma 3.7. *Suppose that for all $i \ge 0$ there exist constants $\lambda_i \in [0,1)$ and $b_i \in [0,\infty)$ such that*

$$\sup_{\theta \in \mathcal{R}_i} P_\theta V(x) \le \lambda_i V(x) + b_i \qquad \text{for all } x \in \mathsf{X}, \tag{3.5}$$

and that both $(\lambda_i)_{i\ge 0}$ and $(b_i)_{i \ge 0}$ are non-decreasing. Then, for any $(\theta, x) \in \mathcal{R}_0 \times \mathsf{X}$ and $i \ge 0$, the bound $\mathbb{E}_{\theta,x}[V(X_{i+1})] \le (1-\lambda_i)^{-1}(b_i \vee V(x))$ holds.

Proof. By construction, for all $i \ge 1$ we have $\mathbb{E}_{\theta,x}[V(X_i)|\mathcal{F}_{i-1}] = P_{\theta_{i-1}} V(X_{i-1})$ and $\theta_{i-1} \in \mathcal{R}_{i-1}$, so we may use (3.5) iteratively to obtain

$$\mathbb{E}_{\theta,x}\big[V(X_{i+1})\big] \le \mathbb{E}_{\theta,x}\big[\lambda_i V(X_i) + b_i\big] \le \cdots \le \big(b_i \vee V(x)\big) \sum_{k=0}^{i} \lambda_i^k \le \frac{b_i \vee V(x)}{1-\lambda_i}.$$

□

Let us consider next a case where the ergodicity rates in each projection set $\mathcal{R}_i$ are controlled by the sequence ξ_i.

Condition 3.8. *Suppose Condition 3.5 holds with constants $M_{\theta,r}$, $\rho_{\theta,r}$ satisfying*

$$\sup_{\theta \in \mathcal{R}_i} M_{\theta,r} \le c_r \xi_i^{\alpha_M} \quad \text{and} \quad \sup_{\theta \in \mathcal{R}_i} (1-\rho_{\theta,r})^{-1} \le c_r \xi_i^{\alpha_\rho}$$

for some constants $\alpha_M, \alpha_\rho \in [0, \infty)$, *and a constant* $c_r \in [0, \infty)$ *depending only on* r.

Proposition 3.9. *If Condition 3.8 holds, then Condition 3.1(iii) holds with* $\alpha_g = \alpha_H + \alpha_M + \alpha_\rho$ *and* $\beta_g = \beta_H$.

Proof. Corollary of Proposition 3.6 with $r = \beta_g$. □

Finally, we shall state a result similar to [25], Lemma 3, yielding Condition 3.5 from simultaneous, but θ-dependent, drift and minorisation conditions. These conditions can be verified for random-walk Metropolis kernels with a target distribution having super-exponential tail decay and sufficiently regular tail contours [3,18,25,29].

Condition 3.10. *Suppose that* P *is an irreducible and aperiodic Markov kernel with invariant distribution* π, *that there exists a Borel set* $C \subset \mathsf{X}$, *a probability measure* ν *concentrated on* C, *constants* $\lambda \in [0, 1)$, $b < \infty$ *and* $\delta \in (0, 1]$ *such that* $v := \sup_{x \in C} V(x) < \infty$ *and*

$$PV(x) \leq \lambda V(x) + b\mathbb{I}\{x \in C\} \qquad \text{for all } x \in \mathsf{X},$$
$$P(x, A) \geq \delta \nu(A) \qquad \text{for all } x \in C \text{ and any Borel set } A \subset \mathsf{X}.$$

Proposition 3.11. *Assume Condition 3.10. Then, for any* $r \in (0, 1]$ *there exists a constant* $c_r^* \in [1, \infty)$ *depending only on* r *such that for all* $\|f\|_{V^r} < \infty$ *and* $k \geq 1$

$$\left\| P^k(x, f) - \pi(f) \right\|_{V^r} \leq V^r(x) M_r \rho_r^k \|f\|_{V^r},$$

where the constants $M_r \in [1, \infty)$ *and* $\rho_r \in (0, 1)$ *are defined in terms of the constants in Condition 3.10 as follows*

$$\rho_r := 1 - \left[c_r^*(1 - \lambda)^{-4} \delta^{-13} \bar{b}^6\right]^{-1},$$
$$M_r := c_r^*(1 - \lambda)^{-4} \delta^{-15} \bar{b}^7,$$

where $\bar{b} := b \vee v \geq 1$.

The proof of Proposition 3.11 is given in Appendix A.

3.2. Smooth family of Markov kernels

In many practically interesting settings, the mapping $\theta \mapsto P_\theta$, possibly restricted to a suitable set, satisfies a Hölder continuity condition. This continuity allows one to establish Condition 3.1(iv) in a natural way [3,4,8]. We restate these results in a quantitative manner below, so that they are directly applicable in the present setting. The Hölder continuity condition is given as follows.

Condition 3.12. *Suppose Condition 3.5 holds and for any $\theta, \theta' \in \hat{\Theta}$, there exist a constant $D_{\theta,\theta',r} \in [0,\infty)$ and a constant $\beta_D \in (0,\infty)$ independent of θ, θ' and r such that for any function $\|f\|_{V^r} < \infty$*

$$\|P_\theta f - P_{\theta'} f\|_{V^r} \le \|f\|_{V^r} D_{\theta,\theta',r} |\theta - \theta'|^{\beta_D}.$$

We consider below only the case when P_θ and $P_{\theta'}$ admit the same stationary measure; this is a commonly encountered in adaptive Markov chain Monte Carlo. The general case is slightly more involved, but can be handled as well; we refer the reader to [4] for details. We start by a lemma characterising the difference of the iterates of the kernels.

Lemma 3.13. *Assume Condition 3.12 holds and f is a measurable function with $\|f\|_{V^r} < \infty$ and that $\pi_\theta = \pi_{\theta'} =: \pi$. Then, for any $k \ge 0$*

$$\left\| P_\theta^k f - P_{\theta'}^k f \right\|_{V^r} \le M_{\theta,r} M_{\theta',r} D_{\theta,\theta',r} k (\rho_{\theta,r} \vee \rho_{\theta',r})^{k-1} \left|\theta - \theta'\right|^{\beta_D} \|f\|_{V^r}.$$

Proof. We use the following telescoping decomposition

$$P_\theta^k f - P_{\theta'}^k f = \sum_{j=1}^{k} P_\theta^{k-j} (P_\theta - P_{\theta'}) P_{\theta'}^{j-1} f = \sum_{j=1}^{k} \big(P_\theta^{k-j} - \Pi\big)(P_\theta - P_{\theta'})\big(P_{\theta'}^{j-1} f - \pi(f)\big),$$

where $\Pi(x, A) := \pi(A)$ for all $x \in \mathsf{X}$ and all measurable $A \subset \mathsf{X}$.

By Condition 3.5 and Condition 3.12,

$$\begin{aligned}\left\|(P_\theta - P_{\theta'})\big(P_{\theta'}^{j-1} f - \pi(f)\big)\right\|_{V^r} &\le \left\| P_{\theta'}^{j-1} f - \pi(f) \right\|_{V^r} D_{\theta,\theta',r} \left|\theta - \theta'\right|^{\beta_D} \\ &\le D_{\theta,\theta',r} M_{\theta',r} \rho_{\theta',r}^{j-1} \|f\|_{V^r} \left|\theta - \theta'\right|^{\beta_D}.\end{aligned}$$

Writing then

$$\left\| P_\theta^k f - P_{\theta'}^k f \right\|_{V^r} \le k \sup_{1 \le j \le k} \left\| \big(P_\theta^{k-j} - \Pi\big)(P_\theta - P_{\theta'})\big(P_{\theta'}^{j-1} f - \pi(f)\big) \right\|_{V^r},$$

and applying Condition 3.5 once more yields the claim. □

Proposition 3.14. *Assume Condition 3.12 holds, $\pi_\theta = \pi_{\theta'} =: \pi$ and $\|f_\theta\|_{V^r} \vee \|f_{\theta'}\|_{V^r} < \infty$. Then, the solutions of the Poisson equation defined as $g_\theta := \sum_{k=0}^{\infty} [P_\theta^k f_\theta - \pi_\theta(f_\theta)]$ satisfy*

$$\begin{aligned}\|g_\theta - g_{\theta'}\|_{V^r} \vee \|P_\theta g_\theta - P_{\theta'} g_{\theta'}\|_{V^r} \le{}& \frac{M_{\theta,r} M_{\theta',r} D_{\theta,\theta',r}}{(1 - (\rho_{\theta,r} \vee \rho_{\theta',r}))^2} |\theta - \theta'|^{\beta_D} \|f_\theta\|_{V^r} \\ &+ M_{\theta',r} (1 - \rho_{\theta',r})^{-1} \|f_\theta - f_{\theta'}\|_{V^r}.\end{aligned} \tag{3.6}$$

Proof. With the estimate from Lemma 3.13,

$$\begin{aligned}\|g_\theta - g_{\theta'}\|_{V^r} &\le \sum_{k=0}^{\infty}\bigl(\bigl\|P_\theta^k f_\theta - P_{\theta'}^k f_\theta\bigr\|_{V^r} + \bigl\|P_{\theta'}^k(f_\theta - f_{\theta'}) - \pi(f_\theta - f_{\theta'})\bigr\|_{V^r}\bigr)\\ &\le M_{\theta,r}M_{\theta',r}D_{\theta,\theta',r}\bigl|\theta-\theta'\bigr|^{\beta_D}\|f_\theta\|_{V^r}\sum_{k=0}^{\infty}k(\rho_{\theta,r}\vee\rho_{\theta',r})^{k-1}\\ &\quad + M_{\theta',r}(1-\rho_{\theta',r})^{-1}\|f_\theta - f_{\theta'}\|_{V^r}.\end{aligned}$$

The same bound clearly holds also for $\|P_\theta g_\theta - P_{\theta'}g_{\theta'}\|_{V^r}$ yielding (3.6). □

We shall provide some sufficient conditions to verify Condition 3.1(iv).

