Bernoulli **20**(2), 2014, 457–485
DOI: 10.3150/12-BEJ493

A central limit theorem for adaptive and interacting Markov chains

G. FORT[1,*], E. MOULINES[1,**], P. PRIOURET[2] and P. VANDEKERKHOVE[3]

[1]*LTCI, TELECOM ParisTech & CNRS, 46 rue Barrault, 75634 Paris Cedex 13, France.*
E-mail: [*]*gersende.fort@telecom-paristech.fr;* [**]*eric.moulines@telecom-paristech.fr*
[2]*LPMA, Univ. Pierre et Marie Curie, Boîte courrier 188, 75252 Paris Cedex 05, France.*
E-mail: priouret@ccr.jussieu.fr
[3]*LAMA, Univ. Paris-Est Marne-la-Vallée, Cité Descartes, 5 Bd. Descartes, Champs-sur-Marne, 77454 Marne-la-Vallée Cedex 2, France. E-mail: vandek@univ-mlv.fr*

Adaptive and interacting Markov Chains Monte Carlo (MCMC) algorithms are a novel class of non-Markovian algorithms aimed at improving the simulation efficiency for complicated target distributions. In this paper, we study a general (non-Markovian) simulation framework covering both the adaptive and interacting MCMC algorithms. We establish a central limit theorem for additive functionals of unbounded functions under a set of verifiable conditions, and identify the asymptotic variance. Our result extends all the results reported so far. An application to the interacting tempering algorithm (a simplified version of the equi-energy sampler) is presented to support our claims.

Keywords: interacting MCMC; limit theorems; MCMC

1. Introduction

Markov chain Monte Carlo (MCMC) methods generate samples from distributions known up to a scaling factor.

In the last decade, several non-Markovian simulation algorithms have been proposed. In the so-called adaptive MCMC algorithm, the transition kernel of the MCMC algorithm depends on a finite dimensional *parameter* which is updated at each iteration from the past values of the chain and the parameters. The prototypical example is the adaptive Metropolis algorithm, introduced in [21] (see [32] and the references therein for recent references). Many other examples of adaptive MCMC algorithms are presented in the survey papers by Andrieu and Thoms [6], Rosenthal [31], Atchadé *et al.* [8].

In the so-called *Interacting MCMC*, several processes are simulated in parallel, each targeting different distribution. Each process might interact with the whole past of its neighboring processes. A prototypical example is the equi-energy sampler introduced in [24], where the different processes target a tempered version of the target distribution. The convergence of this algorithm has been considered in a series of papers by Andrieu *et al.* [1,2,4] and in [18]. Different variants of the interacting MCMC algorithm have been later introduced and studied in [12,14] and [13]. These algorithms are so far limited to specific scenarios, and the assumptions used in these papers preclude the applications of their results in the applications considered in this paper.

The analysis of the convergence of these algorithms is involved. Whereas the basic building blocks of these simulation algorithms are Markov kernels, the processes generated by these techniques are no longer Markovian. Indeed, each individual process either interacts with its distant past, or the distant past of some auxiliary processes.

The ergodicity and the consistency of additive functionals for adaptive and interacting Markov Chains have been considered in several recent papers: see [18] and the references therein. Up to now, there are much fewer works addressing central limit theorems (CLT). In [5] the authors establish the asymptotic normality of additive functionals for a special class of adaptive MCMC algorithms in which a finite dimensional parameter is adapted using a stochastic approximation procedure. Atchadé [10] established a CLT for general adaptive MCMC samplers under stronger conditions than in [5], by assuming simultaneous ergodicity of the transition kernels involved in the adaptive algorithm. Some of the theoretical limitations of Andrieu and Moulines [5] have been alleviated by Saksman and Vihola [32] for the so-called adaptive Metropolis algorithm, which established a CLT for additive functionals for the Adaptive Metropolis algorithm (with a proof specially tailored for this algorithm). The results presented in this contribution contain as special cases these three earlier results.

The theory for interacting MCMC algorithms is up to now quite limited, despite the clear potential of this class of methods to sample complicated multimodal target distributions. The law of large numbers for additive functionals have been established in [3] for some specific interacting algorithm. A wider class of interacting Markov chains has been considered in [14]. This paper establishes the consistency of a form of interacting tempering algorithm and provides non-asymptotic L^p-inequalities. The assumptions under which the results are derived are restrictive and the results do not cover the interacting MCMC algorithms considered in this paper. More recently, Fort *et al.* [18] have established the ergodicity and law of large numbers for a wide class of interacting MCMC, under the weakest conditions known so far.

A functional CLT was derived in [12] for a specific class of interacting Markov Chains but their assumptions do not cover the interactive MCMC considered in this paper (and in particular, the interacting MCMC algorithm). A CLT for additive functionals is established by Atchadé [9] for the interacting tempering algorithm; the proof of the main result in this paper, Theorem 3.3, contains a serious gap (page 865) which seems difficult to correct, see the supplemental material [20].

This paper aims at providing a theory removing the limitations mentioned above and covering both adaptive and interacting MCMC in a common unifying framework. The paper is organized as follows. In Section 2, we derive our main theorem (Theorem 2.3) which establishes CLTs for adaptive and interacting MCMC algorithms. These results are applied in Section 3.2 to the 2-chain interacting tempering algorithm which is a simplified version of the Equi-Energy sampler. All the proofs are postponed in Section 4.

Notations

Let $(\mathsf{X}, \mathcal{X})$ be a general state space and P be a Markov transition kernel (see, e.g., [27], Chapter 3). P acts on bounded functions f on X and on σ-finite positive measures μ on $\mathcal{X}$ via

$$Pf(x) \stackrel{\text{def}}{=} \int P(x, \mathrm{d}y) f(y), \qquad \mu P(A) \stackrel{\text{def}}{=} \int \mu(\mathrm{d}x) P(x, A).$$

We denote by P^n the n-iterated transition kernel defined inductively

$$P^n(x, A) \stackrel{\text{def}}{=} \int P^{n-1}(x, \mathrm{d}y) P(y, A) = \int P(x, \mathrm{d}y) P^{n-1}(y, A);$$

where P^0 is the identity kernel. For a function $V : \mathsf{X} \to [1, +\infty)$, define the V-norm of a function $f : \mathsf{X} \to \mathbb{R}$ by

$$|f|_V \stackrel{\text{def}}{=} \sup_{x \in \mathsf{X}} \frac{|f|(x)}{V(x)}.$$

When $V = 1$, the V-norm is the supremum norm denoted by $|f|_\infty$. Let $\mathcal{L}_V$ be the set of measurable functions such that $|f|_V < +\infty$. For μ a finite signed measure on $(\mathsf{X}, \mathcal{X})$ and $V : \mathsf{X} \to [1, \infty)$ such that $|\mu|(V) < \infty$ where $|\mu|$ is the variation of μ, we define $\|\mu\|_V$ the V-norm of μ as

$$\|\mu\|_V = \sup_{f \in \mathcal{L}_V, |f|_V \le 1} \bigl|\mu(f)\bigr|.$$

When $V \equiv 1$, the V-norm corresponds to the total variation norm.

For finite signed kernels P on $(\mathsf{X}, \mathcal{X})$ and $V : \mathsf{X} \to [1, \infty)$ such that $|P(x, \cdot)|(V) < \infty$ for any $x \in \mathsf{X}$, define

$$\|P\|_V \stackrel{\text{def}}{=} \sup_{x \in \mathsf{X}} V^{-1}(x) \bigl\| P(x, \cdot) \bigr\|_V. \tag{1}$$

Let $(x_n)_{n \in \mathbb{N}}$ be a sequence. For $p \le q \in \mathbb{N}^2$, $x_{p:q}$ denotes the vector $(x_p, \dots, x_q)$.

2. Main results

Let $(\Theta, \mathcal{T})$ be a measurable space. Let $\{P_\theta, \theta \in \Theta\}$ be a collection of Markov transition kernels on $(\mathsf{X}, \mathcal{X})$ indexed by a parameter $\theta \in \Theta$. From here on, it is assumed that for any $A \in \mathcal{X}$, $(x, \theta) \mapsto P_\theta(x, A)$ is $\mathcal{X} \otimes \mathcal{T} / \mathcal{B}([0, 1])$ measurable, where $\mathcal{B}([0, 1])$ denotes the Borel σ-field. From here on, Θ is not necessarily a finite-dimensional vector space. It might be a function space or a space of measures. We consider a $\mathsf{X} \times \Theta$-valued process $\{(X_n, \theta_n)\}_{n \in \mathbb{N}}$ on a filtered probability space $(\Omega, \mathcal{A}, \{\mathcal{F}_n, n \ge 0\}, \mathbb{P})$. It is assumed that

A1 The process $\{(X_n, \theta_n)\}_{n \in \mathbb{N}}$ is $(\mathcal{F}_n)_{n \in \mathbb{N}}$-adapted and for any bounded measurable function h,

$$\mathbb{E}\bigl[h(X_{n+1}) | \mathcal{F}_n\bigr] = P_{\theta_n} h(X_n). \tag{2}$$

Assumption A1 implies that conditional to the past (subsumed in the σ-algebra $\mathcal{F}_n$), the distribution of the next sample X_{n+1} is governed by the current value X_n and the current parameter θ_n. This assumption covers any adaptive and interacting MCMC algorithms; see [6,8,18] for examples. This assumption on the adaptation of the parameter $(\theta_n)_{n \in \mathbb{N}}$ is quite weak since it only requires the parameter to be adapted to the filtration. In practice, it frequently occurs that the joint process $\{(X_n, \theta_n)\}_{n \in \mathbb{N}}$ is Markovian but assumption A1 covers more general adaptation rules.

We assume that the transition kernels $\{P_\theta, \theta \in \Theta\}$ satisfy a Lyapunov drift inequality and smallness conditions:

A2 For all $\theta \in \Theta$, P_θ is phi-irreducible, aperiodic and there exists a function $V : \mathsf{X} \to [1, +\infty)$, and for any $\theta \in \Theta$ there exist some constants $b_\theta \in (1, +\infty)$, $\lambda_\theta \in (0, 1)$ such that for any $x \in \mathsf{X}$,

$$P_\theta V(x) \le \lambda_\theta V(x) + b_\theta.$$

In addition, for any $d \ge 1$ and any $\theta \in \Theta$, the level sets $\{V \le d\}$ are m-small for P_θ that is, for any $\theta \in \Theta$, there exist $\kappa_\theta > 0$ and a probability ν_θ such that for all $x \in \{V \le d\}$, $P_\theta^m(x, A) \ge \kappa_\theta \nu_\theta(A)$ for all $A \in \mathcal{X}$.

In many examples considered so far (see [4,5,18,32]), this condition is satisfied. All the results below can be established under assumptions insuring that the drift inequality and/or the smallness condition are satisfied for some m-iterated P_θ^m. Note that checking assumption on the iterated kernel P_θ^m is prone to be difficult because the expression of the m-iterated kernel is most often rather involved.

A2 implies that, for any $\theta \in \Theta$, P_θ possesses an invariant probability distribution π_θ and the kernel P_θ is geometrically ergodic [27], Chapter 15. The following lemma summarizes the properties of the family $\{P_\theta, \theta \in \Theta\}$ used hereafter (see, e.g., [16] and references therein for the explicit control of ergodicity; and [27], Proposition 17.4.1 for the Poisson equation). For $\theta \in \Theta$, denote by Λ_θ the operator which associates to any function $f \in \mathcal{L}_{V^\alpha}$ the function $\Lambda_\theta f$ given by:

$$\Lambda_\theta f \stackrel{\text{def}}{=} \sum_{n \ge 0} P_\theta^n f - \pi_\theta(f). \tag{3}$$

Lemma 2.1. *Assume* A2. *Then for any* $\theta \in \Theta$, *there exists a probability distribution* π_θ *such that* $\pi_\theta P_\theta = \pi_\theta$ *and* $\pi_\theta(V) \le b_\theta(1 - \lambda_\theta)^{-1}$. *In addition, for any* $\alpha \in (0, 1]$, *the following property holds.*

P[α] *For any* $\theta \in \Theta$, *there exist* $C_\theta < \infty$ *and* $\rho_\theta \in (0, 1)$ *such that, for any* $\gamma \in [\alpha, 1]$,

$$\left\| P_\theta^n - \pi_\theta \right\|_{V^\gamma} \le C_\theta \rho_\theta^n.$$

For any $\alpha \in (0, 1)$ *and* $f \in \mathcal{L}_{V^\alpha}$, *the function* $\Lambda_\theta f$ *exists and is in* $\mathcal{L}_{V^\alpha}$. *The function* $\Lambda_\theta f$ *is the unique solution up to an additive constant of the Poisson equation*

$$\Lambda_\theta f - P_\theta \Lambda_\theta f = f - \pi_\theta(f). \tag{4}$$

It has been shown in [18], that under appropriate assumptions, when the sequence $(\theta_k)_{k \in \mathbb{N}}$ converges to $\theta_\star \in \Theta$ in an appropriate sense, $n^{-1} \sum_{k=1}^n f(X_k)$ converges almost surely to $\pi_{\theta_\star}(f)$, for any functions f belonging to a suitable class of functions $\mathcal{M}$. The objective of this paper is to derive a CLT for $n^{-1/2} \sum_{k=1}^n \{f(X_k) - \pi_{\theta_\star}(f)\}$ for functions f belonging to $\mathcal{M}$. To that goal, consider the following decomposition

$$n^{-1/2} \sum_{k=1}^n \bigl\{f(X_k) - \pi_{\theta_\star}(f)\bigr\} = S_n^{(1)}(f) + S_n^{(2)}(f),$$

where $S_n^{(1)}(f)$ and $S_n^{(2)}(f)$ are given by

$$S_n^{(1)}(f) \stackrel{\text{def}}{=} n^{-1/2} \sum_{k=1}^{n} \{ f(X_k) - \pi_{\theta_{k-1}}(f) \}, \tag{5}$$

$$S_n^{(2)}(f) \stackrel{\text{def}}{=} n^{-1/2} \sum_{k=0}^{n-1} \{ \pi_{\theta_k}(f) - \pi_{\theta_\star}(f) \}. \tag{6}$$

We consider these two terms separately. For the first term, we use a classical technique based on the Poisson decomposition; this amounts to writing $S_n^{(1)}(f)$ as the sum of a martingale difference and of a remainder term converging to zero in probability; see [5,7,14,18,32] for law of large numbers for adaptive and interacting MCMC. Then we apply a classical CLT for martingale difference array; see, for example, [22], Theorem 3.2.

