CONVERGENCE OF ADAPTIVE AND INTERACTING MARKOV CHAIN MONTE CARLO ALGORITHMS

BY G. FORT[1], E. MOULINES[1] AND P. PRIOURET

*LTCI, TELECOM ParisTech-CNRS, LTCI, TELECOM ParisTech-CNRS
and LPMA, Université Pierre et Marie Curie*

Adaptive and interacting Markov chain Monte Carlo algorithms (MCMC) have been recently introduced in the literature. These novel simulation algorithms are designed to increase the simulation efficiency to sample complex distributions. Motivated by some recently introduced algorithms (such as the adaptive Metropolis algorithm and the interacting tempering algorithm), we develop a general methodological and theoretical framework to establish both the convergence of the marginal distribution and a strong law of large numbers. This framework weakens the conditions introduced in the pioneering paper by Roberts and Rosenthal [*J. Appl. Probab.* **44** (2007) 458–475]. It also covers the case when the target distribution π is sampled by using Markov transition kernels with a stationary distribution that differs from π.

1. Introduction. Markov chain Monte Carlo (MCMC) methods generate samples from an arbitrary distribution π known up to a scaling factor; see Robert and Casella (2004). The algorithm consists in sampling a Markov chain $\{X_n, n \geq 0\}$ on a general state space X with Markov transition kernel P admitting π as its unique *invariant* distribution.

In most implementations of MCMC algorithms, the transition kernel P of the Markov chain depends on a tuning parameter θ defined on a space Θ which can be either finite dimensional or infinite dimensional.

Consider, for example, the Metropolis algorithm [Metropolis et al. (1953)]. Here $X = \mathbb{R}^d$ and the stationary distribution is assumed to have a density, also denoted by π with respect to a measure. At the iteration n, a move $Z_{n+1} = X_n + U_{n+1}$ is proposed, where U_{n+1} is drawn independently from $X_0, \ldots, X_n$ from a symmetric distribution on $\mathbb{R}^d$. This move is accepted with probability $\alpha(X_n, Z_{n+1})$, where $\alpha(x, y) = 1 \wedge (\pi(y)/\pi(x))$. A frequently advocated choice of the increment distribution q is the multivariate normal with zero-mean and covariance ma-

trix $(2.38^2/d)\Gamma_\star$, where $\Gamma_\star$ is the covariance matrix of the target distribution π [see Gelman, Roberts and Gilks (1996)].

Of course $\Gamma_\star$ is unknown. In Haario, Saksman and Tamminen (1999), the authors have proposed an *adaptive Metropolis* (AM) algorithm in which the covariance Γ_n is updated at each iteration using the past values of the simulations [see also Haario, Saksman and Tamminen (2001), Haario et al. (2004, 2006), Laine and Tamminen (2008) for applications].

The adaptive Metropolis is an example in which a parameter θ_{n+1} is updated at each iteration from the values of the chain $\{X_0, \ldots, X_{n+1}\}$ and the past values of the parameters $\{\theta_0, \ldots, \theta_n\}$. Many other examples of such adaptive MCMC algorithms are presented in Andrieu and Thoms (2008), Rosenthal (2009) and Atchadé et al. (2011).

When attempting to simulate from a density with multiple modes, the Markov kernel might mix very slowly. A useful solution to that problem is to introduce a temperature parameter. This idea is exploited in parallel tempering: several Metropolis algorithms are run at different temperatures [see Geyer (1991), Atchade, Roberts and Rosenthal (2011)]. One of the simulations, corresponding to $T_1 = 1$ is the desired target probability distribution. The other simulations correspond to the family of the target distribution π^{1/T_i}, $i \in \{1, \ldots, K\}$, created by gradually increasing the temperature.

The *interacting tempering* algorithm, a simplified form of the equi-energy sampler introduced Kou, Zhou and Wong (2006), exploits the parallel tempering idea. Both the algorithms run several chains in parallel, but the interacting tempering algorithm allows more general interactions between chains. The interacting tempering algorithm provides an example in which the process of interest interacts with the past samples of a family of auxiliary processes. Other examples of such interacting schemes are presented in Andrieu et al. (2007) [see also Brockwell, Del Moral and Doucet (2010)].