Condition 3.15. *Condition 3.12 holds with constants satisfying* $\sup_{(\theta,\theta')\in\mathcal{R}_i^2} D_{\theta,\theta',r} \le c_r^D \xi_i^{\alpha_D}$ *for some constant* $c_r^D \in [0,\infty)$ *depending only on* $r \in (0,1]$, *Condition 3.1(i) and (ii) hold with constants* α_H, β_H *and* α_V, *and there exist constants* $c<\infty$, $\alpha_\Delta \in [0,\infty)$ *and* $\beta_\Delta > 0$ *such that*

$$\sup_{(\theta,\theta')\in\mathcal{R}_i^2}\bigl\|H(\theta,\cdot) - H\bigl(\theta',\cdot\bigr)\bigr\|_{V^{\beta_H}} \le c\xi_i^{\alpha_\Delta}\bigl|\theta-\theta'\bigr|^{\beta_\Delta}.$$

Proposition 3.16. *Suppose Conditions 3.1(i) and (ii), 3.8 and 3.15 hold, the constants* $\beta_D, \beta_\Delta \in (0, 1/\beta_H - 1]$, *for any* $i \ge 0$ *the step size* Γ_i *is independent of* X_i *and the projections satisfy* $|\theta_{i+1} - \theta_i| \le |\theta_{i+1}^* - \theta_i|$. *Then, the solutions* g_θ *to the Poisson equation* $g_\theta - P_\theta g_\theta = \bar{H}(\theta,\cdot)$ *exist for all* $\theta \in \hat{\Theta}$, *and there is a constant* $c<\infty$ *such that for all* $(\theta,x)\in\mathcal{R}_0\times\mathsf{X}$

$$\begin{aligned}&\mathbb{E}_{\theta,x}\bigl|P_{\theta_i}g_{\theta_i}(X_i) - P_{\theta_{i-1}}g_{\theta_{i-1}}(X_i)\bigr|\\ &\quad\le c\mathbb{E}\bigl[\Gamma_i^{\beta_D}\bigr]\xi_i^{2\alpha_M+2\alpha_\rho+\alpha_D+(\beta_D+1)(\beta_H\alpha_V+\alpha_H)}V^{(\beta_D+1)\beta_H}(x)\\ &\qquad + c\mathbb{E}\bigl[\Gamma_i^{\beta_\Delta}\bigr]\xi_i^{\alpha_M+\alpha_\rho+\alpha_\Delta+\beta_\Delta\alpha_H+(\beta_\Delta+1)\beta_H\alpha_V}V^{(\beta_\Delta+1)\beta_H}(x).\end{aligned}$$

Proof. By assumption, both θ_i and θ_{i-1} are in $\mathcal{R}_i$, so $|\theta_i - \theta_{i-1}| \le \Gamma_i|H(\theta_{i-1},X_i)| \le c\Gamma_i\xi_i^{\alpha_H}V^{\beta_H}(X_i)$. Proposition 3.14 yields, with $r=\beta_H$ and denoting $H_\theta(x) := H(\theta,x)$,

$$\begin{aligned}&\|P_{\theta_i}g_{\theta_i} - P_{\theta_{i-1}}g_{\theta_{i-1}}\|_{V^{\beta_H}}\\ &\quad\le M_{\theta_i,\beta_H}M_{\theta_{i-1},\beta_H}D_{\theta_i,\theta_{i-1},\beta_H}\bigl(1-(\rho_{\theta_i,\beta_H}\vee\rho_{\theta_{i-1},\beta_H})\bigr)^{-2}|\theta_i-\theta_{i-1}|^{\beta_D}\|H_{\theta_i}\|_{V^{\beta_H}}\\ &\qquad + M_{\theta_{i-1},\beta_H}(1-\rho_{\theta_{i-1},\beta_H})^{-1}\|H_{\theta_i} - H_{\theta_{i-1}}\|_{V^{\beta_H}}\\ &\quad\le c\xi_i^{2\alpha_M+2\alpha_\rho+\alpha_D}|\theta_i-\theta_{i-1}|^{\beta_D}\|H_{\theta_i}\|_{V^{\beta_H}} + c\xi_i^{\alpha_M+\alpha_\rho}\|H_{\theta_i} - H_{\theta_{i-1}}\|_{V^{\beta_H}}\\ &\quad\le c\xi_i^{2\alpha_M+2\alpha_\rho+\alpha_D+\alpha_H(1+\beta_D)}\Gamma_i^{\beta_D}V^{\beta_D\beta_H}(X_i) + c\xi_i^{\alpha_M+\alpha_\rho+\alpha_\Delta+\beta_\Delta\alpha_H}\Gamma_i^{\beta_\Delta}V^{\beta_\Delta\beta_H}(X_i).\end{aligned}$$

The independence of Γ_i and X_i and Condition 3.1(ii) with Jensen's inequality (we have $(1 + (\beta_D \vee \beta_\Delta))\beta_H \in (0, 1]$) imply the claim. □

Now, we shall consider the common case where $(\Gamma_i)_{i\geq 1}$ is a deterministic power sequence. Then, Condition 3.1 can be established.

Proposition 3.17. *Suppose $\Gamma_i \equiv ci^{-\eta}$ for all $i \geq 1$ with some $c < \infty$ and $\eta \in (1/2, 1]$. Then, if the conditions of Proposition 3.16 hold and*

$$\sum_{i=1}^{\infty} i^{-(1+\beta_D)\eta} \xi_i^{\alpha_w+2\alpha_M+2\alpha_\rho+\alpha_D+(\beta_D+1)(\beta_H+\alpha_V+\alpha_H)} < \infty, \tag{3.7}$$

$$\sum_{i=1}^{\infty} i^{-(1+\beta_\Delta)\eta} \xi_i^{\alpha_M+\alpha_\rho+\alpha_\Delta+\beta_\Delta\alpha_H+(\beta_\Delta+1)\beta_H\alpha_V} < \infty, \tag{3.8}$$

$$\sum_{i=1}^{\infty} i^{-2\eta} \xi_i^{2\alpha_w+2(\alpha_H+\alpha_M+\alpha_\rho+\beta_H+\alpha_V)} < \infty, \tag{3.9}$$

then, Condition 3.1 holds.

Proof. Condition 3.1(i) and (ii) hold by assumption. Propositions 3.9 and 3.16 imply Condition 3.1(iii) with $\alpha_g = \alpha_H + \alpha_M + \alpha_\rho$ and $\beta_g = \beta_H$. Condition 3.1(iv) follows from Proposition 3.16 with (3.7) and (3.8).

Observe then that $\Gamma_{i+1}\Gamma_i \leq \Gamma_i^2 = c^2 i^{-2\eta}$ and by the mean value theorem $|\Gamma_{i+1} - \Gamma_i| = c\eta(i + h_i)^{-\eta-1} \leq c\eta i^{-\eta-1} \leq \eta\Gamma_i^2$ where $h_i \in [0, 1]$. Conditions 3.1(v)–(vii) follow easily from (3.9), by the fact $\alpha_g = \alpha_H + \alpha_M + \alpha_\rho$ and $\beta_g = \beta_H$. □

3.3. Non-smooth family of Markov kernels

When the mapping $\theta \to P_\theta$ does not admit (local) Hölder-continuity as discussed above, establishing Condition 3.1 is more involved, but possible using a random step size sequence which, in intuitive terms, enforce continuity in a stochastic manner. We focus on a specific step size sequence given as $\Gamma_i := \gamma_i \mathbb{I}\{U_i \leq p_i\}$ where the U_i are independent uniform $[0, 1]$ random variables and both sequences γ_i and p_i decay to zero. It will be clear later on that these sequences must satisfy $\sum_i \gamma_i p_i = \infty$, $\sum_i \gamma_i^2 p_i < \infty$ and $\sum_i \gamma_i p_i^2 < \infty$; for simplicity of exposition, we shall consider below the particular example where γ_i and p_i decay with a power law.

The definition of $(\Gamma_i)_{i\geq 1}$ above will result in practice in keeping the value of θ_i fixed for longer and longer (random) periods. We remark that one could consider inducing such a behaviour also in a deterministic manner, but we do not pursue this here.

Proposition 3.18. *Assume Conditions 2.1 and 3.8 hold and for all $i \geq 1$ the step size Γ_i is independent of X_i. Suppose also that Condition 3.1(i) holds with $\alpha_H \in [0, \infty)$ and $\beta_H \in [0, 1/2]$, and Condition 3.1(ii) holds with $\alpha_V \in [0, \infty)$.*

Then, the solutions g_θ to the Poisson equation $g_\theta - P_\theta g_\theta = \bar{H}(\theta, \cdot)$ exist for all $\theta \in \hat{\Theta}$, and there exists a constant $c < \infty$ such that for any $(\theta, x) \in \mathcal{R}_0 \times \mathsf{X}$

$$\mathbb{E}_{\theta,x}\big[\big|P_{\theta_i} g_{\theta_i}(X_i) - P_{\theta_{i-1}} g_{\theta_{i-1}}(X_i)\big|\big] \le c\mathbb{P}(\Gamma_i \neq 0)\xi_i^{\alpha_M+\alpha_\rho+\alpha_H+\beta_H\alpha_V} V^{\beta_H}(x).$$

Proof. The solutions g_θ to the Poisson equation exist by Proposition 3.6. If $\Gamma_i = 0$ then clearly $\theta_i = \theta_{i-1}$ and so

$$\begin{aligned}
&\big|P_{\theta_i} g_{\theta_i}(X_i) - P_{\theta_{i-1}} g_{\theta_{i-1}}(X_i)\big| \\
&\quad = \mathbb{I}\{\Gamma_i \neq 0\}\big|P_{\theta_i} g_{\theta_i}(X_i) - P_{\theta_{i-1}} g_{\theta_{i-1}}(X_i)\big| \\
&\quad \le c\mathbb{I}\{\Gamma_i \neq 0\}\big(\xi_i^{\alpha_M+\alpha_\rho}\big\|H(\theta_i, \cdot)\big\|_{V^{\beta_H}} + \xi_{i-1}^{\alpha_M+\alpha_\rho}\big\|H(\theta_{i-1}, \cdot)\big\|_{V^{\beta_H}}\big)V^{\beta_H}(X_i),
\end{aligned}$$

by Proposition 3.6. The claim follows by Conditions 3.1(i) and (ii), and by the independence of Γ_i and X_i. □

Next, we shall consider the particular case where $(\Gamma_i)_{i\ge1}$ is defined by two sequences with a power decay.

Proposition 3.19. *Let $(U_i)_{i\ge1}$ be a sequence of independent and uniformly distributed random variables on $[0, 1]$, and assume $\Gamma_i \equiv \gamma_i \mathbb{I}\{U_i \le p_i\}$, where the constant sequences $(\gamma_i)_{i\ge1} \subset (0, 1)$ and $(p_i)_{i\ge1} \subset [0, 1]$ are defined as $\gamma_i := c_\gamma i^{-\eta_\gamma}$ and $p_i := c_p i^{-\eta_p}$ for some $c_\gamma, c_p \in (0, \infty)$ and $\eta_\gamma, \eta_p \in (0, 1)$ such that $\eta_\gamma + \eta_p \le 1$, $2\eta_\gamma + \eta_p > 1$ and $\eta_\gamma + 2\eta_p > 1$.*

If Conditions 3.1(i) and (ii) and Condition 3.8 hold, and

$$\sum_{i=1}^{\infty} i^{-\eta_\gamma-2\eta_p}\xi_i^{\alpha_w+\alpha_M+\alpha_\rho+\alpha_H+\beta_H\alpha_V} < \infty, \tag{3.10}$$

$$\sum_{i=1}^{\infty} i^{-2\eta_\gamma-\eta_p}\xi_i^{2(\alpha_w+\alpha_H+\alpha_M+\alpha_\rho+\beta_H\alpha_V)} < \infty, \tag{3.11}$$

then, Condition 3.1 is satisfied.

Proof. Proposition 3.9 implies Condition 3.1(iii) with $\beta_g = \beta_H$ and $\alpha_g = \alpha_H + \alpha_M + \alpha_\rho$. Compute $\mathbb{E}[\Gamma_{i+1}]\mathbb{P}(\Gamma_i \neq 0) = \gamma_{i+1}p_{i+1}p_i \le ci^{-\eta_\gamma-2\eta_p}$. Then, Proposition 3.18 with (3.10) imply Condition 3.1(iv).

Let us then compute $\mathbb{E}[\Gamma_i^2] = \gamma_i^2 p_i = ci^{-2\eta_\gamma-\eta_p}$, and observe that $\mathbb{E}[\Gamma_{i+1}\Gamma_i] = ci^{-2\eta_\gamma-2\eta_p} \le ci^{-2\eta_\gamma-\eta_p}$ and that $|\mathbb{E}[\Gamma_{i+1} - \Gamma_i]| \le ci^{-\eta_\gamma-\eta_p-1} \le ci^{-2\eta_\gamma-\eta_p}$. With these bounds, (3.11) implies Conditions 3.1(v)–(vii). □

Remark 3.20. We emphasise that while our conditions on $(\Gamma_i)_{i\ge1}$ are only sufficient, it is necessary that the random step sizes decay to zero, that is $\limsup_{i\to\infty} \Gamma_i = 0$. Otherwise, the procedure might not converge; see [24], Example 4, for a related result in the context of adaptive Markov chain Monte Carlo.