The second term vanishes when $\pi_\theta = \pi_{\theta_\star}$ for all $\theta \in \Theta$ which is the case for example, for the adaptive Metropolis algorithm [21]. In scenarios where $\theta \mapsto \pi_\theta$ is a non-trivial function of θ, the weak convergence $S_n^{(2)}(f)$ relies on conditions which are quite problem specific. The application detailed in Section 3.2, an elementary version of the interacting tempering algorithm, is a situation in which $\pi_{\theta_\star}$ is known but the expression of π_θ, $\theta \neq \theta_\star$, is unknown, except in very simple examples. The Wang–Landau algorithm [25,36] is an example of adaptive MCMC algorithm in which $\theta \mapsto \pi_\theta$ is explicit. The results in this paper cover the case when the expression of π_θ is unknown: we rewrite $S_n^{(2)}(f)$ by showing that the leading term of the difference $\pi_{\theta_k}(f) - \pi_{\theta_\star}(f)$ is $\pi_{\theta_\star}(P_{\theta_k} - P_{\theta_\star})\Lambda_{\theta_\star} f$ where $\Lambda_{\theta_\star}$ is the operator defined by (3). Our approach covers much more general set-up than the one outlined in [12].

The convergence of $S_n^{(1)}(f)$ is addressed under the following assumptions which are related to the regularity in the parameter $\theta \in \Theta$ of the ergodic behavior of the kernels $\{P_\theta, \theta \in \Theta\}$.

A3 There exist $\alpha \in (0, 1/2)$ and a subset of measurable functions $\mathcal{M}_{V^\alpha} \subseteq \mathcal{L}_{V^\alpha}$ satisfying the two following conditions
(a) for any $f \in \mathcal{M}_{V^\alpha}$,

$$n^{-1/2} \sum_{k=1}^{n} |P_{\theta_k} \Lambda_{\theta_k} f - P_{\theta_{k-1}} \Lambda_{\theta_{k-1}} f|_{V^\alpha} V^\alpha(X_k) \xrightarrow{\mathbb{P}} 0.$$

(b) $n^{-1/2\alpha} \sum_{k=0}^{n-1} L_{\theta_k}^{2/\alpha} P_{\theta_k} V(X_k) \xrightarrow{\mathbb{P}} 0$ where L_θ is defined by (9) for the constants C_θ, ρ_θ given by P[α].

A3(a) controls the regularity in the parameter θ of the Poisson solution $\Lambda_\theta f$. By [18], Lemma 4.2,

$$\|P_\theta \Lambda_\theta - P_{\theta'} \Lambda_{\theta'}\|_{V^\alpha} \le 5(L_\theta \vee L_{\theta'})^6 \pi_\theta(V^\alpha) D_{V^\alpha}(\theta, \theta'), \tag{7}$$

where

$$D_V(\theta, \theta') \stackrel{\text{def}}{=} \|P_\theta - P_{\theta'}\|_V, \tag{8}$$

$$L_\theta \stackrel{\text{def}}{=} C_\theta \vee (1 - \rho_\theta)^{-1}, \tag{9}$$

and $\|P_\theta - P_{\theta'}\|_V$ is defined by (1) and C_θ and ρ_θ are introduced in Lemma 2.1. This upper bound relates the regularity in θ of the function $\theta \mapsto P_\theta \Lambda_\theta f$ to the ergodicity constants C_θ and ρ_θ and to the regularity in θ of the function $\theta \mapsto P_\theta$ from the parameter space Θ to the space of Markov transition kernels equipped with the V-operator norm. Therefore, A3(a) corresponds to a diminishing adaptation condition (see [29]).

A3(b) is a kind of containment condition (see [29]): when the ergodic behavior A2 is uniform in θ so that λ_θ, b_θ and the minorization constant of the P_θ-smallness condition do not depend on θ, then the constant L_θ does not depend on θ and by A1 and the drift inequality A2,

$$n^{-1/2\alpha} \sum_{k=0}^{n-1} \mathbb{E}\big[V(X_{k+1})\big] \le n^{1-1/2\alpha} \big\{\mathbb{E}\big[V(X_0)\big] + (1-\lambda)^{-1} b\big\} \to 0.$$

Therefore, condition A3(b) holds provided the ergodic constant L_{θ_k} is controlled by a slowly-increasing function of k. Lemma A.2 in Appendix A provides sufficient conditions to obtain upper bounds of $\theta \mapsto L_\theta$ in terms of the constants appearing in the drift inequality A2.

We finally introduce a condition allowing to obtain a closed-form expression for the asymptotic variance of $S_n^{(1)}(f)$. For $\theta \in \Theta$ and $f \in \mathcal{L}_{V^\alpha}$ define

$$F_\theta \stackrel{\text{def}}{=} P_\theta (\Lambda_\theta f)^2 - [P_\theta \Lambda_\theta f]^2. \tag{10}$$

A4 For any $f \in \mathcal{M}_{V^\alpha}$, $n^{-1} \sum_{k=0}^{n-1} F_{\theta_k}(X_k) \xrightarrow{\mathbb{P}} \sigma^2(f)$, where $\sigma^2(f)$ is a deterministic constant.

Assumption A4 is typically established by using the Law of Large Numbers (LLN) for adaptive and interacting Markov Chain derived in [18]; see also Theorem B.1 in Appendix B. Under appropriate regularity conditions on the Markov kernels $\{P_\theta, \theta \in \Theta\}$, it is proved that $n^{-1} \sum_{k=0}^{n-1} \{F_{\theta_k}(X_k) - \int \pi_{\theta_k}(\mathrm{d}x) F_{\theta_k}(x)\}$ converges in probability to zero. The second step consists in showing that $n^{-1} \sum_{k=0}^{n-1} \int \pi_{\theta_k}(\mathrm{d}x) F_{\theta_k}(x)$ converges to a (deterministic) constant $\sigma^2(f)$: when π_θ is not explicitly known and the set X is Polish, Lemma A.3 in Appendix A is useful to check this convergence. In practice, this may introduce a restriction of the set of functions $f \in \mathcal{L}_{V^\alpha}$ for which this limit holds (see, e.g., the example detailed in Section 3.2 where $\mathcal{M}_{V^\alpha} \neq \mathcal{L}_{V^\alpha}$).

We can now state conditions upon which $S_n^{(1)}(f)$ is asymptotically normal.

Theorem 2.2. *Assume A1 to A4. For any $f \in \mathcal{M}_{V^\alpha}$,*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \big\{ f(X_k) - \pi_{\theta_{k-1}}(f) \big\} \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, \sigma^2(f)\big).$$

The proof is in Section 4.1.1. When $\pi_\theta = \pi$ for any θ, Theorem 2.2 provides sufficient conditions for a CLT for additive functionals to hold.

When π_θ is a function of $\theta \in \Theta$, we need now to obtain a joint CLT for $(S_n^{(1)}(f), S_n^{(2)}(f))$ (see (5) and (6)). To that goal, we replace A1 by the following assumption which implies that, conditionally to the process $(\theta_k)_{k\in\mathbb{N}}$, $(X_k)_{k\in\mathbb{N}}$ is an inhomogeneous Markov chain with transition kernels $(P_{\theta_j}, j \geq 0)$:

A5 There exists an initial distribution ν such that for any bounded measurable function $f : \mathsf{X}^{n+1} \to \mathbb{R}$,

$$\mathbb{E}\left[f(X_{0:n})|\theta_{0:n}\right] = \int \cdots \int \nu(\mathrm{d}x_0) f(x_{0:n}) \prod_{j=1}^{n} P_{\theta_{j-1}}(x_{j-1}, \mathrm{d}x_j).$$

Assumption A5 is satisfied when $\{(X_n, \theta_n)\}_{n\in\mathbb{N}}$ is an interacting MCMC algorithm. Note that A5 implies A1.

The first step in the proof of the joint CLT consists in linearizing the difference $\pi_{\theta_n} - \pi_{\theta_\star}$. Under A2, $\pi_\theta(g)$ exists for any $g \in \mathcal{L}_{V^\alpha}$ and $\theta \in \Theta$ (see Lemma 2.1), and we have

$$\pi_\theta(g) - \pi_{\theta_\star}(g) = \pi_\theta P_\theta g - \pi_{\theta_\star} P_{\theta_\star} g = \pi_\theta(P_\theta - P_{\theta_\star})g + (\pi_\theta - \pi_{\theta_\star})P_{\theta_\star} g,$$

which implies that $(\pi_\theta - \pi_{\theta_\star})(\mathrm{I} - P_{\theta_\star})g = \pi_\theta(P_\theta - P_{\theta_\star})g$. Let $f \in \mathcal{L}_{V^\alpha}$. Then $\Lambda_{\theta_\star} f \in \mathcal{L}_{V^\alpha}$ and by applying the previous equality with $g = \Lambda_{\theta_\star} f$, we have by (4)

$$\pi_\theta(f) - \pi_{\theta_\star}(f) = \pi_\theta(P_\theta - P_{\theta_\star})\Lambda_{\theta_\star} f. \tag{11}$$

We can iterate this decomposition, writing

$$\pi_\theta(f) - \pi_{\theta_\star}(f) = \pi_{\theta_\star}(P_\theta - P_{\theta_\star})\Lambda_{\theta_\star} f + \pi_\theta\big((P_\theta - P_{\theta_\star})\Lambda_{\theta_\star} f\big) - \pi_{\theta_\star}\big((P_\theta - P_{\theta_\star})\Lambda_{\theta_\star} f\big).$$

Applying again (11), we obtain

$$\pi_\theta(f) - \pi_{\theta_\star}(f) = \pi_{\theta_\star}(P_\theta - P_{\theta_\star})\Lambda_{\theta_\star} f + \pi_\theta(P_\theta - P_{\theta_\star})\Lambda_{\theta_\star}(P_\theta - P_{\theta_\star})\Lambda_{\theta_\star} f.$$

The first term in the RHS of the previous equation is the leading term of the error $\pi_{\theta_k} - \pi_{\theta_\star}$, whereas the second term is a remainder. This decomposition naturally leads to the following assumption.

A6 For any function $f \in \mathcal{M}_{V^\alpha}$,
(a) there exists a positive constant $\gamma^2(f)$ such that

$$n^{-1/2} \sum_{k=1}^{n} \pi_{\theta_\star}(P_{\theta_k} - P_{\theta_\star})\Lambda_{\theta_\star} f \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, \gamma^2(f)\big). \tag{12}$$

(b) $n^{-1/2} \sum_{k=1}^{n} \pi_{\theta_k}(P_{\theta_k} - P_{\theta_\star})\Lambda_{\theta_\star}(P_{\theta_k} - P_{\theta_\star})\Lambda_{\theta_\star} f \xrightarrow{\mathbb{P}} 0$.

Theorem 2.3. *Assume A2 to A6. For any function $f \in \mathcal{M}_{V^\alpha}$,*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \{f(X_k) - \pi_{\theta_\star}(f)\} \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, \sigma^2(f) + \gamma^2(f)\big).$$

The proof of Theorem 2.3 is postponed to Section 4.1.2. It is worthwhile to note that, as a consequence of A5, the variance is additive. This result extends Bercu *et al.* [12] which addresses the case when $P_\theta(x, A) = P_\theta(A)$, that is, the case when conditionally to the adaptation process $(\theta_n)_{n\in\mathbb{N}}$, the random variables $(X_n)_{n\in\mathbb{N}}$ are independent (see [12], Equation (1.4)). Our result, applied in this simpler situation, yields the same asymptotic variance.

3. Applications

3.1. Adaptive Metropolis (after Saksman and Vihola [32])

In this example, $\mathsf{X} = \mathbb{R}^d$ and the densities are assumed to be w.r.t. the Lebesgue measure. For $x \in \mathbb{R}^d$, $|x|$ denotes the Euclidean norm. For $\kappa > 0$, let $\mathcal{C}_\kappa^d$ be the set of symmetric and positive definite $d \times d$ matrices whose minimal eigenvalue is larger than κ. The parameter set $\Theta = \mathbb{R}^d \times \mathcal{C}_\kappa^d$ is endowed with the norm $|\theta|^2 \stackrel{\text{def}}{=} |\mu|^2 + \operatorname{Tr}(\Gamma^T\Gamma)$, where $\theta = (\mu, \Gamma)$.