The two examples discussed above can be put into a common unifying framework (see Section 2). The purpose of this work is to analyze these general classes of adaptive and interacting MCMC. This paper complements recent surveys on this topic by Andrieu and Thoms (2008), Rosenthal (2009) and Atchadé et al. (2011) which are devoted to the design of these algorithms. We focus in this paper on two problems: the ergodicity of the sampler (under which condition the marginal distribution of the process converges to the target distribution π) and the strong law of large numbers (SLLN) for additive and unbounded functionals.

Ergodicity of the marginal distributions for adaptive MCMC has been studied by Andrieu and Moulines (2006) for a particular class of samplers in which the parameter is adapted using a stochastic approximation algorithm. These results have later been extended by Roberts and Rosenthal (2007) to handle more general adaptation strategies, but under conditions which are in some respects more stringent. Most of these works assume a form of geometric ergodicity; these conditions are

relaxed in Atchadé and Fort (2010) which addresses Markov chains with subgeometric rate of convergence.

A strong law of large number for the adaptive Metropolis algorithm was established by Haario, Saksman and Tamminen (2001) (for bounded functions and a compact parameter space Θ), using mixingales techniques; these results have later been extended by Atchadé and Rosenthal (2005) to unbounded functions and compact parameter space Θ. The LLN for unbounded functions and noncompact set Θ has been established recently in Saksman and Vihola (2010). Andrieu and Moulines (2006) have established the consistency and the asymptotic normality of $n^{-1} \sum_{k=1}^{n} f(X_k)$ for bounded and unbounded functions for adaptive MCMC algorithms combined with a stochastic approximation procedure [see Atchadé and Fort (2010) for extensions]. The procedure involves projections on a family of increasing compact subsets of the parameter space, and did not include the results obtained for the AM by Saksman and Vihola (2010).

Roberts and Rosenthal (2007) prove a weak law of large numbers for bounded functions for general adaptive MCMC samplers but under technical conditions which are stringent.

The analysis of interacting MCMC algorithms started more recently and the theory is still less developed. The original result in Kou, Zhou and Wong [(2006), Theorem 2], as already noted in the discussion paper [Atchadé and Liu (2006), Section 3] and carefully explained in Andrieu et al. [(2008), Section 3.1] does not amount to a proof. Andrieu et al. (2008) presents a proof of convergence of a simple version of the interacting tempering sampler with $K = 2$ stages. The proofs in Andrieu et al. (2008) (uniformly ergodic case) and in Andrieu et al. (2011) (geometrically ergodic case) are based on the convergence of U-statistics, which explains why the results obtained for $K = 2$ stages cannot easily be extended.

SLLN was established by Atchadé (2010) for a simple version of the interacting tempering algorithm for a transition kernel which is geometrically ergodic with uniformly controlled ergodicity constants, but the proof in this paper is not convincing [see Fort, Moulines and Priouret (2011), Section 1].

Finally, a functional Central Limit theorem was derived in Bercu, Del Moral and Doucet (2009) for a class of interacting Markov chains for uniformly ergodic Markov kernels.

This paper aims at providing a theory weakening some of the limitations mentioned above. Let $\{P_\theta, \theta \in \Theta\}$ be a family of transition kernels on X. We address the general framework when the target density π is approximated by the process $\{X_n, n \geq 0\}$ such that the conditional distribution of X_{n+1} given the past is given by $P_{\theta_n}(X_n, \cdot)$; $\{\theta_n, n \geq 0\}$ is the adapted process. There are two main contributions. First, we cover the case when the ergodicity of the transition kernels $\{P_\theta, \theta \in \Theta\}$ is not uniform along the path $\{\theta_n, n \geq 0\}$. The second novelty is that we address the case when the P_θ has an invariant distribution π_θ depending upon the parameter θ; in this case, the adaptation has to be such that $\{\pi_{\theta_n}, n \geq 0\}$ converges weakly to π (almost surely) and we provide sufficient conditions for this property to hold based

on the (almost sure) weak convergence of the transition kernels $\{P_{\theta_n}, n \geq 0\}$. Such conditions are crucial in many applications where π_θ is known to exist but has no explicit expression. Therefore, to generalize the results and include more realistic conditions, a more complex approach is required.

The paper is organized as follows. In Section 2, we establish the convergence of the marginal distribution and the strong law of large numbers for additive functionals for adaptive and interacting MCMC algorithms. These general results are applied to a running example, namely the adaptive Metropolis algorithm. The novel contribution is the application to the convergence of the interacting tempering algorithm [Kou, Zhou and Wong (2006)] in Section 3.