4. Convergence

Up to this point, we have only considered the stability of the stochastic approximation process with expanding projections. Indeed, after showing the stability we know that the projections can occur only finitely often (almost surely), and the noise sequence can typically be controlled. Given this, the stochastic approximation literature provides several alternatives to show the convergence (e.g., [7–9,11,21]).

In some special cases, one can employ our stability results directly to establish convergence; namely, if the strict drift condition (2.7) holds outside an arbitrary small neighbourhood of the zeros of h. We believe, however, that such a result has only a limited applicability, because we suspect that it is often useful to consider two different Lyapunov functions w and $\hat{w}$ to establish the stability and convergence, respectively.

In many practical scenarios, the 'true' Lyapunov function $\hat{w}$, which would yield convergence, cannot be given in a closed form. It is also possible that $\hat{w}$ does not satisfy Condition 2.1 at all. We believe that it is often possible to find a simpler 'approximate Lyapunov function' w satisfying Condition 2.1, which yields a suitable drift away from the *boundary* of the space, but does not necessarily qualify as a true Lyapunov function to establish the convergence.

We formulate below a more general convergence result following [4] for reader's convenience.

Condition 4.1. *The set $\Theta \subset \mathbb{R}^d$ is open, the mean field $h : \Theta \to \mathbb{R}^d$ is continuous, and there exists a continuously differentiable function $\hat{w} : \Theta \to [0, \infty)$ such that*

(i) *there exists a constant $M_0 > 0$ such that*

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

(ii) *there exists $M_1 \in (M_0, \infty]$ such that $\{\theta \in \Theta : \hat{w}(\theta) \leq M_1\}$ is compact,*
(iii) *for all $\theta \in \Theta \setminus \mathcal{L}$, the inner product $\langle \nabla \hat{w}(\theta), \hat{h}(\theta) \rangle < 0$, and*
(iv) *the closure of $\hat{w}(\mathcal{L})$ has an empty interior.*

Theorem 4.2. *Assume Condition 4.1 holds, and let $\mathcal{K} \subset \Theta$ be a compact set intersecting $\mathcal{L}$, that is, $\mathcal{K} \cap \mathcal{L} \neq \varnothing$. Suppose that $(\gamma_i)_{i \geq 1}$ is a sequence of non-negative real numbers satisfying $\lim_{i \to \infty} \gamma_i = 0$ and $\sum_{i=1}^{\infty} \gamma_i = \infty$. Consider the sequence $(\theta_i)_{i \geq 0}$ taking values in Θ and defined through the recursion $\theta_i = \theta_{i-1} + \gamma_i h(\theta_{i-1}) + \gamma_i \varepsilon_i$ for all $i \geq 1$, where $(\varepsilon_i)_{i \geq 1}$ take values in $\mathbb{R}^d$.*

If there exists an integer i_0 such that $\{\theta_i\}_{i \geq i_0} \subset \mathcal{K}$ and $\lim_{m \to \infty} \sup_{n \geq m} |\sum_{i=m}^{n} \gamma_i \varepsilon_i| = 0$, then $\lim_{n \to \infty} \inf_{x \in \mathcal{L} \cap \mathcal{K}} |\theta_n - x| = 0$.

Proof. Theorem 4.2 is a restatement of [4], Theorem 2.3, but without the monotonicity assumption on the sequence $(\gamma_i)_{i \geq 1}$. The proof of [4], Theorem 2.3, applies unchanged, but the reader can also consult [5], Theorem 5, which is a slight generalisation of Theorem 4.2. □

Remark 4.3. The stability results of the present paper ensure that θ_i are eventually contained in a level set of w which can usually be assumed compact. Then, one can take $\mathcal{K} = \mathcal{W}_{M'}$ for some

(random) $M' > 0$, and the trajectories of $(\theta_i)_{i\geq 0}$ are eventually contained within $\mathcal{K}$, and there are only finitely many projections, almost surely. To employ Theorem 4.2, it then suffices to show that for any M in the possible range of w

$$\lim_{m\to\infty} \sup_{n\geq m} \left| \sum_{i=m}^{n} \Gamma_i \bar{H}(\theta_i, X_{i+1}) \mathbb{I}\{\theta_i \in \mathcal{W}_M\} \right| = 0. \tag{4.1}$$

For the sake of completeness and because our setting involves the random step sizes $(\Gamma_i)_{i\geq 1}$, we give a detailed theorem to establish this noise condition, by a straightforward modification of Theorem 3.3.

Theorem 4.4. *Suppose that for all $i \geq 1$, the step size Γ_i is independent of $\mathcal{F}_{i-1}$ and X_i, and the sums $\sum_{i\geq 1} \mathbb{E}[\Gamma_i^2]$ and $\sum_{i\geq 1} |\mathbb{E}[\Gamma_{i+1} - \Gamma_i]|$ are finite. Let $\mathcal{R} \subset \hat{\Theta}$ be a compact set such that there exists a constant $c < \infty$ so that for any $(\theta, x) \in \mathcal{R} \times \mathsf{X}$*

$$\sup_{i\geq 0} \mathbb{E}_{\theta,x}\left[V(X_{i+1}) \mathbb{I}\left\{A^i_{\mathcal{R}}\right\}\right] \leq cV(x), \tag{4.2}$$

$$\sup_{\theta\in\mathcal{R}} \left[\left|g_\theta(x)\right| + \left|P_\theta g_\theta(x)\right|\right] \leq cV^{\beta_g}(x), \tag{4.3}$$

$$\sum_{i=1}^{\infty} \mathbb{E}[\Gamma_{i+1}] \mathbb{E}_{\theta,x}\left[\left|P_{\theta_i} g_{\theta_i}(X_i) - P_{\theta_{i-1}} g_{\theta_{i-1}}(X_i)\right| \mathbb{I}\left\{A^i_{\mathcal{R}}\right\}\right] < \infty, \tag{4.4}$$

where g_θ is the solution of the Poisson equation as in Proposition 3.6 and $A^i_{\mathcal{R}} := \bigcap_{n=0}^{i} \{\theta_n \in \mathcal{R}\}$. Then, (4.1) holds for $\mathbb{P}_{\theta,x}$-almost every $\omega \in \bigcap_{i\geq 0} A^i_{\mathcal{R}}$.

The proof of Theorem 4.4 is given in Appendix B.

Remark 4.5. The condition (4.4) may be checked in practice either with Proposition 3.16 or with Proposition 3.19. To apply Theorem 4.2 in the case of random step sizes, one must check also that $\sum_{i=1}^{\infty} \Gamma_i$ diverges almost surely. Assuming the conditions of Theorem 4.4, it is sufficient to ensure that $\sum_{i=1}^{\infty} \mathbb{E}[\Gamma_i] = \infty$, because $Z_n := \sum_{i=1}^{n} (\Gamma_i - \mathbb{E}[\Gamma_i])$ form an a.s. convergent L^2-martingale.

5. Application: Particle independent Metropolis–Hastings expectation maximisation

We consider a stochastic approximation expectation maximisation (EM) algorithm [14] for static parameter maximum likelihood estimation in time series models, employing a particle indepen-

dent Metropolis–Hastings (PIMH) sampler [2] in order to approximate the expectation step of the EM algorithm. We present the generic algorithm in Section 5.1. Then, we focus on a specific example involving a Poisson count model with an intensity determined by a latent process. The model is given in Section 5.2 and the employed particle filter is discussed in Section 5.3. We establish the stability of the algorithm in Section 5.4 and conclude with a brief numerical experiment in Section 5.5.

5.1. Generic PIMH-EM algorithm

We assume a state space setting where a latent process $X_{1:n} := (X_1, X_2, \dots, X_n)$ defined on some measurable space $\mathcal{X}$ gives rise to an observation process $Y_{1:n} := (Y_1, Y_2, \dots, Y_n)$ taking values in a measurable space $\mathcal{Y}$ and assumed to consist of independent random variables given the latent process $X_{1:n}$. The process $X_{1:n}$ typically follows a Markov model parameterised by a vector ζ taking values in a measurable parameter space Ξ. The conditional marginal distributions of the observations given the latent process are also assumed to be parameterised by ζ. This allows one to define the so-called complete-data likelihood $p_\zeta(x_{1:n}, y_{1:n})$ for any $x_{1:n} \in \mathcal{X}^n$ and $y_{1:n} \in \mathcal{Y}^n$ and, when applicable, the EM algorithm allows one to iteratively maximise the likelihood $p_\zeta(y_{1:n})$. We will assume below that for any $x_{1:n} \in \mathcal{X}^n$ and $y_{1:n} \in \mathcal{Y}^n$ there exists a unique parameter value $\hat{\zeta} \in \Xi$ maximising the complete-data likelihood, which is also assumed to be uniquely determined through a vector of sufficient statistics taking values in an open set $\Theta \subset \mathbb{R}^d$.

Application of the EM algorithm requires one to compute the expectation of the complete-data log-likelihood with respect to $p_\zeta(\mathrm{d}x_{1:n}|y_{1:n})$. When this is not possible analytically one resorts to numerical methods, and we focus here on the use of Markov chain Monte Carlo (MCMC) algorithms. More precisely, we focus on the use of a methodology recently introduced in [2] which combines MCMC and particle filters and is particularly well suited to sampling in state-space models. Let us denote by $(\tilde{\mathbf{X}}, \mathbf{A}) \sim \mathrm{PF}(y_{1:n}, \zeta)$ the full output of a particle filter targeting the conditional distribution $p_\zeta(\mathrm{d}x_{1:n}|y_{1:n})$ of the model with the parameter value ζ. This output consists of all the random variables generated by the particle filter, that is, the state variables before resampling $\tilde{\mathbf{X}} \in \mathcal{X}^{n\times N}$ and the ancestor indices $\mathbf{A} \in \mathbb{N}^{(n-1)\times N}$; see [2] for details. The sample trajectories relevant to the approximation of quantities dependent on $p_\zeta(\mathrm{d}x_{1:n}|y_{1:n})$, denoted $X_{1:n,k} \in \mathcal{X}^n$ hereafter, and the associated weights $W_k \in [0, 1]$ for $k = 1, \dots, N$ can be recovered from $\tilde{\mathbf{X}}$ and $\mathbf{A}$ through functions $\bar{x}_{1:n} : \mathcal{X}^{n\times N} \times \mathbb{N}^{(n-1)\times N} \times \mathbb{N} \to \mathcal{X}^n$ and $\bar{w} : \mathcal{X}^{n\times N} \times \mathbb{N}^{(n-1)\times N} \times \mathbb{N} \to [0, 1]$, such that

$$X_{1:n,k} := \bar{x}_{1:n}(\tilde{\mathbf{X}}, \mathbf{A}, k) \quad \text{and} \quad W_k := \bar{w}(\tilde{\mathbf{X}}, \mathbf{A}, k).$$

We also introduce a 'sufficient statistics' function $t : \mathcal{X}^n \times \mathcal{Y}^n \to \Theta$ which, given a set of observations and one trajectory of the latent state variables, returns the sufficient statistics underpinning the complete-data likelihood. From our earlier assumption, we can define the function $\hat{\zeta} : \Theta \to \Xi$ which returns the parameter value maximising the conditional likelihood given some sufficient statistics $\theta \in \Theta$.