At each iteration, $X_{n+1} \sim P_{\theta_n}(X_n, \cdot)$, where P_θ is defined by

$$
\begin{aligned}
P_\theta(x, A) \stackrel{\text{def}}{=} & \int_A \left(1 \wedge \frac{\pi(y)}{\pi(x)}\right) q_\Gamma(y - x)\,\mathrm{d}y \\
& + \mathbb{1}_A(x)\left[1 - \int \left(1 \wedge \frac{\pi(y)}{\pi(x)}\right) q_\Gamma(y - x)\,\mathrm{d}y\right],
\end{aligned} \tag{13}
$$

with q_Γ the density of a Gaussian random variable with zero mean and covariance matrix $(2.38)^2 d^{-1}\Gamma$, and π is a density on $\mathbb{R}^d$. The parameter $\theta_n = (\mu_n, \Gamma_n) \in \Theta$ is the sample mean and covariance matrix

$$
\mu_{n+1} = \mu_n + \frac{1}{n+1}(X_{n+1} - \mu_n), \qquad \mu_0 = 0, \tag{14}
$$

$$
\Gamma_{n+1} = \frac{n}{n+1}\Gamma_n + \frac{1}{n+1}\left\{(X_{n+1} - \mu_n)(X_{n+1} - \mu_n)^T + \kappa \mathrm{I}_d\right\}, \tag{15}
$$

where I_d is the identity matrix, $\Gamma_0 \geq 0$ and κ is a positive constant.

By construction, for any $\theta \in \Theta$, π is the stationary distribution for P_θ so that $\pi_\theta = \pi$ for any θ. As in [32], we consider the following assumption:

M1 π is positive, bounded, differentiable and

$$
\lim_{r\to\infty} \sup_{|x|\geq r} \frac{x}{|x|^\rho} \cdot \nabla \log \pi(x) = -\infty
$$

for some $\rho > 1$. Moreover, π has regular contours, that is, for some $R > 0$,

$$
\sup_{|x|\geq R} \frac{x}{|x|} \cdot \frac{\nabla\pi(x)}{|\nabla\pi(x)|} < 0.
$$

Saksman and Vihola [32], Proposition 15, establishes A2: the drift function V is proportional to π^{-s}, (for any) $s \in (0,1)$; the constant b_θ does not depend upon θ; and any level set of V is 1-small for P_θ. Saksman and Vihola [32], Propositions 15, also establishes that there exists a non-negative constant C such that for any $\theta \in \Theta$,

$$\kappa_\theta^{-1} \vee (1-\lambda_\theta)^{-1} \le C|\theta|^{d/2}.$$

This upper bound combined with [18], Lemma 2.3, implies that there exist finite constants C and γ such that for any $\theta \in \Theta$,

$$L_\theta \le C|\theta|^\gamma, \tag{16}$$

where L_θ is defined by (9).

We now prove that A3 holds. Let $\alpha \in (0,1/2)$ and set $\mathcal{M}_{V^\alpha} = \mathcal{L}_{V^\alpha}$. By (7) and (16), there exist positive constants $C, \bar{\gamma}$ such that for any $f \in \mathcal{L}_{V^\alpha}$,

$$\begin{aligned} n^{-1/2} &\sum_{k=1}^n |P_{\theta_k}\Lambda_{\theta_k} f - P_{\theta_{k-1}}\Lambda_{\theta_{k-1}} f|_{V^\alpha} V^\alpha(X_k) \\ &\le cn^{-1/2}\sum_{k=1}^n \bigl(1+|\theta_k|+|\theta_{k-1}|\bigr)^{\bar{\gamma}} D_{V^\alpha}(\theta_k,\theta_{k-1}) V^\alpha(X_k). \end{aligned}$$

In [32], Lemma 12, it is proved that under M1, the rate of growth of the parameters $\{\theta_n, n \ge 0\}$ is controlled. Namely, for any $\tau > 0$,

$$\sup_{n\ge 1} n^{-\tau}|\theta_n| < +\infty, \qquad \mathbb{P}\text{-a.s.} \tag{17}$$

In addition, it is established in [18], Equation (12), that there exists a constant $C < \infty$ such that for any $n \ge 1$,

$$D_{V^\alpha}(\theta_n,\theta_{n-1}) \le \frac{C}{n}\left\{1 + \frac{\ln n}{n-1}\sum_{j=1}^{n-1} \ln^2 V(X_j) + \ln n\bigl(\ln^2 V(X_n) + \ln^2 V(X_{n-1})\bigr)\right\}.$$

Combining the above results show that A3(a) holds provided

$$\frac{1}{\sqrt{n}}\sum_{k=2}^n \frac{\ln k}{k^{1-\tau\bar{\gamma}}}\left(\frac{1}{k-1}\sum_{j=1}^{k-1}\ln^2 V(X_j) + \ln^2 V(X_k) + \ln^2 V(X_{k-1})\right) V^\alpha(X_k) \xrightarrow{\mathbb{P}} 0 \tag{18}$$

for some $\tau > 0$. We prove that such a convergence occurs in L^1. To that goal, observe that the drift inequality $P_\theta V \le V + b$ implies that $\mathbb{E}[V(X_n)] \le \mathbb{E}[V(X_0)] + nb$, which in turn yields, by the Jensen inequality, $\sup_j (\ln^p j)^{-1}\mathbb{E}[\ln^p V(X_j)] < \infty$ for any $p \ge 2$. Then, by the Hölder inequality,

$$\sup_k \bigl(\ln^2 k k^\alpha\bigr)^{-1}\mathbb{E}\left[\left(\frac{1}{k-1}\sum_{j=1}^{k-1}\ln^2 V(X_j) + \ln^2 V(X_k) + \ln^2 V(X_{k-1})\right)V^\alpha(X_k)\right] < \infty.$$

Since $\alpha \in (0, 1/2)$ and τ can be chosen arbitrarily small, (18) is established and thus yields the condition A3(a).

We now consider A3(b). By (17), it is sufficient to prove that for some $\tau > 0$ and any $t > 0$,

$$n^{-1/2\alpha} \sum_{k=0}^{n-1} L_{\theta_k}^{2/\alpha} P_{\theta_k} V(X_k) \mathbb{1}\Big\{\sup_{\ell \geq 1} \ell^{-\tau} |\theta_\ell| \leq t\Big\} \xrightarrow{\mathbb{P}} 0. \tag{19}$$

By [18], Lemma 2.5, there exist a constant C (depending upon τ and t) such that

$$\mathbb{E}\Big[V(X_n) \mathbb{1}\Big\{\sup_{\ell \leq n-1} \ell^{-\tau} |\theta_\ell| \leq t\Big\}\Big] \leq C\big(\mathbb{E}\big[V(X_0)\big] + n^{\tau\gamma}\big),$$

where γ is defined in (16). Since $1/(2\alpha) > 1$, Equations (16) and (17) imply (19). This concludes the proof of A3(b).

Let us consider A4. The proof of this condition is a consequence of the convergence of $\{\theta_n, n \geq 0\}$ and the regularity in θ of F_θ. Under M1 and the condition $\mathbb{E}[V(X_0)] < \infty$, $n^{-1} \sum_{k=1}^{n} f(X_k) \xrightarrow{\text{a.s.}} \pi(f)$ for any $f \in \mathcal{L}_{V^a}$ and $a \in (0, 1)$ (see [18], Theorem 2.10). Since under M1 $\liminf_{|x| \to \infty} \ln V(x)/|x| > 0$, this implies that the strong Law of Large Numbers holds for functions f with quadratic growth at infinity. Therefore, $\{\theta_n, n \geq 0\}$ converges w.p.1 to $\theta_\star = (\mu_\star, \Gamma_\star)$ given by

$$\mu_\star \stackrel{\text{def}}{=} \int x\pi(x)\,\mathrm{d}x, \qquad \Gamma_\star \stackrel{\text{def}}{=} \int (x - \mu_\star)(x - \mu_\star)' \pi(x)\,\mathrm{d}x + \kappa \mathrm{I}.$$

Set

$$\sigma^2(f) \stackrel{\text{def}}{=} \int F_{\theta_\star}(x)\,\mathrm{d}x = \int \big(P_{\theta_\star}(\Lambda_{\theta_\star} f)^2(x) - [P_{\theta_\star} \Lambda_{\theta_\star} f]^2(x)\big)\,\mathrm{d}x. \tag{20}$$

The proof of A4 is given in the supplementary material [20]. Combining the results above yields the following theorem.

Theorem 3.1. *Assume* M1 *and* $\mathbb{E}[V(X_0)] < +\infty$. *Then, for any* $\alpha \in (0, 1/2)$ *and any* $f \in \mathcal{L}_{V^\alpha}$

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \big\{f(X_k) - \pi(f)\big\} \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, \sigma_f^2\big),$$

where $\sigma^2(f)$ *is given by* (20).

3.2. Interacting tempering algorithm

We consider the simplified version of the equi-energy sampler [24] introduced in [4]. This version is referred to as the Interacting-tempering (IT) sampler. Recently, convergence of the marginals and strong law of large numbers results have been established under general conditions (see [18]). In this section, we derive a CLT under similar assumptions.

Let $\{\pi^{\beta_k}, k \in \{1, \ldots, K\}\}$ be a sequence of tempered densities on X, where $0 < \beta_1 < \cdots < \beta_K = 1$. At the first level, a process $(Y_k)_{k\in\mathbb{N}}$ with stationary distribution proportional to π^{β_1} is run. At the second level, a process $(X_k)_{k\in\mathbb{N}}$ with stationary distribution proportional to π^{β_2} is constructed: at each iteration the next value is obtained from a Markov kernel depending on the occupation measure of the chain $(Y_k)_{k\in\mathbb{N}}$ up to the current time-step. This 2-stages mechanism is then repeated to design a process targeting π^{β_k} by using the occupation measure of the process targeting $\pi^{\beta_{k-1}}$.

For ease of exposition, it is assumed that $(\mathsf{X}, \mathcal{X})$ is a Polish space equipped with its Borel σ-field, and the densities are w.r.t. some σ-finite measure on $(\mathsf{X}, \mathcal{X})$. We address the case $K = 2$ and discuss below possible extensions to the case $K > 2$.

We start with a description of the IT (case $K = 2$). Denote by Θ the set of the probability measures on $(\mathsf{X}, \mathcal{X})$ equipped with the Borel sigma-field $\mathcal{T}$ associated to the topology of weak convergence. Let P be a transition kernel on $(\mathsf{X}, \mathcal{X})$ with unique invariant distribution π (typically, P is chosen to be a Metropolis–Hastings kernel). Denote by $\varepsilon \in (0, 1)$ the probability of interaction. Let $(Y_k)_{k\in\mathbb{N}}$ be a discrete-time (possibly non-stationary) process and denote by θ_n the empirical probability measure:

$$\theta_n \stackrel{\text{def}}{=} \frac{1}{n}\sum_{k=1}^{n} \delta_{Y_k}. \tag{21}$$

Choose $X_0 \sim \nu$. At the nth iteration of the algorithm, two actions may be taken:

1. with probability $(1 - \varepsilon)$, the state X_{n+1} is sampled from the Markov kernel $P(X_n, \cdot)$,
2. with probability ε, a tentative state Z_{n+1} is drawn uniformly from the past of the auxiliary process $\{Y_k, k \leq n\}$. This move is accepted with probability $r(X_n, Z_{n+1})$, where the acceptance ratio r is given by

$$r(x, z) \stackrel{\text{def}}{=} 1 \wedge \frac{\pi(z)\pi^{1-\beta}(x)}{\pi^{1-\beta}(z)\pi(x)} = 1 \wedge \frac{\pi^{\beta}(z)}{\pi^{\beta}(x)}. \tag{22}$$

Define the family of Markov transition kernels $\{P_\theta, \theta \in \Theta\}$ by

$$\begin{aligned} P_\theta(x, A) &\stackrel{\text{def}}{=} (1 - \varepsilon)P(x, A) \\ &\quad + \varepsilon\left(\int_A r(x, y)\theta(\mathrm{d}y) + \mathbb{1}_A(x)\int \{1 - r(x, y)\}\theta(\mathrm{d}y)\right). \end{aligned} \tag{23}$$

Then, the above algorithmic description implies that the bivariate process $\{(X_n, \theta_n)\}_{n\in\mathbb{N}}$ is such that for any bounded function h on X^{n+1}

$$\mathbb{E}\big[h(X_{0:n})|\theta_{0:n}\big] = \int \nu(\mathrm{d}x_0)P_{\theta_0}(x_0, \mathrm{d}x_1)\cdots P_{\theta_{n-1}}(x_{n-1}, \mathrm{d}x_n)h(x_{0:n}).$$

We apply the results of Section 2 in order to prove that the IT process $(X_k)_{k\in\mathbb{N}}$ satisfies a CLT. To that goal, it is assumed that the target density π and the transition kernel P satisfy the following conditions:

I1 π is a continuous positive density on X and $|\pi|_\infty < +\infty$.

I2 (a) P is a phi-irreducible aperiodic Feller transition kernel on $(\mathsf{X}, \mathcal{X})$ such that $\pi P = \pi$.

(b) There exist $\tau \in (0,1)$, $\lambda \in (0,1)$ and $b < +\infty$ such that

$$PV(x) \le \lambda V(x) + b \qquad \text{with } V(x) \stackrel{\text{def}}{=} \bigl(\pi(x)/|\pi|_\infty\bigr)^{-\tau}. \tag{24}$$

(c) For any $p \in (0, |\pi|_\infty)$, the sets $\{\pi \ge p\}$ are 1-small (w.r.t. the transition kernel P).

(d) For any $\gamma \in (0, 1/2)$ and any equicontinuous set of functions $\mathcal{F} \subseteq \mathcal{L}_{V^\gamma}$, the set of functions $\{Ph : h \in \mathcal{F}, |h|_{V^\gamma} \le 1\}$ is equicontinuous.