Notation.   Let $(\mathsf{X}, \mathcal{X})$ be a general state space [see, e.g., Meyn and Tweedie (2009), Chapter 3] and P be a Markov transition kernel. P acts on bounded functions f on X and on σ-finite positive measures μ on $\mathcal{X}$ via

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

For $n \in \mathbb{N}$, we will denote by P^n the n-iterated transition kernel defined by induction

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

with the convention that 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 and will be denoted by $\|f\|_\infty$. Let $\mathcal{L}_V$ be the set of functions such that $\|f\|_V < +\infty$. For two probability distributions μ_1, μ_2 on X, define the V-distance

$$\|\mu_1 - \mu_2\|_V \stackrel{\text{def}}{=} \sup_{\{f, \|f\|_V \leq 1\}} |\mu_1(f) - \mu_2(f)|.$$

When $V = 1$, the V-distance is the total variation distance and is denoted by $\|\mu_1 - \mu_2\|_{\text{TV}}$.

Denote by $\mathsf{C}_b(\mathsf{X})$ the class of bounded continuous functions from X to $\mathbb{R}$. Recall that a Markov transition kernel P on $(\mathsf{X}, \mathcal{X})$ is (weak) Feller if it maps $\mathsf{C}_b(\mathsf{X})$ to $\mathsf{C}_b(\mathsf{X})$.

A measurable set $A \in \mathcal{A}$ on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is said to be a $\mathbb{P}$-full set if $\mathbb{P}(A) = 1$.

2. Main results. Let $(\Theta, \mathcal{T})$ be a measurable space and $(X, \mathcal{X})$ a general state space. Let $\{P_\theta, \theta \in \Theta\}$ be a collection of Markov transition kernels indexed by θ in Θ, which can be either finite or infinite dimensional. We consider a $X \times \Theta$-valued process $\{(X_n, \theta_n), n \geq 0\}$ on a filtered probability space $(\Omega, \mathcal{A}, \{\mathcal{F}_n, n \geq 0\}, \mathbb{P})$. It is assumed that (X_n, θ_n) is $\mathcal{F}_n$-adapted and for any bounded measurable function f

$$(1) \qquad \mathbb{E}[f(X_{n+1})|\mathcal{F}_n] = P_{\theta_n} f(X_n).$$

2.1. *Ergodicity.* For $V : X \to [1, \infty)$ and $\theta, \theta' \in \Theta$, denote by $D_V(\theta, \theta')$ the V-variation of the kernels P_θ and $P_{\theta'}$

$$(2) \qquad D_V(\theta, \theta') \overset{\text{def}}{=} \sup_{x \in X} \frac{\|P_\theta(x, \cdot) - P_{\theta'}(x, \cdot)\|_V}{V(x)}.$$

When $V \equiv 1$, we use the simpler notation $D(\theta, \theta')$. Consider the following assumption:

A1 For any $\theta \in \Theta$, there exists a probability distribution π_θ such that $\pi_\theta P_\theta = \pi_\theta$.
A2 (a) For any $\varepsilon > 0$, there exists a nondecreasing sequence $\{r_\varepsilon(n), n \geq 0\}$ in $\mathbb{N} \setminus \{0\}$, such that $\limsup_{n \to \infty} r_\varepsilon(n)/n = 0$ and

$$\limsup_{n \to \infty} \mathbb{E}\big[\big\| P_{\theta_{n-r_\varepsilon(n)}}^{r_\varepsilon(n)}\big(X_{n-r_\varepsilon(n)}, \cdot\big) - \pi_{\theta_{n-r_\varepsilon(n)}}\big\|_{\text{TV}}\big] \leq \varepsilon.$$

(b) For any $\varepsilon > 0$, $\lim_{n \to \infty} \sum_{j=0}^{r_\varepsilon(n)-1} \mathbb{E}[D(\theta_{n-r_\varepsilon(n)+j}, \theta_{n-r_\varepsilon(n)})] = 0$, where D is defined in (2).