We can now summarise our PIMH-EM algorithm with the projections $\Pi_{\mathcal{R}_i} : \Theta \to \mathcal{R}_i$ to the sets $\mathcal{R}_0 \subset \mathcal{R}_1 \subset \cdots \subset \Theta$ as follows.

Algorithm 5.1. *Choose an initial value for the parameters $\zeta_0 \in \Xi$ and set*

$$\big(\tilde{\mathbf{X}}^{(0)}, \mathbf{A}_*^{(0)}\big) \sim \mathrm{PF}(y_{1:n}, \zeta_0), \tag{5.1}$$

$$\theta_0 := \Pi_{\mathcal{R}_0}\left[\sum_{k=1}^{N} W_k^{(0)} t\big(X_{1:n,k}^{(0)}, y_{1:n}\big)\right]. \tag{5.2}$$

For $i \geq 1$, proceed recursively as follows:

$$\big(\tilde{\mathbf{X}}_*^{(i)}, \mathbf{A}_*^{(i)}\big) \sim \mathrm{PF}\big(y_{1:n}, \hat{\zeta}(\theta_{i-1})\big), \tag{5.3}$$

$$\big(\tilde{\mathbf{X}}^{(i)}, \mathbf{A}^{(i)}\big) := \begin{cases} \big(\tilde{\mathbf{X}}_*^{(i)}, \mathbf{A}_*^{(i)}\big), & \text{with probability } \min\left\{1, \dfrac{\hat{Z}_{\hat{\zeta}(\theta_{i-1})}(\tilde{\mathbf{X}}_*^{(i)})}{\hat{Z}_{\hat{\zeta}(\theta_{i-1})}(\tilde{\mathbf{X}}^{(i-1)})}\right\}, \\ \big(\tilde{\mathbf{X}}^{(i-1)}, \mathbf{A}^{(i-1)}\big), & \text{otherwise}, \end{cases} \tag{5.4}$$

$$\theta_i := \Pi_{\mathcal{R}_i}\left[\theta_{i-1} + \Gamma_i\left(\sum_{k=1}^{N} W_k^{(i)} t\big(X_{1:n,k}^{(i)}, y_{1:n}\big) - \theta_{i-1}\right)\right], \tag{5.5}$$

where the step (5.4) *implements an accept-reject mechanism, and $\hat{Z}_\zeta(\hat{\mathbf{X}})$ stands for the estimate of the likelihood $p_\zeta(y_{1:n})$ computed with the given particles $\hat{\mathbf{X}}$* [2] *and $(\Gamma_i)_{i\geq 1}$ is a random step size sequence taking values in* $[0, \infty)$.

We can rewrite the steps (5.3) and (5.4) as $(\tilde{\mathbf{X}}^{(i)}, \mathbf{A}^{(i)}) \sim P^{\mathrm{PIMH}}_{\hat{\zeta}(\theta_{i-1})}((\tilde{\mathbf{X}}^{(i-1)}, \mathbf{A}^{(i-1)}), \cdot)$, in terms of a Markov kernel P^{PIMH}_ζ with the invariant distribution $\pi^{\mathrm{PIMH}}_\zeta(\mathrm{d}\tilde{\mathbf{x}}, \mathrm{d}\mathbf{a})$. As shown in [2], $\pi^{\mathrm{PIMH}}_\zeta(\mathrm{d}\tilde{\mathbf{x}}, \mathrm{d}\mathbf{a})$ has the property that for any function $f : \mathcal{X}^n \to \mathbb{R}$

$$\int \sum_{k=1}^{N} \bar{w}(\tilde{\mathbf{x}}, \mathbf{a}, k) f\big(\bar{x}_{1:n}(\tilde{\mathbf{x}}, \mathbf{a}, k)\big) \pi^{\mathrm{PIMH}}_\zeta(\mathrm{d}\tilde{\mathbf{x}}, \mathrm{d}\mathbf{a}) = \int f(x_{1:n}) p_\zeta(\mathrm{d}x_{1:n} | y_{1:n}),$$

whenever the integrals above are well-defined. Note that it is possible to further improve on this scheme by using smoothing procedures within the particle filtering procedure, but we do not consider such a possibility here. Given this, we define $H(\theta, (\tilde{\mathbf{x}}, \mathbf{a})) := \sum_{k=1}^{N} \bar{w}(\tilde{\mathbf{x}}, \mathbf{a}, k) t(\bar{x}_{1:n}(\tilde{\mathbf{x}}, \mathbf{a}, k)) - \theta$. Assuming $\Pi_{\mathcal{R}_i}(\theta) = \theta$ for all $\theta \in \mathcal{R}_i$, we can rewrite (5.3)–(5.5) in our generic stochastic approximation framework as follows

$$\begin{aligned} &\mathfrak{X}_i \sim P_{\theta_{i-1}}(\mathfrak{X}_{i-1}, \cdot), \\ &\theta_i^* = \theta_{i-1} + \Gamma_i H(\theta_{i-1}, \mathfrak{X}_i), \\ &\theta_i = \theta_i^* \mathbb{I}\big\{\theta_i^* \in \mathcal{R}_i\big\} + \theta_i^{\mathrm{proj}} \mathbb{I}\big\{\theta_i^* \notin \mathcal{R}_i\big\}, \end{aligned} \tag{5.6}$$

where $\mathfrak{X}_i := (\tilde{\mathbf{X}}^{(i)}, \mathbf{A}^{(i)})$ stands for the state variable, $P_{\theta_i} := P^{\mathrm{PIMH}}_{\hat{\zeta}(\theta_i)}$ and $\theta_i^{\mathrm{proj}} = \Pi_{\mathcal{R}_i}(\theta_i^*)$. Note also that the initial value θ_0 computed in (5.1) and (5.2) belongs to the initial projection set $\mathcal{R}_0$.

Remark 5.2. A similar algorithm to our PIMH-EM algorithm has been independently developed recently by Donnet and Samson [15]. They apply the algorithm to the problem of maximum likelihood estimation of static parameters in continuous-time diffusion models. Our work differs in various ways: at a theoretical level, Donnet and Samson [15] (essentially) assume a compact state space $\mathcal{X}$, which, among other things, eliminates the need to establish the stability of the recursion. At a methodological level, apart from the stabilisation procedure through the expanding projections scheme, our algorithm differs in that we use a random step size sequence, which allows us to consider families of Markov kernels $\{P_\theta\}_{\theta\in\Theta}$ which do not satisfy Hölder-continuity as discussed in Section 3.2.

5.2. Example: Poisson count model with random intensity

Our specific example is a Poisson count model with an intensity determined by a autoregressive process [10,16,31]. The latent stationary AR(1) process is determined by an initial distribution $X_1 \sim N(0, (1-\rho^2)^{-1}\sigma^2)$ and for $2 \le k \le n$ through

$$X_k = \rho X_{k-1} + \sigma\epsilon_k,$$

where ϵ_k are independent standard Gaussian random variables. The observations are conditionally independent following the law

$$Y_k | X_k \sim \mathrm{Poisson}\big(\mathrm{e}^{\alpha+X_k}\big).$$

For brevity, we keep $\rho \in (-1, 1)$ and $\sigma^2 > 0$ fixed, so that the unknown parameter of the model is $\zeta := \alpha \in \Xi := \mathbb{R}$.

The complete data log-likelihood for the model considered satisfies $\log(p_\zeta(x_{1:n}, y_{1:n})) = L(x_{1:n}, \zeta) + c$ where $c = c(\rho, \sigma^2) \in \mathbb{R}$ is a constant and

$$L(x_{1:n}, \zeta) := \sum_{i=1}^{n}\big[y_i(\alpha + x_i) - \mathrm{e}^{\alpha+x_i}\big] - \frac{1}{2\sigma^2}\left[x_1^2 + x_n^2 + \big(1+\rho^2\big)\sum_{i=2}^{n-1} x_i^2 - 2\rho\sum_{i=2}^{n} x_i x_{i-1}\right].$$

Let us introduce a sufficient statistics function $t(x_{1:n}, y_{1:n}) := t(x_{1:n}) := \sum_{i=1}^n \mathrm{e}^{x_i}$ taking values in $\Theta := (0, \infty)$. Then, denoting with $\mathbb{E}_\zeta$ the expectation with respect to $p_\zeta(\mathrm{d}x_{1:n}|y_{1:n})$, we can write the mean field of the stochastic approximation as

$$h(\theta) = \mathbb{E}_{\hat{\zeta}(\theta)}\big(t(X_{1:N})\big) - \theta.$$

It is straightforward to check that the unique parameter value maximising the complete-data likelihood is $\hat{\zeta}(\theta) := \hat{\alpha}(\theta) = \log(\frac{\bar{y}}{\theta})$, where $\bar{y} := \sum_{i=1}^n y_i$.

5.3. Particle filter for the example

We use the AR(1) process prior as a proposal distribution in our particle filter, that is,

$$q_\zeta(x_i|x_{1:i-1}, y_{1:i}) := p_\zeta(x_i|x_{i-1}) = N\big(x_i; \rho x_{i-1}, \sigma^2\big). \tag{5.7}$$

For our convenience, we augment the state space by adding an artificial initial state $X_0 \sim N(0, (1-\rho^2)^{-1}\sigma^2)$ with no associated observations, which we sample perfectly.

For our analysis, we need to quantify the dependence on ζ of the (geometric) rates of ergodicity of the PIMH kernel for a particular drift function. We shall see that for this it is sufficient to upper bound the weights of the particle filter and to lower bound the true likelihood.

Proposition 5.3. *The weights of the particle filter for* $1 \le i \le n$

$$w_\zeta(x_i, x_{i-1}) := \frac{p_\zeta(y_i|x_i)\, p_\zeta(x_i|x_{i-1})}{q_\zeta(x_i|x_{1:i-1}, y_{1:i})} \tag{5.8}$$

with the proposal distribution $q_\zeta(x_i|x_{1:i-1}, y_{1:i})$ *given in* (5.7)*, applied to the model described in Section* 5.2 *satisfy for all* $i \ge 1$

$$\sup_{(x_i, x_{i-1}) \in \mathbb{R}^2} w_\zeta(x_i, x_{i-1}) \le 1. \tag{5.9}$$

Proof. Because we use the prior proposal, the particle weights are determined by the likelihood. The observations are discrete, so the likelihood is upper bounded by one. □

Proposition 5.4. *The log-likelihood of the model satisfies, with* $\bar{y} := \sum_{i=1}^n y_i$*, the bound*

$$\log p_\zeta(y_{1:n}) \ge -\sum_{i=1}^n \log y_i! + \bar{y}\alpha - n \exp\left(\alpha + \frac{\sigma^2}{2(1-\rho^2)}\right). \tag{5.10}$$

Proof. We may write the log-likelihood in terms of an expectation with respect to the stationary latent process $X_{1:n}$, and use Jensen's inequality to obtain

$$\begin{aligned}\log p_\zeta(y_{1:n}) &= \log \mathbb{E}\left[\prod_{i=1}^n p(y_i|X_i, \zeta)\right] \ge \sum_{i=1}^n \mathbb{E}\big[\log p(y_i|X_i, \zeta)\big] \\ &= \sum_{i=1}^n \mathbb{E}\big[y_i(\alpha + Z) - \mathrm{e}^{\alpha+Z} - \log(y_i!)\big],\end{aligned}$$

where Z follows the stationary distribution of $X_{1:n}$, that is, Z is zero-mean Gaussian with the variance $\sigma_Z^2 := (1-\rho^2)^{-1}\sigma^2$. By recalling that the mean of a log-Gaussian random variable e^Z is $\exp(\sigma_Z^2/2)$, we obtain the desired bound (5.10). □

We now turn to the particle independent Metropolis–Hastings (PIMH) kernel in this context. Denote by q_ζ^{PF} the overall distribution of the random variables $(\tilde{\mathbf{X}}, \mathbf{A})$ generated by the particle filter with the proposal distribution $q_\zeta(x_i|x_{1:i-1}, y_{1:i})$ given in (5.7) and targeting $p_\zeta(x_{1:n}, y_{1:n})$. The PIMH is nothing but an ordinary independent Metropolis–Hastings algorithm with the proposal distribution q_ζ^{PF} and the target distribution $\pi_\zeta^{\mathrm{PIMH}}$.