From the expression of the acceptance ratio r (see Equation (22)) and the assumption I2(a), it holds

$$\pi P_{\theta_\star} = \pi, \tag{25}$$

where $\theta_\star \propto \pi^{1-\beta}$. Therefore, when θ_n converges to $\theta_\star$, it is expected that $(X_k)_{k\in\mathbb{N}}$ behaves asymptotically as π; see [18].

Drift conditions for the symmetric random walk Metropolis (SRWM) algorithm are discussed in [30], [23] and [32]. Under conditions which imply that the target density π is super-exponential in the tails and have regular contours, Jarner and Hansen [23] and Saksman and Vihola [32] show that any functions proportional to π^{-s} with $s \in (0,1)$ satisfies a Foster–Lyapunov drift inequality [23], Theorems 4.1 and 4.3. Under this condition, I2(b) is satisfied with any τ in the interval $(0,1)$. Assumptions I2(c) and I2(d) hold for the SRWM kernel under weak conditions on the symmetric proposal distribution: the minorization condition is verified whenever the proposal is positive and continuous (see, e.g., [26], Lemma 1.2) and the following lemma gives sufficient conditions for I2(d). The proof is in Section 4.2.1.

Lemma 3.2. *Assume* I1. *Let P be a Metropolis kernel with invariant distribution π and a symmetric proposal distribution $q : \mathsf{X} \times \mathsf{X} \to \mathbb{R}^+$ such that $\sup_{(x,y)\in\mathsf{X}^2} q(x,y) < +\infty$ and the function $x \mapsto q(x,\cdot)$ is continuous from $(\mathsf{X}, |\cdot|)$ to the set of probability densities equipped with the total variation norm. Then P satisfies* I2(d) *with any function $V \propto \pi^{-\tau}$, $\tau \in [0,1)$, such that $\pi(V) < +\infty$.*

For a measurable function $f : \mathsf{X} \to \mathbb{R}$ such that $\theta_\star(|f|) < +\infty$, define the following sequence of random processes on $[0,1]$:

$$t \mapsto S_n(f;t) = n^{-1/2} \sum_{j=1}^{\lfloor nt \rfloor} \bigl\{ f(Y_j) - \theta_\star(f) \bigr\}. \tag{26}$$

It is assumed that the auxiliary process $\{Y_n, n \ge 0\}$ converges to the probability distribution $\theta_\star$ in the following sense:

I3 (a) $\theta_\star(V) < +\infty$ and $\sup_n \mathbb{E}[V(Y_n)] < +\infty$.

(b) There exists a space $\mathcal{N}$ of real-valued measurable functions defined on X such that $V \in \mathcal{N}$ and for any function $f \in \mathcal{N}$, $\theta_n(f) \xrightarrow{\text{a.s.}} \theta_\star(f)$.

(c) For any function $f \in \mathcal{N}$, the sequence of processes $(S_n(f,t), n \geq 1, t \in [0,1])$ converges in distribution to $(\tilde{\gamma}(f)B(t), t \in [0,1])$, where $\tilde{\gamma}(f)$ is a non-negative constant and $(B(t): t \in [0,1])$ is a standard Brownian motion.

(d) For any $\alpha \in (0, 1/2)$, there exist constants ϱ_0 and ϱ_1 such that, for any integers $n, k \geq 1$, for any measurable function $h : \mathsf{X}^k \to \mathbb{R}$ satisfying $|h(y_1, \dots, y_k)| \leq \sum_{j=1}^k V^\alpha(y_j)$,

$$\mathbb{E}\left(\int \cdots \int \prod_{j=1}^{k} \left[\theta_n(\mathrm{d}y_j) - \theta_\star(\mathrm{d}y_j)\right] h(y_1, \dots, y_k)\right)^2 \leq A_k n^{-k},$$

with $\limsup_k \ln A_k/(k \ln k) < \infty$.

I3 is satisfied when $(Y_k)_{k\in\mathbb{N}}$ is i.i.d. with distribution $\theta_\star$ such that $\theta_\star(V) < +\infty$. In that case, I3(b) to I3(c) hold for any measurable function f such that $\theta_\star(|f|^2) < +\infty$. I3(d) is satisfied using [33], Lemma A, pages 190.

I3 is also satisfied when $(Y_k)_{k\in\mathbb{N}}$ is a geometrically ergodic Markov chain with transition kernel Q. In that case, I3(a) to I3(c) are satisfied for any measurable function f such that $\theta_\star(|f[(I-Q)^{-1}f]|) < +\infty$ (see, e.g., [27], Chapter 17). Condition I3(d) for a (non-stationary) geometrically ergodic Markov chain is established in the supplementary paper [19].

The following proposition shows that under I1 and I2, condition A2 holds with the drift function V given by A2(b). It also provides a control of the ergodicity constants C_θ, ρ_θ in Lemma 2.1. The proof is a direct consequence of [18], Proposition 3.1, Corollary 3.2, Lemmas 2.1 and A.2, and is omitted.

Proposition 3.3. *Assume* I1 *and* I2(a)–(c). *For any $\theta \in \Theta$, P_θ is phi-irreducible, aperiodic. In addition, there exist $\tilde{\lambda} \in (0,1)$ and $\tilde{b} < +\infty$ such that, for any $\theta \in \Theta$,*

$$P_\theta V(x) \leq \tilde{\lambda} V(x) + \tilde{b}\theta(V) \qquad \text{for all } x \in \mathsf{X}. \tag{27}$$

The property P[α] *holds for any $\alpha \in (0, 1/2)$, and there exists C such that for any $\theta \in \Theta$, $L_\theta \leq C\theta(V)$.*

Assume in addition I3(a) *and* $\mathbb{E}[V(X_0)] < +\infty$. *Then,* $\sup_{n\geq 0} \mathbb{E}[V(X_n)] < +\infty$.

The next step is to check assumptions A3 and A4.

Proposition 3.4. *Assume* I1, I2, I3(a)–(b) *and* $\mathbb{E}[V(X_0)] < +\infty$. *For any $\alpha \in (0, 1/2)$, set $\mathcal{M}_{V^\alpha}$ be the set of continuous functions belonging to $\mathcal{L}_{V^\alpha} \cap \mathcal{N}$. Then, for any $\alpha \in (0, 1/2)$, the conditions* A3 *and* A4 *hold with*

$$\sigma^2(f) \stackrel{def}{=} \int \pi(\mathrm{d}x) F_{\theta_\star}(x), \tag{28}$$

where F_θ is given by (10).

The proof is postponed to Section 4.2.2. We can now apply Theorem 2.3 and prove a CLT for the 2-levels IT.

Theorem 3.5. *Assume* I1, I2, I3 *and* $\mathbb{E}[V(X_0)] < +\infty$. *Then, for any* $\alpha \in (0, 1/2)$ *and any continuous function* $f \in \mathcal{L}_{V^\alpha} \cap \mathcal{N}$ *such that the function* G_f *given by*

$$G_f(z) \stackrel{def}{=} \varepsilon \int \pi(\mathrm{d}x) r(x, z) \big(\Lambda_{\theta_\star} f(z) - \Lambda_{\theta_\star} f(x)\big),$$

is in $\mathcal{N}$:

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \big(f(X_k) - \pi(f)\big) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, \sigma^2(f) + 2\tilde{\gamma}^2(G_f)\big),$$

where $\sigma^2(f)$ *and* $\tilde{\gamma}^2(G_f)$ *are given by* (28) *and* I3(c).

The proof is postponed to Appendix 4.2.3.

It may be possible to repeat the above argument to show a CLT for the K-level IT when $K > 2$ (see [18] for a similar approach in the proof of the ergodicity and the LLN for IT). Nevertheless, the main difficulty is to iterate the control of the L^2-moment for the V-statistics (see I3(d)) when $(Y_k)_{k\in\mathbb{N}}$ is not a Markov chain or, more generally, a process satisfying some mixing conditions. A similar difficulty has been reported in [4].

Theorem 3.5 shows that the asymptotic variance of sample path averages of the process $\{X_n, n \geq 0\}$ for the functional f is the sum of two terms. The first term $\sigma^2(f)$ is the asymptotic variance of sample path averages of a Markov chain with transition kernel $P_{\theta_\star}$ and functional f (see, e.g., [27], Chapter 17). The second term $\tilde{\gamma}^2(G_f)$ is the asymptotic variance of sample path averages of the auxiliary process $\{Y_n, n \geq 0\}$ for the functional G_f. The expression of this asymptotic variance can help in the choice of the probability of interaction ε. For example, given the kernel P, a question is: is the asymptotic variance reduced when replacing the classical MCMC chain with kernel P by the interacting process satisfying (2) with P_θ given by (23)? to answer this question, first note that the derivative with respect to ε of $\sigma^2(f) + 2\tilde{\gamma}^2(G_f)$ at $\varepsilon = 0$ is equal to the derivative of $\sigma^2(f)$ at $\varepsilon = 0$. In addition, this derivative is of the sign of

$$-\int \pi(\mathrm{d}x) \bar{h}(x) \Lambda_{\theta_\star} (P - K_{\theta_\star}) \Lambda_{\theta_\star} \bar{h}(x) = -\big\langle \Lambda_{\theta_\star} \bar{h}, (P - K_{\theta_\star}) \Lambda_{\theta_\star} \bar{h} \big\rangle_{L^2(\pi)},$$

where K_θ is defined by $P_\theta = (1 - \varepsilon)P + \varepsilon K_\theta$ and $\bar{h} = h - \pi(h)$. Therefore, if $P - K_{\theta_\star}$ is a positive operator on $L_0^2(\pi) \stackrel{\text{def}}{=} \{h : \pi(h) = 0, \pi(h^2) < \infty\}$, the 2-level IT algorithm with ε small enough will improve on the MCMC sampler P. A sufficient condition for $P - K_{\theta_\star}$ to be a positive operator is $P \leq K_{\theta_\star}$ in the Peskun ordering of transition kernels (see, e.g., [34], Lemma 3). Note that under this Peskun order assumption on P and $K_{\theta_\star}$, the function $\varepsilon \mapsto \sigma^2(f)$ is non-increasing on $[0, 1]$ for any function $f \in L_0^2(\pi)$ (see the proof of [34], Theorem 4). Figure 1 below shows that this non-increasing property is balanced by the behavior of $\varepsilon \mapsto 2\tilde{\gamma}^2(G_f)$. Figure 1 displays an estimation of the variance of $\sqrt{N} \sum_{k=1}^{N} \{f(X_k) - \pi(f)\}$, obtained from 300 independent run of the process $\{X_n, n \geq 0\}$. In this numerical application, $N = 400k$; π is a mixture of five $\mathbb{R}^5$-valued Gaussian distribution with means drawn in the range $[-3; 3]^5$ and covariance matrix identity; $f : \mathbb{R}^5 \to \mathbb{R}$ is defined by $x = (x_1, \ldots, x_5) \mapsto x_5$; P is a SRWM algorithm with proposal kernel $q(x, \cdot) \sim \mathcal{N}_5(x, \mathrm{I})$; and $\{Y_k, k \geq 0\}$ is a SRWM with proposal kernel $q(x, \cdot) \sim \mathcal{N}_5(x, \mathrm{I})$ and

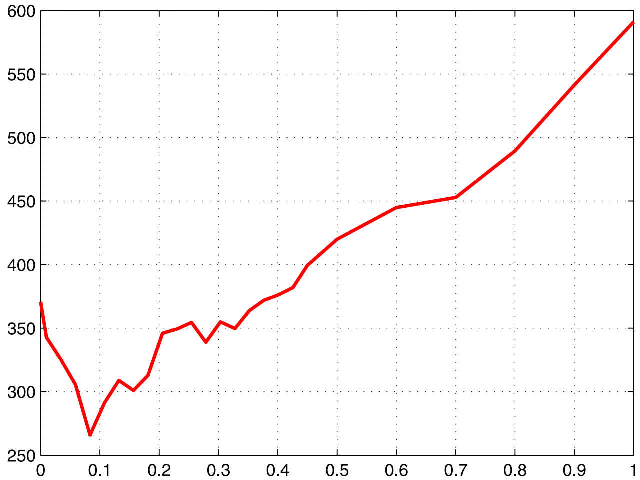

Figure 1. Estimation of the variance of $\sqrt{N}\sum_{k=1}^{N}\{f(X_k)-\pi(f)\}$, as a function of the probability of interaction ε. The plots have been obtained with 20 linearly spaced values of ε in the range $[0, 0.45]$; and 6 linearly spaced values in the range $[0.5, 1]$.

invariant measure $\pi_{\theta_\star}^{0.2}$. Figure 1 shows that the variance is minimal for some ε in $[0.05; 0.15]$ and corroborates previous empirical results on the choice of ε (see, e.g., [24]).

4. Proofs

Note that under A2, for any $\alpha \in (0, 1]$, any $f \in \mathcal{L}_{V^\alpha}$ and any $\theta \in \Theta$,

$$|\Lambda_\theta f|_{V^\alpha} \le |f|_{V^\alpha} L_\theta^2, \tag{29}$$

where L_θ is defined by (9).

4.1. Proofs of the results in Section 2

4.1.1. *Proof of Theorem* 2.2

Let $f \in \mathcal{M}_{V^\alpha}$. Equation (4) yields $S_n^{(1)}(f) = \Xi_n(f) + R_n^{(1)}(f) + R_n^{(2)}(f)$ with

$$\Xi_n(f) \stackrel{\text{def}}{=} \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\bigl\{\Lambda_{\theta_{k-1}} f(X_k) - P_{\theta_{k-1}}\Lambda_{\theta_{k-1}} f(X_{k-1})\bigr\},$$

$$R_n^{(1)}(f) \stackrel{\text{def}}{=} n^{-1/2}\sum_{k=1}^{n}\bigl\{P_{\theta_k}\Lambda_{\theta_k} f(X_k) - P_{\theta_{k-1}}\Lambda_{\theta_{k-1}} f(X_k)\bigr\},$$

$$R_n^{(2)}(f) \stackrel{\text{def}}{=} n^{-1/2} P_{\theta_0}\Lambda_{\theta_0} f(X_0) - n^{-1/2} P_{\theta_n}\Lambda_{\theta_n} f(X_n).$$

We first show that the two remainders terms $R_n^{(1)}(f)$ and $R_n^{(2)}(f)$ converge to zero in probability. We have

$$\left|P_\theta \Lambda_\theta f(x) - P_{\theta'} \Lambda_{\theta'} f(x)\right| \le \left|P_\theta \Lambda_\theta f(x) - P_{\theta'} \Lambda_{\theta'} f(x)\right|_{V^\alpha} V^\alpha(x).$$

Assumption A3 implies that $R_n^{(1)}(f)$ converges to zero in probability. The drift inequality A2 combined with the Jensen's inequality imply $P_\theta V^\alpha \le \lambda_\theta^\alpha V^\alpha + b_\theta^\alpha$. By (29) and this inequality,

$$\left|P_\theta \Lambda_\theta f(x)\right| \le |f|_{V^\alpha} L_\theta^2 P_\theta V^\alpha(x) \le |f|_{V^\alpha} L_\theta^2 \left(V^\alpha(x) + b_\theta^\alpha\right).$$

Then, $P_{\theta_0} \Lambda_{\theta_0} f(X_0)$ is finite w.p.1 and $n^{-1/2} P_{\theta_0} \Lambda_{\theta_0} f(X_0) \xrightarrow{\text{a.s.}} 0$. By A3(b) and (29), $n^{-1/2} P_{\theta_n} \Lambda_{\theta_n} f(X_n) \xrightarrow{\mathbb{P}} 0$. Hence, $R_n^{(2)}(f) \xrightarrow{\mathbb{P}} 0$.

We now consider $\Xi_n(f)$. Set $D_k(f) \stackrel{\text{def}}{=} \Lambda_{\theta_{k-1}} f(X_k) - P_{\theta_{k-1}} \Lambda_{\theta_{k-1}} f(X_{k-1})$. Observe that under A1, $D_k(f)$ is a martingale-increment w.r.t. the filtration $\{\mathcal{F}_k, k \ge 0\}$. The limiting distribution for $\Xi_n(f)$ follows from martingale CLT (see, e.g., [22], Corollary 3.1). We check the conditional Lindeberg condition. Let $\varepsilon > 0$. Under A2, we have by (29)

$$D_k(f) \le |f|_{V^\alpha} \left| L_{\theta_{k-1}}^2 \left\{V^\alpha(X_k) + P_{\theta_{k-1}} V^\alpha(X_{k-1})\right\}\right|.$$

Set $\tau \stackrel{\text{def}}{=} 1/\alpha - 2 > 0$.

$$\begin{aligned}
&\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\left[D_k^2(f)\mathbb{1}_{|D_k(f)|\ge\varepsilon\sqrt{n}} | \mathcal{F}_{k-1}\right] \\
&\le \left(\frac{1}{\varepsilon\sqrt{n}}\right)^{\tau} \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\left[D_k^{2+\tau}(f) | \mathcal{F}_{k-1}\right] \\
&\le |f|_{V^\alpha}^{2+\tau} \left(\frac{1}{\varepsilon\sqrt{n}}\right)^{\tau} \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\left[L_{\theta_{k-1}}^{2(2+\tau)} \left\{V^\alpha(X_k) + P_{\theta_{k-1}} V^\alpha(X_{k-1})\right\}^{2+\tau} | \mathcal{F}_{k-1}\right] \\
&\le 2^{2+\tau} |f|_{V^\alpha}^{2+\tau} \left(\frac{1}{\varepsilon\sqrt{n}}\right)^{\tau} \frac{1}{n}\sum_{k=0}^{n-1} L_{\theta_k}^{2(2+\tau)} P_{\theta_k} V(X_k).
\end{aligned}$$

Under A3(b), the RHS converges to zero in probability thus concluding the proof of the conditional Lindeberg condition. For the limiting variance condition, observe that

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\left[D_k^2(f) | \mathcal{F}_{k-1}\right] = \frac{1}{n}\sum_{k=0}^{n-1} F_{\theta_k}(X_k),$$

where F_θ is given by (10) and, under A4, $n^{-1}\sum_{k=1}^{n} \mathbb{E}[D_k^2 | \mathcal{F}_{k-1}] \xrightarrow{\mathbb{P}} \sigma^2(f)$. This concludes the proof.

4.1.2. *Proof of Theorem* 2.3

We start by establishing a joint CLT for $(S_n^{(1)}(f), S_n^{(2)}(f))$, where $S_n^{(1)}(f)$ and $S_n^{(2)}(f)$ are defined in (5) and (6), respectively. Similar to the proof of Theorem 2.2, we write $S_n^{(1)}(f) = \Xi_n(f) + R_n^{(1)}(f) + R_n^{(2)}(f)$ and prove that $R_n^{(1)}(f) + R_n^{(2)}(f) \xrightarrow{\mathbb{P}} 0$. We thus consider the convergence of $\Xi_n(f) + S_n^{(2)}(f)$. Set $\mathcal{F}_n^\theta \stackrel{\text{def}}{=} \sigma(\theta_k, k \le n)$. Under A5,

$$\mathbb{E}\big[\mathrm{e}^{\mathrm{i}(u_1 \Xi_n(f) + u_2 S_n^{(2)}(f))}\big] = \mathbb{E}\big[\mathbb{E}\big[\mathrm{e}^{\mathrm{i}u_1 \Xi_n(f)} | \mathcal{F}_n^\theta\big] \mathrm{e}^{\mathrm{i}u_2 S_n^{(2)}(f)}\big].$$

Applying the conditional CLT [15], Theorem A.3, with the filtration $\mathcal{F}_{n,k} \stackrel{\text{def}}{=} \sigma(Y_1, \ldots, Y_n, X_1, \ldots, X_k)$, yields:

$$\lim_{n\to\infty} \mathbb{E}\big[\mathrm{e}^{\mathrm{i}u_1 \Xi_n(f)} | \mathcal{F}_n^\theta\big] \xrightarrow{\mathbb{P}} \mathrm{e}^{-u_1^2 \sigma^2(f)/2}; \tag{30}$$

observe that under A5, the conditions (31) and (32) in [15] can be proved following the same lines as in the proof of Theorem 2.2; details are omitted. Therefore,

$$\begin{aligned}\mathbb{E}\big[\mathrm{e}^{\mathrm{i}(u_1 \Xi_n(f) + u_2 S_n^{(2)}(f))}\big] &= \mathbb{E}\big[\big(\mathbb{E}\big[\mathrm{e}^{\mathrm{i}u_1 \Xi_n(f)} | \mathcal{F}_n^\theta\big] - \mathrm{e}^{-u_1^2 \sigma^2(f)/2}\big) \mathrm{e}^{\mathrm{i}u_2 S_n^{(2)}(f)}\big] \\ &\quad + \mathrm{e}^{-u_1 \sigma^2(f)/2} \mathbb{E}\big[\mathrm{e}^{\mathrm{i}u_2 S_n^{(2)}(f)}\big].\end{aligned}$$

By (30), the first term in the RHS of the previous equation converges to zero. Under A6, $\lim_{n\to\infty} \mathbb{E}[\mathrm{e}^{\mathrm{i}u_2 S_n^{(2)}(f)}] = \mathrm{e}^{-u_2^2 \gamma^2(f)/2}$ and this concludes the proof.

4.2. Proofs of Section 3.2

Note that by (25), $\pi_{\theta_\star} = \pi$.

4.2.1. *Proof of Lemma* 3.2

Let $\gamma \in (0, 1/2)$ and $\mathcal{F}$ be an equicontinuous set of functions in $\mathcal{L}_{V^\gamma}$. Let $h \in \mathcal{F}$, $|h|_{V^\gamma} \le 1$. By construction, the transition kernel of a symmetric random walk Metropolis with proposal transition density $q(x, \cdot)$ and target density π may be expressed as

$$Ph(x) = \int r(x, y) h(y) q(x, y)\, \mathrm{d}y + h(x) \int \big\{1 - r(x, y)\big\} q(x, y)\, \mathrm{d}y,$$

where $r(x, y) \stackrel{\text{def}}{=} 1 \wedge (\pi(y)/\pi(x))$ is the acceptance ratio. Therefore, the difference $Ph(x) - Ph(x')$ may be bounded by

$$\begin{aligned}\big|Ph(x) - Ph\big(x'\big)\big| &\le 2\big|h(x) - h\big(x'\big)\big| \\ &\quad + \int \big|h(y) - h\big(x'\big)\big|\big|r(x, y) - r\big(x', y\big)\big| q(x, y)\, \mathrm{d}y \\ &\quad + \left|\int \big(h(y) - h\big(x'\big)\big) r\big(x', y\big)\big(q(x, y) - q\big(x', y\big)\big)\, \mathrm{d}y\right|.\end{aligned}$$

Since $|r(x,y)-r(x',y)| \le \pi(y)|\pi^{-1}(x)-\pi^{-1}(x')|$,

$$\begin{aligned}
&\int \left|h(y)-h(x')\right|\left|r(x,y)-r(x',y)\right| q(x,y)\,\mathrm{d}y \\
&\quad\le \left|\pi^{-1}(x)-\pi^{-1}(x')\right| \int \left|h(y)-h(x')\right| \pi(y) q(x,y)\,\mathrm{d}y \\
&\quad\le \Bigl(\sup_{(x,y)\in \mathsf{X}^2} q(x,y)\Bigr) \left|\pi^{-1}(x)-\pi^{-1}(x')\right| \bigl(\pi(V^\gamma)+V^\gamma(x')\bigr).
\end{aligned}$$

In addition,

$$\begin{aligned}
&\left|\int \bigl(h(y)-h(x')\bigr) r(x',y)\bigl(q(x,y)-q(x',y)\bigr)\,\mathrm{d}y\right| \\
&\quad= \left|\int_{\{y:\pi(y)\le\pi(x')\}} \bigl(h(y)-h(x')\bigr)\frac{\pi(y)}{\pi(x')}\bigl(q(x,y)-q(x',y)\bigr)\,\mathrm{d}y\right| \\
&\qquad+ \left|\int_{\{y:\pi(y)>\pi(x')\}} \bigl(h(y)-h(x')\bigr)\bigl(q(x,y)-q(x',y)\bigr)\,\mathrm{d}y\right| \\
&\quad\le 4\pi^{-1}(x')\left\|q(x,\cdot)-q(x',\cdot)\right\|_{\mathrm{TV}} \sup_{y\in\mathsf{X}}\left|h(y)\pi(y)\right|.
\end{aligned}$$

Since $V \propto \pi^{-\tau}$ and $\tau \in (0,1)$, $\sup_{\mathsf{X}} |h|\pi \le 1$ under I1. Therefore, there exists a constant C such that for any $h \in \{h \in \mathcal{F}, |h|_{V^\gamma} \le 1\}$ and any $x, x' \in \mathsf{X}$,

$$\begin{aligned}
\left|Ph(x)-Ph(x')\right| &\le 2\left|h(x)-h(x')\right| \\
&\quad+ C\bigl(\left|\pi^{-1}(x)-\pi^{-1}(x')\right| + \left\|q(x,\cdot)-q(x',\cdot)\right\|_{\mathrm{TV}}\bigr)\bigl(V^\gamma(x')+\pi^{-1}(x')\bigr),
\end{aligned}$$

thus concluding the proof.

4.2.2. *Proof of Proposition* 3.4

The proof is prefaced by several lemmas. The proof of Lemma 4.1 is omitted for brevity and can be found in the supplementary material [20]. The proof of Lemma 4.2 is adapted from [18], Lemma 5.1, and is omitted.

Lemma 4.1. *Let* $\alpha \in (0,1)$. *Assume* I1, I2(a)–(c), I3(a)–(b), *and* $\mathbb{E}[V(X_0)] < +\infty$. *Then for any* $\gamma, \gamma' \in (0,1)$ *and any* $\delta > \gamma$,

$$n^{-\delta}\sum_{k=1}^{n} D_{V^\gamma}(\theta_k,\theta_{k-1}) V^{\gamma'}(X_k) \xrightarrow{\mathbb{P}} 0.$$

Lemma 4.2. *For any $\theta \in \Theta$, any measurable function $f : \mathsf{X} \to \mathbb{R}$ in $\mathcal{L}_{V^\alpha}$ and any $x, x' \in \mathsf{X}$ such that $\pi(x) \le \pi(x')$*

$$\left|P_\theta f(x) - P_\theta f\left(x'\right)\right| \le \left|Pf(x) - Pf\left(x'\right)\right| + \left|f(x) - f\left(x'\right)\right| \\ + \sup_{\mathsf{X}} \pi |f|_{V^\alpha} \left|\pi^{-\beta}(x) - \pi^{-\beta}\left(x'\right)\right| \left(V^\alpha\left(x'\right) + \theta\left(V^\alpha\right)\right).$$

Proof of Proposition 3.4. Let $\alpha \in (0, 1/2)$. By Proposition 3.3, A2 and P[α] hold. By I3(b),

$$\limsup_n L_{\theta_n} < +\infty, \qquad \mathbb{P}\text{-a.s.}, \tag{31}$$

where L_θ is given by (9) with C_θ, ρ_θ defined by P[α].