Proposition 5.5. *The ratio of the overall distribution of the particle filter and the target density satisfies the bound*

$$\inf_{(\tilde{\mathbf{x}},\mathbf{a})\in\mathsf{X}} \frac{\mathrm{d}q_\zeta^{\mathrm{PF}}}{\mathrm{d}\pi_\zeta^{\mathrm{PIMH}}}(\tilde{\mathbf{x}}, \mathbf{a}) \geq c_1 \exp\bigl[\bar{y}\alpha - c_2 \mathrm{e}^{\alpha}\bigr], \tag{5.11}$$

with constants $c_1 = c_1(y_{1:n}) > 0$ and $c_2 = c_2(\rho, \sigma^2, n) > 0$.

Proof. In case of the Particle IMH, [2], page 299,

$$\frac{\mathrm{d}\pi_\zeta^{\mathrm{PIMH}}}{\mathrm{d}q_\zeta^{\mathrm{PF}}}(\tilde{\mathbf{x}}, \mathbf{a}) = \frac{\hat{Z}_\zeta(\tilde{\mathbf{x}}, \mathbf{a})}{Z_\zeta} = \prod_{k=1}^{n} \frac{1}{N} \sum_{i=1}^{N} w_\zeta\bigl(\tilde{x}_{k,i}, \tilde{x}_{k-1,i}^{a}\bigr) \Big/ p_\zeta(y_{1:n}),$$

where N is the number of particles, w_ζ are the unnormalised particle weights given in (5.8) and $\tilde{x}_{k,i}$ and $\tilde{x}_{k-1,i}^{a}$ stand for the ith particle at time k and its ancestor, respectively. The bound (5.11) follows directly from the bounds (5.9) and (5.10) established in Propositions 5.3 and 5.4, respectively. □

The bound on the ratio of the proposal and target densities in Proposition 5.5 ensures a uniform ergodicity of the PIMH sampler. We, however, must be able to analyse the ergodic behaviour of the algorithm for unbounded functions. Therefore, we consider geometric ergodicity with a certain 'drift' function V, which will allow us to control averages of functions f such that $\sup_{x\in\mathsf{X}} |f(x)|/V(x) < \infty$.

Proposition 5.6. *Let $q_\zeta^{\mathrm{PF}}(\mathrm{d}\tilde{\mathbf{x}}, \mathrm{d}\mathbf{a})$ stand for the overall proposal density of the particle filter with the one-step proposal density $q_\zeta(x_i|x_{1:i-1}, y_{1:i})$ given in (5.7) and denote*

$$V(\tilde{\mathbf{x}}, \mathbf{a}) := \sum_{i=1}^{n} \sum_{j=1}^{N} \mathrm{e}^{2|\tilde{x}_i^j|}.$$

Then, the following bounds hold

$$q_\zeta(V) \leq 2nN^n \exp\left(\frac{2\sigma^2}{1-\rho^2}\right), \tag{5.12}$$

$$\sup_{(\tilde{\mathbf{x}},\mathbf{a})\in\mathsf{X}} \frac{H(\theta, (\tilde{\mathbf{x}}, \mathbf{a}))}{V^{1/2}(\tilde{\mathbf{x}}, \mathbf{a})} \leq \sqrt{n}N + \frac{|\theta|}{V^{1/2}(\tilde{\mathbf{x}}, \mathbf{a})}. \tag{5.13}$$

Proof. The overall proposal density of the particle filter without selection $\hat{q}_\zeta(x_{1:n})$ is in fact the finite-dimensional distribution of the stationary AR(1) prior. Denote by $\hat{X}_{1:n} \sim \hat{q}_\zeta$. We obtain by a crude bound

$$q_\zeta(V) \le \sum_{i=1}^{n} N^i \mathbb{E}\big[\mathrm{e}^{2|\hat{X}_i|}\big] \le n N^n \sup_{1\le i\le n} \mathbb{E}\big[\mathrm{e}^{-2\hat{X}_i} + \mathrm{e}^{2\hat{X}_i}\big].$$

Our $\hat{X}_i$ are Gaussian with zero mean and variance $\sigma^2/(1-\rho^2)$, and $\mathbb{E}[\exp(\pm\hat{X}_i)] = \exp(\mathrm{Var}(\hat{X}_i)/2)$. We obtain (5.12).

Consider then (5.13). Because $|\bar{w}| \le 1$, we have

$$\big|H(\theta, (\tilde{\mathbf{x}}, \mathbf{a})\big| \le N \sup_{1\le k\le N} \big|t\big(\bar{x}_{1:n}(\tilde{\mathbf{x}}, \mathbf{a}, k)\big)\big| + |\theta|.$$

Because $\bar{x}_{1:n}$ only chooses a path among the state variables $\tilde{\mathbf{x}}$ and the sufficient statistics of the chosen paths satisfy

$$t\big(\bar{x}_{1:n}(\tilde{\mathbf{x}}, \mathbf{a}, k)\big)^2 = \left(\sum_{i=1}^{n} \exp\big(\bar{x}_i(\tilde{\mathbf{x}}, \mathbf{a}, k)\big)\right)^2 \le n \sum_{i=1}^{n} \exp\big(2\bar{x}_i(\tilde{\mathbf{x}}, \mathbf{a}, k)\big),$$

where $\bar{x}_i(\tilde{\mathbf{x}}, \mathbf{a}, k) = \tilde{x}_{i,j(k,i)}$ for some integer $1 \le j(k,i) \le N$. Therefore, $|t(\bar{x}_{1:n}(\tilde{\mathbf{x}}, \mathbf{a}, k))| \le \sqrt{n} V^{1/2}(\tilde{\mathbf{x}}, \mathbf{a})$, and we get (5.13). □

5.4. Stability of the PIMH-EM

We already have most of the ingredients to establish the stability of the PIMH-EM algorithm with expanding projections applied to our example Poisson count model with random intensity. What remains is to identify a Lyapunov function w for the sufficient statistic. For this purpose, we study the properties of the mean field $h(\theta)$.

Proposition 5.7. *For any constant $c \in (1, \infty)$ there exists a $c_\theta = c_\theta(c, \sigma^2, \rho, y_{1:n}) \in (0, 1]$ such that*

$$h(\theta) \ge c\theta^{1-(1/2)\mathbf{1}^T\Sigma^{-1}\mathbf{1}\log\theta} \qquad \text{for all } \theta \in \big(0, c_\theta\big], \tag{5.14}$$

$$h(\theta) \le -c^{-1}\theta \qquad \text{for all } \theta \in \big[c_\theta^{-1}, \infty\big). \tag{5.15}$$

Proof. Observe first that we may write, up to a constant,

$$p_\zeta(x_{1:n}, y_{1:n}) = \det\big(\Sigma^{-1/2}\big) \exp\left(-\frac{1}{2} x_{1:n}^T \Sigma^{-1} x_{1:n} + \sum_{i=1}^{n} \big[y_i(\alpha + x_i) - \mathrm{e}^{\alpha+x_i}\big]\right),$$

where $\Sigma^{-1} = \Sigma^{-1}(\rho, \sigma^2) \in \mathbb{R}^{n\times n}$ is a symmetric and positive definite matrix with all elements equal to zero except the diagonal elements which satisfy $\Sigma^{-1}_{1,1} = \Sigma^{-1}_{n,n} = 1/\sigma^2$ and $\Sigma^{-1}_{2,2} = \cdots =$

$\Sigma^{-1}_{n-1,n-1} = (1+\rho^2)/\sigma^2$, and the first diagonal above and below the main diagonal which are such that $\Sigma^{-1}_{i,i-1} = \Sigma^{-1}_{i-1,i} = -\rho/\sigma^2$ for $i = 2, \dots, n$.

We may write the mean field as

$$\begin{aligned} h(\theta) &= \int_{\mathbb{R}^n} \left(\sum_{i=1}^n \mathrm{e}^{x_i} - \theta \right) \frac{p_{\hat{\alpha}(\theta)}(x_{1:n}, y_{1:n})}{p_{\hat{\alpha}(\theta)}(y_{1:n})} \, \mathrm{d}x_{1:n} \\ &= \theta \int_{\mathbb{R}^n} \exp\left(-\frac{1}{2} x^T \Sigma^{-1} x + \sum_{i=1}^n y_i x_i - \bar{y} \sum_{i=1}^n \frac{\mathrm{e}^{x_i}}{\theta} \right) \left(\sum_{i=1}^n \frac{\mathrm{e}^{x_i}}{\theta} - 1 \right) \mathrm{d}x \qquad (5.16) \\ &\quad \Big/ \int_{\mathbb{R}^n} \exp\left(-\frac{1}{2} x^T \Sigma^{-1} x + \sum_{i=1}^n y_i x_i - \bar{y} \sum_{i=1}^n \frac{\mathrm{e}^{x_i}}{\theta} \right) \mathrm{d}x. \end{aligned}$$

For (5.15), it is enough to observe that by dominated convergence $\lim_{\theta\to\infty} h(\theta)/\theta = -1$.