We first check A3(a). Let $f \in \mathcal{N} \cap \mathcal{L}_{V^\alpha}$. By Lemma A.1,

$$|P_{\theta_k} \Lambda_{\theta_k} f - P_{\theta_{k-1}} \Lambda_{\theta_{k-1}} f|_{V^\alpha} \le 5(L_{\theta_k} \vee L_{\theta_{k-1}})^6 \pi_{\theta_k}\left(V^\alpha\right) D_{V^\alpha}(\theta_k, \theta_{k-1}) |f|_{V^\alpha}.$$

By Lemma 2.1, Proposition 3.3 and Assumptions I1, I2 and I3(b),

$$\limsup_{n\to\infty} \pi_{\theta_n}(V) \le \tilde{b}(1 - \tilde{\lambda})^{-1} \limsup_{n\to\infty} \theta_n(V) < \infty, \qquad \mathbb{P}\text{-a.s.} \tag{32}$$

Therefore, by (31) and (32), it suffices to prove that

$$n^{-1/2} \sum_{k=1}^{n} D_{V^\alpha}(\theta_k, \theta_{k-1}) V^\alpha(X_k) \xrightarrow{\mathbb{P}} 0,$$

which follows from Lemma 4.1. We now check A3(b). By Proposition 3.3, it holds

$$n^{-1/(2\alpha)} \sum_{k=1}^{n} L_{\theta_k}^{2/\alpha} P_{\theta_k} V(X_k) \le n^{-1/(2\alpha)} \sum_{k=1}^{n} L_{\theta_k}^{2/\alpha} \left[V(X_k) + \tilde{b}\theta_k(V)\right].$$

Under the stated assumptions, $\limsup_n [\theta_n(V) + L_{\theta_n}] < +\infty$ w.p.1 and by Proposition 3.3, $\sup_k \mathbb{E}[V(X_k)] < +\infty$. Since $2\alpha < 1$, this concludes the proof.

The proof of A4 is in two steps: it is first proved that

$$\frac{1}{n} \sum_{k=0}^{n-1} F_{\theta_k}(X_k) - \frac{1}{n} \sum_{k=0}^{n-1} \int \pi_{\theta_k}(\mathrm{d}x) F_{\theta_k}(x) \xrightarrow{\mathbb{P}} 0, \tag{33}$$

and then it is established that

$$\int \pi_{\theta_k}(\mathrm{d}x) F_{\theta_k}(x) \xrightarrow{\text{a.s.}} \int \pi_{\theta_\star}(\mathrm{d}x) F_{\theta_\star}(x). \tag{34}$$

Theorem B.1 in Appendix B applied with $\gamma = 2\alpha$ implies (33). The main tools for checking the assumptions of Theorem B.1 are (31), (32), Lemma 4.1 and Lemmas A.1 and A.3. A detailed proof can be found in the supplementary paper, see [20].

The second step is to prove (34). To that goal, we have to strengthen the conditions on f by assuming that f is continuous. For any $\theta \in \Theta$, $\int \pi_\theta(\mathrm{d}x)F_\theta(x) = \int \pi_\theta(\mathrm{d}x)H_\theta(x)$ with

$$H_\theta(x) \stackrel{\text{def}}{=} (\Lambda_\theta f)^2(x) - (P_\theta \Lambda_\theta f)^2(x). \tag{35}$$

We have to prove that there exists $\Omega_\star$ with $\mathbb{P}(\Omega_\star) = 1$ and for any $\omega \in \Omega_\star$,

(I) for any continuous bounded function h, $\lim_n \pi_{\theta_n(\omega)}(h) = \pi_{\theta_\star}(h)$,
(II) the set $\{H_{\theta_n(\omega)}, n \geq 0\}$ is equicontinuous,
(III) $\sup_n \pi_{\theta_n(\omega)}(|H_{\theta_n(\omega)}|^{1/(2\alpha)}) < +\infty$,
(IV) $\lim_n H_{\theta_n(\omega)}(x) = H_{\theta_\star}(x)$ for any $x \in \mathsf{X}$,
(V) $\pi_{\theta_\star}(|H_{\theta_\star}|) < +\infty$.

The proof is then concluded by application of Lemma A.3. Details of these steps are omitted for brevity and can be found in the supplementary paper, see [20]. $\square$

4.2.3. *Proof of Theorem* 3.5

We check the conditions of Theorem 2.3. A2 to A5 hold (see Propositions 3.3 and 3.4) and we now prove A6. We first check condition A6(a). For any function $f \in \mathcal{L}_{V^\alpha} \cap \mathcal{N}$, define

$$G_f(z) \stackrel{\text{def}}{=} \varepsilon \int\int \bigl(\delta_z(\mathrm{d}z') - \theta_\star(\mathrm{d}z')\bigr)\pi_{\theta_\star}(\mathrm{d}x) r\bigl(x, z'\bigr)\bigl(\Lambda_{\theta_\star} f\bigl(z'\bigr) - \Lambda_{\theta_\star} f(x)\bigr). \tag{36}$$

Let $f \in \mathcal{L}_{V^\alpha} \cap \mathcal{N}$; note that $G_f \in \mathcal{L}_{V^\alpha}$. Recall that by Equation (23), for any θ such that $\theta(V^\alpha) < +\infty$,

$$P_\theta f(x) - P_{\theta_\star} f(x) = \varepsilon \int \bigl[\theta(\mathrm{d}y) - \theta_\star(\mathrm{d}y)\bigr] r(x, y)\bigl(f(y) - f(x)\bigr). \tag{37}$$

Then, using (36),

$$\begin{aligned}
&\pi_{\theta_\star}(P_{\theta_k} - P_{\theta_\star})\Lambda_{\theta_\star} f \\
&\quad = \varepsilon \int\int \pi_{\theta_\star}(\mathrm{d}x)\bigl[\theta_k(\mathrm{d}z) - \theta_\star(\mathrm{d}z)\bigr] r(x, z)\bigl[\Lambda_{\theta_\star} f(z) - \Lambda_{\theta_\star} f(x)\bigr] = \theta_k(G_f).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{k=1}^n \pi_{\theta_\star}(P_{\theta_k} - P_{\theta_\star})\Lambda_{\theta_\star} f \\
&\quad = \frac{1}{n} \sum_{k=1}^n \frac{n}{k} \frac{1}{\sqrt{n}} \sum_{j=1}^k G_f(Y_j) \\
&\quad = \int_0^1 t^{-1} S_n(G_f, t)\, \mathrm{d}t + \sum_{k=1}^{n-1} \int_{k/n}^{(k+1)/n} \left(\frac{n}{k} - \frac{1}{t}\right) S_n(G_f, t)\, \mathrm{d}t + \frac{1}{n} S_n(G_f, 1),
\end{aligned}$$

with $S_n(G_f, t) \stackrel{\text{def}}{=} n^{-1/2} \sum_{j=1}^{\lfloor nt \rfloor} G_f(Y_j)$. Note that

$$\mathbb{E}\left[\left|\sum_{k=1}^{n-1} \int_{k/n}^{(k+1)/n} \left(\frac{n}{k} - \frac{1}{t}\right) S_n(G_f, t)\, \mathrm{d}t\right|\right] \le \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{1}{k+1} \frac{1}{k} \sum_{j=1}^{k} \mathbb{E}\big[\big|G_f(Y_j)\big|\big].$$

Since $G_f \in \mathcal{L}_{V^\alpha}$, I3(a) implies that $\sup_{k \ge 0} \mathbb{E}[|G_f|(Y_k)] < \infty$. Therefore,

$$\sum_{k=1}^{n-1} \int_{k/n}^{(k+1)/n} \left(\frac{n}{k} - \frac{1}{t}\right) S_n(G_f, t)\, \mathrm{d}t + \frac{1}{n} S_n(G_f, 1) \xrightarrow{\mathbb{P}} 0.$$

Using I3(c), I3(d) and the Continuous mapping theorem ([35], Theorem 1.3.6), we obtain

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \pi_{\theta_\star}(P_{\theta_k} - P_{\theta_\star})\Lambda_{\theta_\star} f \xrightarrow{\mathcal{D}} \tilde{\gamma}^2(f) \int_0^1 t^{-1} B_t\, \mathrm{d}t.$$

Since $\int_0^1 t^{-1} B_t\, \mathrm{d}t = \int_0^1 \log(t)\, \mathrm{d}B_t$, $\int_0^1 t^{-1} B_t\, \mathrm{d}t$ is a Gaussian random variable with zero mean and variance $\int_0^1 \log^2(t)\, \mathrm{d}t = 2$.

We now check condition A6(b). Note that

$$n^{-1/2} \sum_{k=1}^{n} \pi_{\theta_k}(P_{\theta_k} - P_{\theta_\star})\Lambda_{\theta_\star}(P_{\theta_k} - P_{\theta_\star})\Lambda_{\theta_\star} f = n^{-1/2} \sum_{k=1}^{n} \pi_{\theta_k}\big(G_{\theta_k}^f\big),$$

where

$$G_\theta^f(x) \stackrel{\text{def}}{=} (P_\theta - P_{\theta_\star})\Lambda_{\theta_\star}(P_\theta - P_{\theta_\star})\Lambda_{\theta_\star} f(x). \tag{38}$$

We write for any $x \in \mathsf{X}$ and any $\ell_k \in \mathbb{N}$,

$$\pi_{\theta_k}\big(G_{\theta_k}^f\big) = \big(\pi_{\theta_k} - P_{\theta_k}^{\ell_k}\big)G_{\theta_k}^f(x) + \big(P_{\theta_k}^{\ell_k} - P_{\theta_\star}^{\ell_k}\big)G_{\theta_k}^f(x) + P_{\theta_\star}^{\ell_k} G_{\theta_k}^f(x).$$

By Proposition 3.3, P[α] holds and there exist C_θ, ρ_θ such that $\|P_\theta^n - \pi_\theta\|_{V^\alpha} \le C_\theta \rho_\theta^n$. Furthermore, Lemma A.2 and I3(b) imply that $\limsup_n C_{\theta_n} < +\infty$ w.p.1 and there exists a constant $\rho \in (0,1)$ such that $\limsup_n \rho_{\theta_n} \le \rho$, w.p.1. Set $\ell_k \stackrel{\text{def}}{=} \lfloor \ell \ln k \rfloor$ with ℓ such that $1/2 + \ell \ln \rho < 0$. Let $x \in \mathsf{X}$ be fixed.

By Lemma 4.3 and I3(b), there exists an almost surely finite random variable C_1 s.t.

$$\left|\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \big(\pi_{\theta_k} - P_{\theta_k}^{\ell_k}\big)G_{\theta_k}^f(x)\right| \le C_1 V^\alpha(x) n^{-1/2} \sum_{k=1}^{n} \rho^{\ell_k}.$$

Since $n^{-1/2} \sum_{k=1}^{n} \rho^{\ell_k} \le \rho^{-1} n^{-1/2} \sum_{k=1}^{n} k^{\ell \ln \rho} \to_{n\to\infty} 0$, it holds

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \big(\pi_{\theta_k} - P_{\theta_k}^{\ell_k}\big)G_{\theta_k}^f(x) \xrightarrow{\text{a.s.}} 0.$$

By Lemma 4.5, there exist some positive constants $C_2, \kappa_\star, a$ such that

$$\mathbb{E}\left[\left(\sum_{k=1}^{n}\{P_{\theta_k}^{\ell_k}-P_{\theta_\star}^{\ell_k}\}G_{\theta_k}^{f}(x)\right)^2\right]^{1/2}\le C_2|f|_{V^\alpha}V^\alpha(x)\sum_{k=1}^{n}\frac{1}{k}\sum_{t=1}^{\ell_k-1}\left(\frac{\kappa_\star\ell_k}{k^{1/(2a)}}\right)^{at}.$$

Since $\lim_k \ell_k^a/k^{1/2}=0$, there exists $k_\star$ such that for $k\ge k_\star$, $(\kappa_\star\ell_k)^a/k^{1/2}\le 1/2$. Then,

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\frac{1}{k}\sum_{t=1}^{\ell_k}\left(\frac{\kappa_\star\ell_k}{k^{1/(2a)}}\right)^{at}\le\frac{1}{\sqrt{n}}\sum_{k=1}^{k_\star}\frac{1}{k}\sum_{t=1}^{\lceil\ell\ln k\rceil}\left(\frac{\kappa_\star\ell_k}{k^{1/(2a)}}\right)^{at}+\frac{2}{\sqrt{n}}\sum_{k=k_\star+1}^{n}\frac{1}{k}.$$

The RHS tends to zero when $n\to+\infty$, which proves that $n^{-1/2}\sum_{k=1}^{n}\{P_{\theta_k}^{\ell_k}-P_{\theta_\star}^{\ell_k}\}G_{\theta_k}^{f}(x)\xrightarrow{\mathbb{P}}0$.

Finally, by Lemma 4.6, there exists a constant C_3 such that

$$\mathbb{E}\left[\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n}P_{\theta_\star}^{\ell_k}G_{\theta_k}^{f}(x)\right)^2\right]^{1/2}\le C_3V^\alpha(x)\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\frac{\ell_k^\alpha}{k}\underset{n\to\infty}{\to}0,$$

thus implying that $n^{-1/2}\sum_{k=1}^{n}P_{\theta_\star}^{\ell_k}G_{\theta_k}^{f}(x)\xrightarrow{\mathbb{P}}0$.