Let us then consider the case where θ is small (5.14). Denote the numerator in (5.16) by N_h, and use the change of variables $u_i := e^{x_i}/\theta$ for all $i = 1, \dots, n$ to write

$$\begin{aligned} N_h = \int_{\mathbb{R}^n_+} &\exp\left(-\frac{1}{2} (\log\theta \times \mathbf{1} + \log u)^T \Sigma^{-1} (\log\theta \times \mathbf{1} + \log u) \right) \left(\sum_{i=1}^n u_i - 1 \right) \\ &\times \exp\left(\sum_{i=1}^n y_i \log(\theta u_i) - \bar{y} \sum_{i=1}^n u_i \right) \frac{\mathrm{d}u}{\prod_{i=1}^n u_i}, \end{aligned}$$

where we use the convention $\log u := [\log u_1, \dots, \log u_n]^T$ and $\mathbf{1} := [1, \dots, 1]^T$. By rearranging the terms, this can be written as

$$N_h = \theta^{\bar{y} - (1/2)\mathbf{1}^T \Sigma^{-1} \mathbf{1} \log\theta} \int_{\mathbb{R}^n_+} \theta^{-\mathbf{1}^T \Sigma^{-1} \log u} \left(\sum_{i=1}^n u_i - 1 \right) g_\Sigma(u) \, \mathrm{d}u, \qquad (5.17)$$

where the function g_Σ is independent of θ and for all $u \in \mathbb{R}^n_+$ and all $\Sigma^{-1} \in \mathbb{R}^{n\times n}$,

$$g_\Sigma(u) := \exp\left(-\frac{1}{2} \log u^T \Sigma^{-1} \log u + \sum_{i=1}^n (y_i - 1) \log u_i - \bar{y} \sum_{i=1}^n u_i \right) > 0.$$

We shall partition the domain $\mathbb{R}^n_+$ according to the sign of the integrand in (5.17) as $I_- := \{u \in \mathbb{R}^n_+ \colon \sum_{i=1}^n u_i < 1\}$ and $I_+ := \mathbb{R}^n_+ \setminus I_-$. Observe that for all $u \in I_-$, the elements of $\log u$ are all negative, and the row sums of Σ^{-1} are all positive. Therefore, $-\mathbf{1}^T \Sigma^{-1} \log u > 0$ for all $u \in I_-$ and because the integral is finite for any fixed $\theta > 0$,

$$\lim_{\theta\to 0+} \int_{I_-} \theta^{-\mathbf{1}^T \Sigma^{-1} \log u} \left(\sum_{i=1}^n u_i - 1 \right) g_\Sigma(u) \, \mathrm{d}u = 0.$$

On the other hand, considering the subset $\hat{I}_+ := \{u \in \mathbb{R}_+^n \colon \forall i = 1, \dots, n \ \log(u_i) > 0\} \subset I_+$, then similarly $-\mathbf{1}^T \Sigma^{-1} \log u < 0$ for all $u \in \hat{I}_+$, whence

$$\lim_{\theta \to 0+} \int_{\hat{I}_+} \theta^{-\mathbf{1}^T \Sigma^{-1} \log u} \left(\sum_{i=1}^n u_i - 1 \right) g_\Sigma(u) \, \mathrm{d}u = \infty.$$

Overall, we deduce that for any constant $c' > 0$ there exists a $c_\theta = c_\theta(c', \Sigma, y_{1:n}) > 0$ such that for all $\theta \in (0, c_\theta)$,

$$N_h \geq c' c_\Sigma \theta^{\bar{y} - (1/2)\mathbf{1}^T \Sigma^{-1} \mathbf{1} \log \theta} > 0.$$

We are left with upper bounding the denominator D_h in (5.16), which we write as an expectation with respect to a random variable $X \sim N(0, \Sigma)$

$$D_h = c_\Sigma \mathbb{E}\left[\exp\left(\sum_{i=1}^n y_i X_i - \frac{\bar{y}}{\theta} \sum_{i=1}^n \mathrm{e}^{X_i} \right) \right].$$

By elementary calculus, one can compute that for $y, \bar{y}, \theta > 0$

$$\sup_{x \in \mathbb{R}} \exp\left(yx - \frac{\bar{y}}{\theta} \mathrm{e}^x \right) = \theta^y \exp\left(y \log \frac{y}{\bar{y}} - y \right),$$

so $D_h \leq c_{y_{1:n}, \Sigma} \theta^{\bar{y}}$, and we deduce (5.14) by choosing c' sufficiently large. □

Now we are ready to establish the stability of the PIMH-EM in our example setting.

Proposition 5.8. *Consider Algorithm 5.1 applied to the model specified in Section 5.2, with the projections (5.6). The projection sets are defined as* $\mathcal{R}_i := \{\theta \in \Theta \colon \underline{\theta}_i \leq \theta \leq \bar{\theta}_i\}$ *and the projections as* $\theta_i^{\mathrm{proj}} := (\underline{\theta}_i \vee \theta_i^*) \wedge \bar{\theta}_i$, *with the constant sequences* $\underline{\theta}_i \downarrow 0$ *and* $\bar{\theta}_i \uparrow \infty$ *satisfying*

$$\liminf_{i \to \infty} \underline{\theta}_i \log(i) = \infty \quad \textit{and} \quad \limsup_{i \to \infty} \frac{\bar{\theta}_i}{i^\epsilon} = 0$$

for all $\epsilon > 0$. *The step sizes are defined as* $\Gamma_i := c_\gamma i^{-\eta_\gamma} \mathbb{I}\{U_i \leq c_p i^{-\eta_p}\}$ *where* $c_\gamma, c_p \in (0, \infty)$, *and the constants* $\eta_\gamma, \eta_p \in (0, 1)$ *satisfy* $\eta_\gamma + \eta_p < 1$, $2\eta_\gamma + \eta_p > 1$ *and* $\eta_\gamma + 2\eta_p > 1$, *and* $(U_i)_{i \geq 1}$ *are uniform* $(0, 1)$ *distributed random variables independent on the history* $\mathcal{F}_{i-1}$ *and* X_i.

Then, there exists $0 < c_1 < c_2 < \infty$ *such that for any* $(\theta, x) \in \mathcal{R}_0 \times \mathsf{X}$,

$$\mathbb{P}_{\theta, x}\left(\bigcup_{m=1}^\infty \bigcap_{n=m}^\infty \{c_1 \leq \theta_i \leq c_2\} \right) = 1.$$

Proof. Let $c_\theta \in (0, 1)$ be the constant from Proposition 5.7 applied with, say, $c = 1$, and define $\hat{w}(\theta) := |\theta - c_\theta^*|$ with $c_\theta^* := (c_\theta + c_\theta^{-1})/2$. Define w as the smoothed version of $\hat{w}$ through the

convolution $w := \hat{w} * \phi$ with a C^∞-mollifier ϕ supported on a sufficiently small $[-\epsilon_\phi, \epsilon_\phi]$, so that $w = \hat{w}$ on $(0, c_\theta] \cup [c_\theta^{-1}, \infty)$. Then, w is twice differentiable with bounded derivatives, $w(\theta) < w(\theta')$ for all $\theta \in \mathcal{W}_{M_0} = [c_\theta, c_\theta^{-1}]$ and $\theta' \in \mathbb{R} \setminus \mathcal{W}_{M_0}$, where $M_0 := c_\theta^* - c_\theta > 0$. To sum up, letting $\xi_i := i \vee 1$ for $i \geq 0$, Conditions 2.1(i), (ii), (iv) and (v) hold with $\alpha_w = 0$ and with some constant $c < \infty$.

Now, we turn into establishing Condition 2.7. The bounds from Proposition 5.7 imply $\delta := \inf_{\theta \geq c_\theta} -\langle h(\theta), \nabla w(\theta) \rangle > 0$ and

$$\begin{aligned}\delta_i &:= \inf_{\theta \in [\underline{\theta}_i, c_\theta^{-1}]} -\big\langle h(\theta), \nabla w(\theta) \big\rangle \geq c \inf_{\theta \in [\underline{\theta}_i, c_\theta^{-1}]} \theta^{1 - c_h \log(\theta)} \\ &= c \underline{\theta}_i^{1 - c_h \log(\underline{\theta}_i)} \geq c_1 (\log i)^{-c_2 \log\log i}\end{aligned}$$

for $i \geq 2$, where $c_1, c_2 \in (0, \infty)$. Therefore, with our choice of the step sizes $\sum_{i=1}^\infty (\delta \wedge \delta_i) \mathbb{E}[\Gamma_i] = \infty$, implying that $\sum_{i=1}^\infty (\delta \wedge \delta_i) \Gamma_i = \infty$ almost surely.[2]

Recalling that $\hat{\alpha}(\theta) = \log(\bar{y}/\theta)$, we bound by Proposition 5.5

$$\hat{\epsilon}(\theta) := \inf_{(\tilde{\mathbf{x}}, \mathbf{a}) \in \mathsf{X}} \frac{\mathrm{d} q^{\mathrm{PF}}_{\hat{\alpha}(\theta)}}{\mathrm{d} \pi^{\mathrm{PIMH}}_{\hat{\alpha}(\theta)}} (\tilde{\mathbf{x}}, \mathbf{a}) \geq c_1 \left(\frac{\mathrm{e}^{-c_2/\theta}}{\theta} \right)^{\bar{y}},$$

where $c_1, c_2 < \infty$ are constants independent of θ. Now, fix an $\varepsilon > 0$. Then, it is straightforward to check that there exists a constant $c < \infty$ such that for all $i \geq 1$

$$\sup_{\theta \in \mathcal{R}_i} \frac{1}{\hat{\epsilon}(\theta)} = \left(\sup_{\theta \in [\underline{\theta}_i, 1]} \frac{1}{\hat{\epsilon}(\theta)} \right) \vee \left(\sup_{\theta \in [1, \bar{\theta}_i]} \frac{1}{\hat{\epsilon}(\theta)} \right) \leq c \xi_i^\varepsilon .$$

Without loss of generality, we may assume $\hat{\epsilon}(\theta) \leq 1/2$, so Corollary C.2 implies that the P_θ is geometrically ergodic with constants $\hat{M} = \hat{M}(\hat{\epsilon}(\theta)) = c \hat{\epsilon}^{-2}(\theta)$ and $\hat{\rho} = \hat{\rho}(\hat{\epsilon}(\theta)) = (1 - \hat{\epsilon}(\theta)/2)$. It is easy to see that then Condition 3.8 holds with $\alpha_M = 2\varepsilon$ and $\alpha_\rho = \varepsilon$.

Let V be defined as in Proposition 5.6. Then, there exists a constant $c < \infty$ such that

$$\sup_{\theta \in \mathcal{R}_i} \big\| H(\theta, \cdot) \big\|_{V^{1/2}} \leq c_2 + \sup_{\theta \in \mathcal{R}_i} |\theta| = c_2 + \bar{\theta}_i \leq c \xi_i^\varepsilon ,$$

implying Condition 3.1(i) with $\beta_H = 1/2$ and $\alpha_H = \varepsilon$. The drift condition assumed in Lemma 3.7 holds with $\lambda_i = 1 - \inf_{\theta \in \mathcal{R}_i} \hat{\epsilon}(\theta)$ and $b_i = b < \infty$ due to Corollary C.2. This implies Condition 3.1(ii) with $\alpha_V = \alpha_\rho = \varepsilon$.

Now, Proposition 3.19 is applicable as soon as we choose $\varepsilon > 0$ above sufficiently small so that

$$\alpha_w + \alpha_M + \alpha_\rho + \alpha_H + \beta_H \alpha_V < (\eta_\gamma + 2\eta_p - 1) \wedge \frac{2\eta_\gamma + \eta_p - 1}{2}.$$

Proposition 3.19 implies Condition 3.1, allowing us to establish the noise condition in Theorem 3.3. Finally, Theorem 2.8 yields the claim with $c_1 = c_\theta$ and $c_2 = c_\theta^{-1}$. □

[2]The random variables $Z_n := \sum_{i=1}^n (\delta \wedge \delta_i)(\Gamma_i - \mathbb{E}[\Gamma_i])$ form an a.s. convergent L^2-martingale.

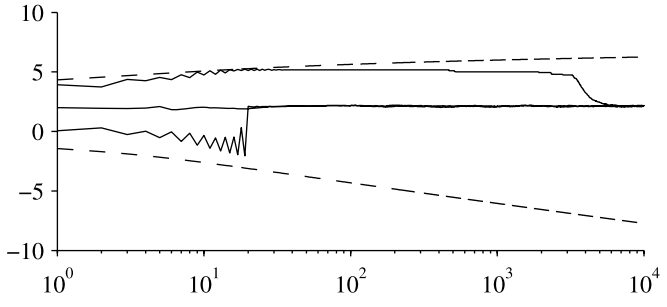

Figure 2. Trajectories of the estimate $\hat{\alpha}(\theta_i)$ corresponding the PIMH-EM started from three different initial values for $\hat{\alpha}_0$. The dashed lines correspond to the boundaries induced to $\hat{\alpha}(\theta_i)$ by $(\underline{\theta}_i)_{i\geq 0}$ and $(\bar{\theta}_i)_{i\geq 0}$. Notice the logarithmic scale on the x-axis (iterations).

We remark that the condition for $\bar{\theta}_i$ in Proposition 5.8 can be relaxed by only assuming it to hold with a certain fixed $\epsilon > 0$ depending on $\bar{y}$, η_γ and η_p.

5.5. Numerical experiment

We illustrate our algorithm briefly in practice in the setup of Proposition 5.8. We consider the same setting as Fort and Moulines [16]: we have $n = 100$ simulated observations of the model of Section 5.2 with parameters $\alpha = 2$, $\rho = 0.4$ and $\sigma^2 = 1$.

We use the following projection sequences to control the sufficient statistic

$$\underline{\theta}_i := \underline{c}\log^{\underline{\epsilon}-1}(i+2) \quad \text{and} \quad \bar{\theta}_i := \bar{c}_1(i+2)^{\bar{c}_2/\log^{\bar{\epsilon}}(i+2)},$$

with the constants $\underline{c} = 0.1m_\theta$, $\bar{c}_1 = 10m_\theta$, $\underline{\epsilon} = \bar{\epsilon} = 0.1$ and $\bar{c}_2 = 1$, where $m_\theta := n\exp(\frac{\sigma^2}{2(1-\rho^2)})$ is the prior expectation of the sufficient statistic. The step size sequence parameters are $c_\gamma = 6$, $c_p = 3$ and $\gamma_\eta = \gamma_p = 0.35$. The number of particles is set to $N = 1000$.

Figure 2 shows the trajectories of the estimates $\hat{\alpha}(\theta_i)$ for 10,000 iterations of the algorithm starting from three different initial values $\hat{\alpha}_0 \in \{0, 2, 4\}$. The final values of the estimates $\hat{\alpha}$ are within 2.10–2.16. The average acceptance rate during the runs varied between 46–72%. Notice the unstable initial behaviour of the estimates in Figure 2, which is controlled by the projections.

Appendix A: Geometric ergodicity from drift condition

Before the proof of Proposition 3.11, we restate the result by Meyn and Tweedie [23] upon which the proof relies.

Theorem A.1 (Meyn and Tweedie [23] Theorem 2.3). *Suppose Condition 3.10 holds. Then, for all $k \geq 0$ and $\|f\|_V < \infty$*

$$\left|P_s^k(x, f) - \pi(f)\right| \leq V(x)(1+\gamma)\frac{\rho}{\rho - \vartheta}\rho^k\|f\|_V$$

for any $\rho > \vartheta = 1 - \tilde{M}^{-1}$, *for*

$$\tilde{M} = \frac{1}{(1-\check{\lambda})^2}\big[1 - \check{\lambda} + \check{b} + \check{b}^2 + \bar{\zeta}\big(\check{b}(1-\check{\lambda}) + \check{b}^2\big)\big],$$

defined in terms of

$$\gamma = \delta^{-2}[4b + 2\delta\lambda v], \qquad \check{\lambda} = (\lambda+\gamma)/(1+\gamma) < 1 \quad \textit{and} \quad \check{b} = v + \gamma < \infty,$$

and the bound

$$\bar{\zeta} \le \frac{4-\delta^2}{\delta^5}\bigg(\frac{b}{1-\lambda}\bigg)^2.$$

Proof of Proposition 3.11. Let us first consider the claim for $r = 1$. Define first

$$\bar{\zeta} := \big(4-\delta^2\big)\delta^{-5}b^2(1-\lambda)^{-2} \le 4\delta^{-5}\bar{b}^2(1-\lambda)^{-2},$$

and observe that $\gamma := \delta^{-2}[4b + 2\delta\lambda v] \le 6\delta^{-2}\bar{b}$. We also have

$$\check{\lambda} := \frac{\lambda+\gamma}{1+\gamma} \le \frac{\lambda + 6\delta^{-2}\bar{b}}{1 + 6\delta^{-2}\bar{b}} \quad \text{implying} \quad \frac{1}{1-\check{\lambda}} \le \frac{1+6\delta^{-2}\bar{b}}{1-\lambda} \le \frac{7\delta^{-2}\bar{b}}{1-\lambda}.$$

We have also $\check{b} := v + \gamma \le 7\delta^{-2}\bar{b}$. Now, we can bound

$$\begin{aligned}\tilde{M} &:= \frac{1}{(1-\check{\lambda})^2}\big[1 - \check{\lambda} + \check{b} + \check{b}^2 + \bar{\zeta}\big(\check{b}(1-\check{\lambda}) + \check{b}^2\big)\big] \\ &\le \frac{1}{(1-\check{\lambda})^2}\bar{\zeta}\big(5\check{b}^2\big) \le 48{,}020(1-\lambda)^{-4}\delta^{-13}\bar{b}^6.\end{aligned}$$

Now we can take $\rho_1 := 1 - [100{,}000(1-\lambda)^{-4}\delta^{-13}\bar{b}^6]^{-1}$ satisfying $\rho_1 > 1 - \tilde{M}^{-1}/2$. Finally, the claim holds with $c_1^* = c^* := 336{,}140$ by setting

$$M_1 := (1+\gamma)\frac{\rho}{\rho - (1-\tilde{M}^{-1})} \le (1+\gamma)2\tilde{M} \le 336{,}140(1-\lambda)^{-4}\delta^{-15}\bar{b}^7.$$

Let us consider then the case $r \in (0,1)$. Observe first that by Jensen's inequality

$$\begin{aligned}PV^r(x) &\le \big(PV(x)\big)^r \le \lambda^r V^r(x) \qquad \text{for all } x \notin C, \\ PV^r(x) &\le \Big(\sup_{z\in C} V(z) + b\Big)^r \le 2^r(v \vee b)^r \qquad \text{for all } x \in C.\end{aligned}$$

That is, Condition 3.10 holds for V^r with $\lambda_r := \lambda^r$, $\bar{b}_r := 2\bar{b}^r$, and $v_r := \sup_{x\in C} V^r(x) = (\sup_{x\in C} V(x))^r = v^r$. Because $t \mapsto t^r$ is concave, $\lambda^r \le 1 - r(1-\lambda)$ and so $(1-\lambda^r)^{-1} \le r^{-1}(1-\lambda)^{-1}$. We may take $c_r^* := (2r^{-1})^4 c^*$. □

Appendix B: Noise condition for convergence theorem

Proof of Theorem 4.4. We give only the required modifications to the proof of Theorem 3.3 regarding (3.2). First, by symbolically substituting $\nabla w \equiv 1$, it is sufficient to show that claim holds for the following four terms in turn:

$$R^1_{m,n} := \sum_{i=m}^{n} \Gamma_{i+1}\big(g_{\theta_i}(X_{i+1}) - P_{\theta_i} g_{\theta_i}(X_i)\big)\mathbb{I}\big\{A^i_{\mathcal{R}}\big\},$$

$$R^2_{m,n} := \sum_{i=m}^{n} \Gamma_{i+1}\big(P_{\theta_i} g_{\theta_i}(X_i) - P_{\theta_{i-1}} g_{\theta_{i-1}}(X_i)\big)\mathbb{I}\big\{A^i_{\mathcal{R}}\big\},$$

$$R^4_{m,n} := \big(\Gamma_m P_{\theta_{m-1}} g_{\theta_{m-1}}(X_m) - \Gamma_{n+1} P_{\theta_n} g_{\theta_n}(X_{n+1})\big)\mathbb{I}\big\{A^n_{\mathcal{R}}\big\},$$

$$R^5_{m,n} := \sum_{i=m}^{n} (\Gamma_{i+1} - \Gamma_i) P_{\theta_{i-1}} g_{\theta_{i-1}}(X_i)\mathbb{I}\big\{A^{i-1}_{\mathcal{R}}\big\}.$$

The first term $R^1_{m,n}$ is a martingale, so by Doob's inequality, (4.2) and (4.3),

$$\Big(\mathbb{E}_{\theta,x}\Big[\sup_{n\geq m}\big|R^1_{m,n}\big|\Big]\Big)^2 \leq C\sum_{i=m}^{\infty}\mathbb{E}_{\theta,x}\big[\Gamma^2_{i+1}\big|g_{\theta_i}(X_{i+1}) - P_{\theta_i} g_{\theta_i}(X_i)\big|^2\mathbb{I}\big\{A^i_{\mathcal{R}}\big\}\big]$$

$$\leq CV^{2\beta_g}(x)\sum_{i=m}^{\infty}\mathbb{E}\big[\Gamma^2_{i+1}\big] \xrightarrow{m\to\infty} 0.$$

The claim for the second term is implied directly by (4.4). For the term $R^4_{m,n}$, it is enough to observe that

$$\mathbb{E}_{\theta,x}\Big[\sup_{n\geq m}\big(R^4_{m,n}\big)^2\Big] \leq 4\sum_{i=m}^{\infty}\mathbb{E}\big[\Gamma^2_i\big]\mathbb{E}_{\theta,x}\big[\big|P_{\theta_{i-1}} g_{\theta_{i-1}}(X_i)\big|^2\mathbb{I}\big\{A^{i-1}_{\mathcal{R}}\big\}\big] \leq CV^{2\beta_g}(x)\sum_{i=m}^{\infty}\mathbb{E}\big[\Gamma^2_i\big].$$

Finally, we may employ Lemma 3.4 for $R^5_{m,n}$ with $U_i := \Gamma_i$ and $B_{i-1} := |P_{\theta_{i-1}} g_{\theta_{i-1}}(X_i)|\mathbb{I}\{A^{i-1}_{\mathcal{R}}\}$ because $\mathbb{E}_{\theta,x}[|B_{i-1}|] \leq CV^{\beta_g}(x)$ and $\mathbb{E}_{\theta,x}[B^2_{i-1}] \leq CV^{2\beta_g}(x)$. □

Appendix C: Geometric ergodicity of IMH

We provide here quantitative bounds for the ergodicity constants for independent Metropolis–Hastings kernels. To our knowledge, the results here are new, and can be useful also in other settings.

Recall that the independent Metropolis–Hastings kernel with target density π and proposal density q on space $\mathsf{X} \subset \mathbb{R}^d$ is defined as

$$P(x, A) := \int_A \alpha(x, y)q(y)\,\mathrm{d}y + \mathbb{I}\{x \in A\}\bigg(1 - \int_{\mathsf{X}} \alpha(x, y)q(y)\,\mathrm{d}y\bigg)$$

for all $x \in \mathsf{X}$ and measurable $A \subset \mathsf{X}$, where the acceptance probability $\alpha(x, y)$ is defined as

$$\alpha(x, y) := \min\left\{1, \frac{\pi(y)/q(y)}{\pi(x)/q(x)}\right\}.$$

Proposition C.1. *Assume P is the independent Metropolis–Hastings kernel with target density π and proposal density q satisfying $\epsilon := \inf_{x\in\mathsf{X}} q(x)/\pi(x) > 0$. Let $V : \mathsf{X} \to [1, \infty)$ be a function with $q(V) < \infty$. Then,*

(i) *the drift inequality*

$$PV(x) \le \rho V(x) + q(V) \qquad \text{for all } x \in \mathsf{X}$$

holds with the constant $\rho := 1 - \epsilon$, and

(ii) *the following bound holds for any measurable function $f : \mathsf{X} \to \mathbb{R}^d$ with $\|f\|_V := \sup_{x\in\mathsf{X}} |f(x)|/V(x) < \infty$, all $k \ge 1$ and all $x \in \mathsf{X}$*

$$\left|P^k f(x) - \pi(f)\right| \le kM(1-\epsilon)^k \|f\|_V V(x),$$

where the constant $M = q(V)[1 + \epsilon^{-1} + (1-\epsilon)^{-1}]$.