Lemma 4.3. *Assume* I1 *and* I2(a)–(c). *Let* $\alpha\in(0,1/2)$. *For any* $f\in\mathcal{L}_{V^\alpha}$ *and* $\theta\in\Theta$,

$$G_\theta^f(x)=\int(\theta-\theta_\star)^{\otimes 2}(\mathrm{d}z_{1:2})F^{(0)}(x,z_1,z_2),$$

where G_θ^f *is defined by* (38); *and there exists a constant* C *such that for any* $x\in\mathsf{X}$,

$$\left|F^{(0)}(x,z_1,z_2)\right|\le C|f|_{V^\alpha}V^{\alpha\wedge(\beta/\tau)}(x)\big(V^\alpha(z_1)+V^\alpha(z_2)\big).$$

In addition, there exists some constant C' *such that for any* $\ell\in\mathbb{N}$, *any* $\theta\in\Theta$ *and any* $f\in\mathcal{L}_{V^\alpha}$,

$$\left|\big(\pi_\theta-P_\theta^\ell\big)G_\theta^f\right|_{V^\alpha}\le C'|f|_{V^\alpha}\left\|P_\theta^\ell-\pi_\theta\right\|_{V^\alpha}\theta\big(V^\alpha\big).$$

Proof. Set $\gamma\stackrel{\text{def}}{=}\alpha\wedge(\beta/\tau)$. Throughout this proof, let L_θ be the constant given by P[γ]. We have

$$\begin{aligned}F^{(0)}(x,z_1,z_2)\stackrel{\text{def}}{=}\varepsilon^2r(x,z_2)\Big[&\int\Lambda_{\theta_\star}(z_2,\mathrm{d}y)r(y,z_1)\big(\Lambda_{\theta_\star}f(z_1)-\Lambda_{\theta_\star}f(y)\big)\\&-\int\Lambda_{\theta_\star}(x,\mathrm{d}y)r(y,z_1)\big(\Lambda_{\theta_\star}f(z_1)-\Lambda_{\theta_\star}f(y)\big)\Big].\end{aligned}$$

Note that $|r(\cdot,z_1)|_{V^\gamma}\le 1$ for any z_1 so that by (29),

$$\left|\int\Lambda_{\theta_\star}(z_2,\mathrm{d}y)r(y,z_1)\Lambda_{\theta_\star}f(z_1)\right|\le L_{\theta_\star}^4|f|_{V^\alpha}V^\alpha(z_1)V^\gamma(z_2).$$

In addition, since $\gamma - \beta/\tau \le 0$, we have by definition of the acceptance ratio r (see (22))

$$r(x, z_2) V^{\gamma}(z_2) \le V^{\gamma}(x).$$

Then, there exists a constant C such that

$$\varepsilon^2 r(x, z_2) \left| \int \Lambda_{\theta_\star}(z_2, \mathrm{d}y) r(y, z_1) \Lambda_{\theta_\star} f(z_1) \right| \le C |f|_{V^\alpha} V^{\alpha}(z_1) V^{\gamma}(x).$$

Similar upper bounds can be obtained for the three remaining terms in $F^{(0)}$, thus showing the upper bounds on $F^{(0)}$.

In addition, by P$[\gamma]$

$$\left|\left(\pi_\theta - P_\theta^{\ell}\right) G_\theta^f f(x)\right|_{V^\alpha} \le \left\|\pi_\theta - P_\theta^{\ell}\right\|_{V^\alpha} \left|G_\theta^f f\right|_{V^\alpha} V^{\alpha}(x).$$

The proof is concluded upon noting that $|G_\theta^f(x)| \le C |f|_{V^\alpha} \theta(V^\alpha)$. □

Lemma 4.4. *Assume* I1 *and* I2(a)–(c). *Let* $\alpha \in (0, 1/2)$. *There exist some constants* $C, \kappa_\star$ *and* $\rho_\star \in (0, 1)$ *such that for any* $t \ge 1$, *any integers* $u_1, \ldots, u_t$ *and any* $f \in \mathcal{L}_{V^\alpha}$,

$$\begin{aligned}
&(P_\theta - P_{\theta_\star})\left(P_{\theta_\star}^{u_t} - \pi_{\theta_\star}\right) \cdots (P_\theta - P_{\theta_\star})\left(P_{\theta_\star}^{u_1} - \pi_{\theta_\star}\right) G_\theta^f(x) \\
&\quad = \int \cdots \int (\theta - \theta_\star)^{\otimes(t+2)}(\mathrm{d}z_{1:t+2}) F_{u_{1:t}}^{(t)}(x, z_1, \ldots, z_{t+2}),
\end{aligned}$$

where G_θ^f *is defined in* (38), *and*

$$\left|F_{u_{1:t}}^{(t)}(x, z_1, \ldots, z_{t+2})\right| \le C |f|_{V^\alpha} \kappa_\star^t \rho_\star^{\sum_{j=1}^t u_j} V^{\alpha \wedge (\beta/\tau)}(x) \sum_{j=1}^{t+2} V^{\alpha}(z_j). \tag{39}$$

Proof. By repeated applications of Equation (37), it can be proved that the functions $F_{u_{1:t}}^{(t)}$ are recursively defined as follows

$$\begin{aligned}
&F_{u_{1:t}}^{(t)}(x, z_1, \ldots, z_{t+2}) \\
&\quad \stackrel{\text{def}}{=} \varepsilon r(x, z_{t+2}) \int \left(P_{\theta_\star}^{u_t}(z_{t+2}, \mathrm{d}y) - P_{\theta_\star}^{u_t}(x, \mathrm{d}y)\right) F_{u_{1:t-1}}^{(t-1)}(y, z_1, \ldots, z_{t+1}),
\end{aligned} \tag{40}$$

where $F_{u_{1:0}}^{(0)} = F^{(0)}$ and $F^{(0)}$ is given by Lemma 4.3.

The proof of the upper bound is by induction. The property holds for $t = 1$. Assume it holds for $t \ge 2$. Set $\gamma \stackrel{\text{def}}{=} \alpha \wedge (\beta/\tau)$; by Proposition 3.3 and the property P$[\gamma]$, there exist some constants

$C_\star$ and $\rho_\star \in (0,1)$ such that $\| P^\ell_{\theta_\star} - \pi_{\theta_\star} \|_{V^\gamma} \le C_{\theta_\star} \rho^\ell_{\theta_\star}$. Then,

$$\begin{aligned}\left|F^{(t)}_{u_{1:t}}(x, z_{1:t+2})\right| &\le C|f|_{V^\alpha} \kappa_\star^{t-1} \rho_{\theta_\star}^{\sum_{j=1}^{t-1} u_j} \left(\sum_{j=1}^{t+1} V^\alpha(z_j)\right) \\ &\quad \times r(x, z_{t+2})\left[\left\| P^{u_t}_{\theta_\star} - \pi_{\theta_\star}\right\|_{V^\gamma} V^\gamma(z_{t+2}) + \left\| P^{u_t}_{\theta_\star} - \pi_{\theta_\star}\right\|_{V^\gamma} V^\gamma(x)\right] \\ &\le C|f|_{V^\alpha} \kappa_\star^{t-1} \varepsilon C_{\theta_\star} \rho_{\theta_\star}^{\sum_{j=1}^{t} u_j} r(x, z_{t+2})\left\{V^\gamma(z_{t+2}) + V^\gamma(x)\right\}.\end{aligned}$$

Since $\gamma \le \beta/\tau$, $r(x, z_{t+2}) V^\gamma(z_{t+2}) \le V^\gamma(x)$ thus showing (39) with $\kappa_\star = 2C_{\theta_\star}\varepsilon$. □

Lemma 4.5. *Assume* I1, I2(a)–(c) *and* I3. *Let* $\alpha \in (0, 1/2)$. *There exist positive constants* C, κ, a *such that for any* $f \in \mathcal{L}_{V^\alpha}$, *any* $k, \ell \ge 1$ *and any* $x \in \mathsf{X}$,

$$\mathbb{E}\left[\left(\left\{P^\ell_{\theta_k} - P^\ell_{\theta_\star}\right\} G^f_{\theta_k}(x)\right)^2\right]^{1/2} \le C|f|_{V^\alpha} \frac{V^\alpha(x)}{k} \sum_{t=1}^{\ell-1} \left(t\kappa k^{-1/(2a)}\right)^{at},$$

where G^f_θ *is given by* (38).

Proof. For any $g \in \mathcal{L}_{V^\alpha}$, $k, \ell \ge 1$ and $x \in \mathsf{X}$,

$$\begin{aligned}&P^\ell_{\theta_k} g(x) - P^\ell_{\theta_\star} g(x) \\ &\quad = \sum_{t=1}^{\ell-1} \sum_{u_{1:t} \in \mathcal{U}_t} P_{\theta_\star}^{\ell - t - \sum_{j=1}^t u_j} (P_{\theta_k} - P_{\theta_\star}) P^{u_t}_{\theta_\star} \cdots (P_{\theta_k} - P_{\theta_\star}) P^{u_1}_{\theta_\star} g(x) \\ &\quad = \sum_{t=1}^{\ell-1} \sum_{u_{1:t} \in \mathcal{U}_t} P_{\theta_\star}^{\ell - t - \sum_{j=1}^t u_j} (P_{\theta_k} - P_{\theta_\star})\left(P^{u_t}_{\theta_\star} - \pi_{\theta_\star}\right) \cdots (P_{\theta_k} - P_{\theta_\star})\left(P^{u_1}_{\theta_\star} - \pi_{\theta_\star}\right) g(x),\end{aligned}$$

where $\mathcal{U}_t = \{u_{1:t}, u_j \in \mathbb{N}, \sum_{j=1}^t u_j \le \ell - t\}$. Fix $t \in \{1, \ldots, \ell - 1\}$ and $u_{1:t} \in \mathcal{U}_t$. Then by Lemma 4.4,

$$\begin{aligned}&P_{\theta_\star}^{\ell - t - \sum_{j=1}^t u_j} (P_{\theta_k} - P_{\theta_\star})\left(P^{u_t}_{\theta_\star} - \pi_{\theta_\star}\right) \cdots (P_{\theta_k} - P_{\theta_\star})\left(P^{u_1}_{\theta_\star} - \pi_{\theta_\star}\right) G^f_{\theta_k}(x) \\ &\quad = \int (\theta_k - \theta_\star)^{\otimes(t+2)}(\mathrm{d}z_{1:t+2}) \int P_{\theta_\star}^{\ell - t - \sum_{j=1}^t u_j}(x, \mathrm{d}y) F^{(t)}_{u_{1:t}}(y, z_1, \ldots, z_{t+2}).\end{aligned}$$

Assumptions I3(b) and I3(d) and Lemma 4.4 show that there exist constants $C, \kappa_\star, \rho_\star \in (0, 1)$ such that

$$\begin{aligned}&\left\| \int (\theta_k - \theta_\star)^{\otimes(t+2)}(\mathrm{d}z_{1:t+2}) \int P_{\theta_\star}^{\ell - t - \sum_{j=1}^t u_j}(x, \mathrm{d}y) F^{(t)}_{u_{1:t}}(y, z_1, \ldots, z_{t+2}) \right\|_2 \\ &\quad \le \frac{C}{k^{1+t/2}} A_t |f|_{V^\alpha} \kappa_\star^t \rho_\star^{\sum_{j=1}^t u_j} P_{\theta_\star}^{\ell - t - \sum_{j=1}^t u_j} V^\alpha(x).\end{aligned}$$

Finally, Proposition 3.3 implies that $\sup_{j\geq 0}|P^j_{\theta_\star}V^\alpha|_{V^\alpha} < +\infty$. By combining these results, we have for some constant C

$$\|P^\ell_{\theta_k}G^f_{\theta_k}(x) - P^\ell_{\theta_\star}G^f_{\theta_k}(x)\|_2 \leq Ck^{-1}|f|_{V^\alpha}V^\alpha(x)\sum_{t=1}^{\ell-1}A_t\kappa_\star^t k^{-t/2}\sum_{u_{1:t}\in\mathcal{U}_t}\rho_\star^{\sum_{j=1}^t u_j}.$$

Note that $\sum_{u_{1:t}\in\mathcal{U}_t}\rho_\star^{\sum_{j=1}^t u_j} \leq (1-\rho_\star)^{-t}$. Furthermore, there exists $a>0$ such that $A_t \leq t^{at}$. Therefore,

$$\begin{aligned}&\|P^\ell_{\theta_k}G^f_{\theta_k}(x) - P^\ell_{\theta_\star}G^f_{\theta_k}(x)\|_2\\ &\quad\leq Ck^{-1}|f|_{V^\alpha}V^\alpha(x)\sum_{t=1}^{\ell-1}\big(t\kappa^{1/a}(1-\rho_\star)^{-1/a}k^{-1/(2a)}\big)^{at}.\end{aligned}$$

This concludes the proof. □

Lemma 4.6. *Assume* I1, I2(a)–(c) *and* I3. *Let* $\alpha\in(0,1/2)$ *and* $f\in\mathcal{L}_{V^\alpha}$. *Then, there exists a constant* C *such that for any* $k,\ell\geq 1$ *and any* $x\in\mathsf{X}$,

$$\mathbb{E}\big[\big(P^\ell_{\theta_\star}G^f_{\theta_k}(x)\big)^2\big]^{1/2} \leq C\ell^\alpha|f|_{V^\alpha}k^{-1}V^\alpha(x).$$

Proof. We have

$$P^\ell_{\theta_\star}G^f_{\theta_k}(x) = \int\int(\theta_k-\theta_\star)^{\otimes 2}(\mathrm{d}z_{1:2})H_\ell(x,z_1,z_2),$$

with $H_\ell(x,z_1,z_2) \stackrel{\text{def}}{=} P^\ell_{\theta_\star}(x,F^{(0)}(\cdot,z_1,z_2))$ where $F^{(0)}$ is given by Lemma 4.3. Lemma 4.3 also implies that there exists a constant C such that

$$\big|H_\ell(x,z_1,z_2)\big| \leq C|f|_{V^\alpha}\big(V^\alpha(z_1)+V^\alpha(z_2)\big)P^\ell_{\theta_\star}V^\alpha(x). \tag{41}$$

By I3, the variance of $P^\ell_{\theta_\star}G^f_{\theta_k}(x)$ is upper bounded by

$$C|f|^2_{V^\alpha}\big(P^\ell_{\theta_\star}V^\alpha(x)\big)^2k^{-2}.$$

The proof is concluded by application of the drift inequality (27) and I3(a). □

Appendix A: Technical lemmas

The following lemma is (slightly) adapted from [18], Lemma 4.2.