Proof. Denote by $r(x) := \pi(x)/q(x)$ so that $\alpha(x, y) = \min\{1, r(x)/r(y)\}$ and compute

$$\frac{PV(x)}{V(x)} - 1 = \frac{\int V(y)\alpha(x, y)q(y)\,\mathrm{d}y}{V(x)} - \int \min\left\{r^{-1}(y), r^{-1}(x)\right\}\pi(y)\,\mathrm{d}y \le \frac{q(V)}{V(x)} - \epsilon.$$

This readily implies (i).

Observe then that for any measurable $A \subset \mathsf{X}$, the following uniform minorisation inequality holds

$$P(x, A) \ge \int_A \alpha(x, y)q(y)\,\mathrm{d}y \ge \epsilon\pi(A).$$

By this inequality, one can define a Markov kernel $Q(x, A) := (1-\epsilon)^{-1}(P(x, A) - \epsilon\pi(A))$. By (i), we have $QV(x) \le (1-\epsilon)^{-1}(\rho V(x) + q(V)) = V(x) + (1-\epsilon)^{-1}q(V)$ so by induction we obtain

$$Q^k V(x) \le V(x) + k(1-\epsilon)^{-1}q(V).$$

Observe that for any probability measure ν with $\nu(V) < \infty$, one has $\nu(|f|) \le \|f\|_V \nu(V)$, and that

$$\pi(V) = \int \frac{\pi(x)}{q(x)} V(x)q(x)\,\mathrm{d}x \le \epsilon^{-1}q(V).$$

Note that $\pi Q = \pi$, whence by denoting $\Pi(x, \cdot) := \pi(\cdot)$ one can compute for any $k \ge 1$

$$\begin{aligned}\left|P^k f(x) - \pi(f)\right| &= \left|(P - \Pi)P^{k-1} f(x)\right| = (1-\epsilon)\left|(Q - \Pi)P^{k-1} f(x)\right| \\ &= (1-\epsilon)\left|QP^{k-1} f(x) - \pi(f)\right| = \cdots = (1-\epsilon)^k \left|Q^k f(x) - \pi(f)\right| \\ &\le (1-\epsilon)^k \left(V(x) + k(1-\epsilon)^{-1} + \epsilon^{-1}\right)\|f\|_V q(V),\end{aligned}$$

establishing (ii). □

Corollary C.2. *In Proposition C.1, the bound (ii) can be replaced with the following*

$$\left|P^k f(x) - \pi(f)\right| \le M'(1-\zeta\epsilon)^k \|f\|_V V(x),$$

where $\zeta \in (0,1)$ *can be chosen arbitrarily and where*

$$M' = \frac{M}{\mathrm{e}}\left[\log\left(\frac{1-\zeta\epsilon}{1-\epsilon}\right)\right]^{-1}.$$

If $\epsilon \le 1/2$, *then* M' *can be taken as* $M' = 2M[\mathrm{e}(1-\zeta)\epsilon]^{-1}$.

Proof. From Proposition C.1, we obtain

$$\begin{aligned}\left|P^k f(x) - \pi(f)\right| &\le kM(1-\epsilon)^k \|f\|_V V(x)\\ &\le M'(1-\zeta\epsilon)^k \|f\|_V V(x),\end{aligned}$$

with

$$M' := M \sup_{k\ge 1} k\left(\frac{1-\epsilon}{1-\zeta\epsilon}\right)^k \le \frac{M}{\mathrm{e}}\left[\log\left(\frac{1-\zeta\epsilon}{1-\epsilon}\right)\right]^{-1},$$

since by a straightforward calculation one obtains for any $a \in (0,1)$ that $\sup_{x>0} xa^x = (\mathrm{e}\log(1/a))^{-1}$. Suppose then that $\epsilon \le 1/2$ and notice that for any $h > 0$ one has $\log(1+h) \ge h - \frac{1}{2}h^2$ and so

$$\log\left(\frac{1-\zeta\epsilon}{1-\epsilon}\right) \ge \frac{(1-\zeta)\epsilon}{1-\epsilon}\left(1 - \frac{1}{2}\frac{(1-\zeta)\epsilon}{1-\epsilon}\right) \ge \frac{1}{2}(1-\zeta)\epsilon.$$ □

Appendix D: Nomenclature

- α_w in Condition 2.1, page 549, related to the growth of $\sup_{\theta\in\mathcal{R}_i} |\nabla w(\theta)|$.
- α_H, β_H in Condition 3.1, page 556, characterise $\sup_{\theta\in\mathcal{R}_i} |H(\theta, x)|$.
- α_V, β_V in Condition 3.1, page 556, characterise $\mathbb{E}_{\theta,x}[V(X_i)]$.
- α_g, β_g in Condition 3.1, page 556, characterise $\sup_{\theta\in\mathcal{R}_i} [|g_\theta(x)| + |P_\theta g_\theta(x)|]$.
- β_D in Condition 3.12, page 564, characterises the Hölder continuity of $\|P_\theta f - P_{\theta'} f\|_{V^r}$.
- $\alpha_\Delta, \beta_\Delta$ in Condition 3.15, page 565, characterise the size of $\sup_{(\theta,\theta')\in\mathcal{R}_i^2} \|H(\theta,\cdot) - H(\theta',\cdot)\|_{V^{\beta_H}}$.
- α_M and α_ρ are defined in Condition 3.8, page 562, and characterise the loss of ergodicity through the growth of geometric ergodicity constants $\sup_{\theta\in\mathcal{R}_i} M_{\theta,r}$ and $\sup_{\theta\in\mathcal{R}_i}(1-\rho_{\theta,r})^{-1}$, respectively.

Acknowledgements

We thank Harriet Bass and the referees for helpful comments. The work of the first author was supported in part by an EPSRC advance research fellowship and a Winton Capital research award. The second author was supported by the Academy of Finland Project 250575, by the Finnish Academy of Science and Letters, Vilho, Yrjö and Kalle Väisälä Foundation, by the Finnish Centre of Excellence in Analysis and Dynamics Research, and by the Finnish Doctoral Programme in Stochastics and Statistics.

References

[1] Andradóttir, S. (1995). A stochastic approximation algorithm with varying bounds. *Oper. Res.* **43** 1037–1048. MR1488889

[2] Andrieu, C., Doucet, A. and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** 269–342. MR2758115

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

[4] Andrieu, C., Moulines, É. and Priouret, P. (2005). Stability of stochastic approximation under verifiable conditions. *SIAM J. Control Optim.* **44** 283–312. MR2177157

[5] Andrieu, C., Moulines, É. and Volkov, S. (2004). Convergence of stochastic approximation for Lyapunov stable dynamics: A proof from first principles Technical report.

[6] Andrieu, C., Tadić, V.B. and Vihola, M. (2012). On the stability of controlled Markov chains and its applications to stochastic approximation with Markovian dynamic. Available at arXiv:1205.4181v1.

[7] Benaïm, M. (1999). Dynamics of stochastic approximation algorithms. In *Séminaire de Probabilités, XXXIII. Lecture Notes in Math.* **1709** 1–68. Berlin: Springer. MR1767993

[8] Benveniste, A., Métivier, M. and Priouret, P. (1990). *Adaptive Algorithms and Stochastic Approximations. Applications of Mathematics* (*New York*) **22**. Berlin: Springer. Translated from the French by Stephen S. Wilson. MR1082341

[9] Borkar, V.S. (2008). *Stochastic Approximation*: *A Dynamical Systems Viewpoint*. Cambridge: Cambridge Univ. Press. MR2442439

[10] Chan, K.S. and Ledolter, J. (1995). Monte Carlo EM estimation for time series models involving counts. *J. Amer. Statist. Assoc.* **90** 242–252. MR1325132

[11] Chen, H.-F. (2002). *Stochastic Approximation and Its Applications. Nonconvex Optimization and Its Applications* **64**. Dordrecht: Kluwer Academic. MR1942427

[12] Chen, H.F., Lei, G. and Gao, A.J. (1988). Convergence and robustness of the Robbins–Monro algorithm truncated at randomly varying bounds. *Stochastic Process. Appl.* **27** 217–231. MR0931029

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

[14] Delyon, B., Lavielle, M. and Moulines, E. (1999). Convergence of a stochastic approximation version of the EM algorithm. *Ann. Statist.* **27** 94–128. MR1701103

[15] Donnet, S. and Samson, A. (2011). EM algorithm coupled with particle filter for maximum likelihood parameter estimation of stochastic differential mixed-effect models. Technical Report hal-00519576 v2, Universite Paris Descartes MAP5.

[16] Fort, G. and Moulines, E. (2003). Convergence of the Monte Carlo expectation maximization for curved exponential families. *Ann. Statist.* **31** 1220–1259. MR2001649

[17] Fort, G., Moulines, E., Roberts, G.O. and Rosenthal, J.S. (2003). On the geometric ergodicity of hybrid samplers. *J. Appl. Probab.* **40** 123–146. MR1953771

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

[19] Kamal, S. (2012). Stabilization of stochastic approximation by step size adaptation. *Systems Control Lett.* **61** 543–548. MR2910330

[20] Kushner, H.J. and Clark, D.S. (1978). *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. *Applied Mathematical Sciences* **26**. New York: Springer. MR0499560

[21] Kushner, H.J. and Yin, G.G. (2003). *Stochastic Approximation and Recursive Algorithms and Applications*, 2nd ed. *Applications of Mathematics* (*New York*): *Stochastic Modelling and Applied Probability* **35**. New York: Springer. MR1993642

[22] Mengersen, K.L. and Tweedie, R.L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24** 101–121. MR1389882

[23] Meyn, S.P. and Tweedie, R.L. (1994). Computable bounds for geometric convergence rates of Markov chains. *Ann. Appl. Probab.* **4** 981–1011. MR1304770

[24] Roberts, G.O. and Rosenthal, J.S. (2007). Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. *J. Appl. Probab.* **44** 458–475. MR2340211

[25] Saksman, E. and Vihola, M. (2010). On the ergodicity of the adaptive Metropolis algorithm on unbounded domains. *Ann. Appl. Probab.* **20** 2178–2203. MR2759732

[26] Sharia, T. (1997). Truncated recursive estimation procedures. *Proc. A. Razmadze Math. Inst.* **115** 149–159. MR1639120

[27] Sharia, T. (2011). Truncated stochastic approximation with moving bounds: Convergence. Available at arXiv:1101.0031v3.

[28] Tadić, V. (1998). Stochastic approximation with random truncations, state-dependent noise and discontinuous dynamics. *Stochastics Stochastics Rep.* **64** 283–326. MR1709288

[29] Vihola, M. (2011). On the stability and ergodicity of adaptive scaling Metropolis algorithms. *Stochastic Process. Appl.* **121** 2839–2860. MR2844543

[30] Younes, L. (1999). On the convergence of Markovian stochastic algorithms with rapidly decreasing ergodicity rates. *Stochastics Stochastics Rep.* **65** 177–228. MR1687636

[31] Zeger, S.L. (1988). A regression model for time series of counts. *Biometrika* **75** 621–629. MR0995107

Received November 2011 and revised September 2012