Lemma A.1. *Assume A2. For any $\alpha \in (0,1)$ and $\theta, \theta' \in \Theta$,*

$$\|\pi_\theta - \pi_{\theta'}\|_{V^\alpha} \le 2(L_{\theta'} \vee L_\theta)^4 \pi_\theta\big(V^\alpha\big) D_{V^\alpha}\big(\theta, \theta'\big),$$

$$\|\Lambda_\theta - \Lambda_{\theta'}\|_{V^\alpha} \le 3(L_\theta \vee L_{\theta'})^6 \pi_\theta\big(V^\alpha\big) D_{V^\alpha}\big(\theta, \theta'\big),$$

$$\|P_\theta\Lambda_\theta - P_{\theta'}\Lambda_{\theta'}\|_{V^\alpha} \le 5(L_\theta \vee L_{\theta'})^6 \pi_\theta\big(V^\alpha\big) D_{V^\alpha}\big(\theta, \theta'\big),$$

where L_θ and Λ_θ are given by (9) and (3).

The following lemma can be obtained from [16,17,28] or [11] (see also the proof of [32], Lemma 3, for a similar result).

Lemma A.2. *Let $\{P_\theta, \theta \in \Theta\}$ be a family of phi-irreducible and aperiodic Markov kernels. Assume that there exist a function $V : \mathsf{X} \to [1, +\infty)$, and for any $\theta \in \Theta$ there exist some constants $b_\theta < +\infty$, $\delta_\theta > 0$, $\lambda_\theta \in (0,1)$ and a probability measure ν_θ on X such that for any $x \in \mathsf{X}$*

$$P_\theta V(x) \le \lambda_\theta V(x) + b_\theta,$$

$$P_\theta(x, \cdot) \ge \delta_\theta \nu_\theta(\cdot) \mathbb{1}_{\{V \le c_\theta\}}(x), \qquad c_\theta \stackrel{def}{=} 2b_\theta(1-\lambda_\theta)^{-1} - 1.$$

Then there exists $\gamma > 0$ and for any θ, there exist some finite constants C_θ and $\rho_\theta \in (0,1)$ such that

$$\big\| P_\theta^n(x, \cdot) - \pi_\theta \big\|_V \le C_\theta \rho_\theta^n V(x)$$

and

$$C_\theta \vee (1-\rho_\theta)^{-1} \le C\big\{b_\theta \vee \delta_\theta^{-1} \vee (1-\lambda_\theta)^{-1}\big\}^\gamma.$$

Lemma A.3 is proved in [18], Section 4.

Lemma A.3. *Let X be a Polish space endowed with its Borel σ-field $\mathcal{X}$. Let μ and $(mu_n)_{n\in\mathbb{N}}$ be probability distributions on $(\mathsf{X}, \mathcal{X})$. Let $(h_n)_{n\in\mathbb{N}}$ be an equicontinuous family of functions from X to $\mathbb{R}$. Assume*

(i) *the sequence $(\mu_n)_{n\in\mathbb{N}}$ converges weakly to μ,*
(ii) *for any $x \in \mathsf{X}$, $\lim_{n\to\infty} h_n(x)$ exists, and there exists $\gamma > 1$ such that $\sup_n \mu_n(|h_n|^\gamma) + \mu(|\lim_n h_n|) < +\infty$.*

Then, $\mu_n(h_n) \to \mu(\lim_n h_n)$.

Appendix B: Weak law of large numbers for adaptive and interacting MCMC algorithms

The proof of the theorem below is along the same lines as the proof of [18], Theorem 2.7, which addresses the strong law of large numbers and details are omitted. Note that in this generalization, we relax the condition $\sup_\theta |F(\cdot, \theta)|_V < +\infty$ of [18]. The proof is provided in the supplementary material [19].

Theorem B.1. *Assume* A1, A2 *and let* $a \in (0,1)$. *Let* $F : \mathsf{X} \times \Theta \to \mathbb{R}$ *be a measurable function. Assume that there exists a sequence of stopping-times* $\{\tau_m, m \geq 1\}$ *such that* $\mathbb{P}(\bigcup_m \{\tau_m = +\infty\}) = 1$ *and*

(i) $\limsup_{n\to\infty} L_{\theta_n} < \infty$, $\mathbb{P}$-*a.s. where* L_θ *is defined in Lemma* 2.1 *applied with the closed interval* $[a, 1]$.
(ii) $\limsup_{n\to\infty} \pi_{\theta_n}(V^a) < \infty$, $\mathbb{P}$-*a.s.*
(iii) $\limsup_{n\to\infty} |F_{\theta_n}|_{V^a} < +\infty$, $\mathbb{P}$-*a.s.*
(iv) *For any* $m \geq 1$, *there exists* $t < 1/a - 1$ *such that* $\sup_{n\geq 1} n^{-t} \mathbb{E}[V(X_n) \mathbb{1}_{\{n-1<\tau_m\}}] < \infty$.
(v) $n^{-1} \sum_{k=1}^n D_{V^a}(\theta_k, \theta_{k-1}) V^a(X_k) \xrightarrow{\mathbb{P}} 0$.
(vi) $n^{-1} \sum_{k=1}^{n-1} |F_{\theta_k} - F_{\theta_{k-1}}|_{V^a} V^a(X_k) \xrightarrow{\mathbb{P}} 0$.

Then,

$$\frac{1}{n} \sum_{k=0}^{n-1} F_{\theta_k}(X_k) - \frac{1}{n} \sum_{k=0}^{n-1} \int \pi_{\theta_k}(\mathrm{d}x) F_{\theta_k}(x) \xrightarrow{\mathbb{P}} 0.$$

Acknowledgement

This work is partially supported by the French National Research Agency, under the Program ANR-08-BLAN-0218 BigMC.

Supplementary Material

Supplement to "A central limit theorem for adaptive and interacting Markov chains" (DOI: 10.3150/12-BEJ493SUPP; pdf). We detail in this supplement: (1) the gap in the proof of Atchade's [9] theorem, (2) the proofs of technical Lemmas 4.1, 4.3, A.1–A.3, (3) some additional proofs of [18], Section 3.1, (4) results on the variance of completely degenerated V-statistics of asymptotically stationary Markov chains, and (5) the weak law of large number for adaptive and interacting Markov chains.

References

[1] Andrieu, C., Jasra, A., Doucet, A. and Del Moral, P. (2007). Convergence of the equi-energy sampler. In *Conference Oxford sur les Méthodes de Monte Carlo Séquentielles*. *ESAIM Proc*. **19** 1–5. Les Ulis: EDP Sci. MR2405642
[2] Andrieu, C., Jasra, A., Doucet, A. and Del Moral, P. (2007). Non-linear Markov chain Monte Carlo. In *Conference Oxford sur les Méthodes de Monte Carlo Séquentielles*. *ESAIM Proc*. **19** 79–84. Les Ulis: EDP Sci. MR2405652
[3] Andrieu, C., Jasra, A., Doucet, A. and Del Moral, P. (2008). A note on convergence of the equi-energy sampler. *Stoch*. *Anal*. *Appl*. **26** 298–312. MR2399737
[4] Andrieu, C., Jasra, A., Doucet, A. and Del Moral, P. (2011). On nonlinear Markov chain Monte Carlo. *Bernoulli* **17** 987–1014. MR2817614

[5] Andrieu, C. and Moulines, É. (2006). On the ergodicity properties of some adaptive MCMC algorithms. *Ann. Appl. Probab.* **16** 1462–1505. MR2260070
[6] Andrieu, C. and Thoms, J. (2008). A tutorial on adaptive MCMC. *Stat. Comput.* **18** 343–373. MR2461882
[7] Atchadé, Y. and Fort, G. (2010). Limit theorems for some adaptive MCMC algorithms with subgeometric kernels. *Bernoulli* **16** 116–154. MR2648752
[8] Atchadé, Y., Fort, G., Moulines, E. and Priouret, P. (2011). Adaptive Markov chain Monte Carlo: Theory and methods. In *Bayesian Time Series Models* 32–51. Cambridge: Cambridge Univ. Press. MR2894232
[9] Atchadé, Y.F. (2010). A cautionary tale on the efficiency of some adaptive Monte Carlo schemes. *Ann. Appl. Probab.* **20** 841–868. MR2680550
[10] Atchadé, Y.F. (2011). Kernel estimators of asymptotic variance for adaptive Markov chain Monte Carlo. *Ann. Statist.* **39** 990–1011. MR2816345
[11] Baxendale, P.H. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.* **15** 700–738. MR2114987
[12] Bercu, B., Del Moral, P. and Doucet, A. (2009). A functional central limit theorem for a class of interacting Markov chain Monte Carlo methods. *Electron. J. Probab.* **14** 2130–2155. MR2550295
[13] Brockwell, A., Del Moral, P. and Doucet, A. (2010). Sequentially interacting Markov chain Monte Carlo methods. *Ann. Statist.* **38** 3387–3411. MR2766856
[14] Del Moral, P. and Doucet, A. (2010). Interacting Markov chain Monte Carlo methods for solving nonlinear measure-valued equations. *Ann. Appl. Probab.* **20** 593–639. MR2650043
[15] Douc, R. and Moulines, E. (2008). Limit theorems for weighted samples with applications to sequential Monte Carlo methods. *Ann. Statist.* **36** 2344–2376. MR2458190
[16] Douc, R., Moulines, E. and Rosenthal, J.S. (2004). Quantitative bounds on convergence of time-inhomogeneous Markov chains. *Ann. Appl. Probab.* **14** 1643–1665. MR2099647
[17] Fort, G. and Moulines, E. (2003). Polynomial ergodicity of Markov transition kernels. *Stochastic Process. Appl.* **103** 57–99. MR1947960
[18] Fort, G., Moulines, E. and Priouret, P. (2011). Convergence of adaptive and interacting Markov chain Monte Carlo algorithms. *Ann. Statist.* **39** 3262–3289.
[19] Fort, G., Moulines, E., Priouret, P. and Vandekerkhove, P. (2012). A simple variance inequality for U-statistics of a Markov chain with applications. *Statist. Probab. Lett.* **82** 1193–1201. MR2915087
[20] Fort, G., Moulines, E., Priouret, P. and Vandekerkhove, P. (2014). Supplement to "A central limit theorem for adaptive and interacting Markov chains." DOI:10.3150/12-BEJ493SUPP.
[21] Haario, H., Saksman, E. and Tamminen, J. (1999). Adaptive proposal distribution for random walk Metropolis algorithm. *Comput. Statist.* **14** 375–395.
[22] Hall, P. and Heyde, C.C. (1980). *Martingale Limit Theory and Its Application. Probability and Mathematical Statistics*. New York: Academic Press [Harcourt Brace Jovanovich Publishers]. MR0624435
[23] Jarner, S.F. and Hansen, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.* **85** 341–361. MR1731030
[24] Kou, S.C., Zhou, Q. and Wong, W.H. (2006). Equi-energy sampler with applications in statistical inference and statistical mechanics. *Ann. Statist.* **34** 1581–1652. MR2283711
[25] Liang, F., Liu, C. and Carroll, R.J. (2007). Stochastic approximation in Monte Carlo computation. *J. Amer. Statist. Assoc.* **102** 305–320. MR2345544
[26] Mengersen, K.L. and Tweedie, R.L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24** 101–121. MR1389882
[27] Meyn, S. and Tweedie, R.L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge: Cambridge Univ. Press. With a prologue by Peter W. Glynn. MR2509253

[28] Roberts, G.O. and Rosenthal, J.S. (2004). General state space Markov chains and MCMC algorithms. *Probab. Surv.* **1** 20–71. MR2095565
[29] 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
[30] Roberts, G.O. and Tweedie, R.L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83** 95–110. MR1399158
[31] Rosenthal, J.S. (2011). Optimal proposal distributions and adaptive MCMC. In *Handbook of Markov Chain Monte Carlo*. *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 93–111. Boca Raton, FL: CRC Press. MR2858446
[32] Saksman, E. and Vihola, M. (2010). On the ergodicity of the adaptive Metropolis algorithm on unbounded domains. *Ann. Appl. Probab.* **20** 2178–2203. MR2759732
[33] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. *Wiley Series in Probability and Mathematical Statistics*. New York: Wiley. MR0595165
[34] Tierney, L. (1998). A note on Metropolis–Hastings kernels for general state spaces. *Ann. Appl. Probab.* **8** 1–9. MR1620401
[35] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*: *With Applications to Statistics*. *Springer Series in Statistics*. New York: Springer. MR1385671
[36] Wang, F. and Landau, D.P. (2001). An efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.* **86** 2050–2053.

Received December 2011 and revised July 2012