Assumption A2(a) is implied by the *containment condition* introduced in Roberts and Rosenthal (2007): for any $\varepsilon > 0$, the sequence $\{M_\varepsilon(X_n, \theta_n), n \geq 0\}$ is bounded in probability, where for $x \in X$, $\theta \in \Theta$,

$$(3) \qquad M_\varepsilon(x, \theta) \overset{\text{def}}{=} \inf\{n \geq 0, \|P_\theta^n(x, \cdot) - \pi_\theta\|_{\text{TV}} \leq \varepsilon\}.$$

In this case, it is easily checked that A2(a) is satisfied by setting $r_\varepsilon(n) = N$ for all $n \geq 0$, where N is large enough. Assumption A2(a) is weaker than the containment condition, because the sequence $\{r_\varepsilon(n), n \geq 0\}$ can grow to infinity. This is important in applications where it is not known a priori that the parameter sequence $\{\theta_n, n \geq 0\}$ stays in a region where the ergodicity constants are controlled uniformly. Examples of such applications are given in a toy example and a more realistic example below.

Assumption A2(b) requires that the amount of change vanishes as n goes to infinity at a rate which is matched with the rate at which the kernel converges to stationarity. If the kernel mixes uniformly fast along any parameter sequence $\{\theta_n, n \geq 0\}$, that is, $r_\varepsilon(n) = N$ for any $n \geq 0$ for some integer N, A2(b) is equivalent to the diminishing adaptation condition introduced in Roberts and Rosenthal (2007): $\{D(\theta_n, \theta_{n-1}), n \geq 1\}$ converges to zero in probability at any rate. On the other hand, if the ergodicity is not uniform along a sequence $\{\theta_n, n \geq 0\}$, then

the rate of convergence of the adaptation should converge to zero but with a fast enough rate. As expected, there is a trade-off between the rate of convergence of the chain and the rate at which the parameter can be adapted. This does not necessarily imply however that the parameter sequence $\{\theta_n, n \geq 0\}$ converges to some fixed value [see, e.g., Roberts and Rosenthal (2007)].

THEOREM 2.1.   *Assume* A1 *and* A2. *Let f be a bounded function such that* $\lim_n \pi_{\theta_n}(f) = \alpha$ $\mathbb{P}$-*a.s. for some constant α. Then*

$$\lim_{n \to \infty} \mathbb{E}[f(X_n)] = \alpha.$$

The proof is in Section 4.1. As a trivial corollary, we have:

COROLLARY 2.2.   *Assume* A1 *and* A2. *Assume* $\{\pi_{\theta_n}, n \geq 0\}$ *converges weakly to π $\mathbb{P}$-a.s. Then,* $\lim_{n \to \infty} \mathbb{E}[f(X_n)] = \pi(f)$ *for any bounded continuous function f.*

When $\pi_\theta = \pi$ for any $\theta \in \Theta$, Theorem 2.1 improves the results of Roberts and Rosenthal (2007) by weakening the conditions on the transition kernels $\{P_\theta, \theta \in \Theta\}$ (the containment condition is not assumed to hold). The following example shows that ergodicity can be achieved even if the containment condition in Roberts and Rosenthal (2007) fails, provided that the adaptation rate is slow enough.

EXAMPLE 1 (Toy example).   Let us consider the following example introduced in Andrieu and Moulines (2006) and thoroughly analyzed in Andrieu and Thoms [(2008), Section 2] and Bai, Roberts and Rosenthal (2011). Let $\{\theta_n, n \geq 0\}$ be a $[0, 1]$-valued deterministic sequence. Consider the nonhomogeneous Markov chain over $\mathsf{X} = \{0, 1\}$ with transition matrix

(4) $$P_\theta = \begin{bmatrix} \theta & 1 - \theta \\ 1 - \theta & \theta \end{bmatrix}, \qquad \theta \in [0, 1].$$

For any $\theta \in [0, 1]$, $\pi = [1/2, 1/2]$ is a stationary distribution; the chain is irreducible if $\theta \in (0, 1)$. In this case, for $\varepsilon > 0$ and $\theta \in (0, 1)$,

$$M_\varepsilon(x, \theta) = \ln(\varepsilon)/\ln|1 - 2\theta|.$$

Assume that, for $n \geq 1$, $\theta_n = n^{-1/4}$. Clearly, for any $\varepsilon > 0$, $\{M_\varepsilon(X_n, \theta_n), n \geq 0\}$ grows to infinity with probability 1 and the containment condition does not hold [see also Bai, Roberts and Rosenthal (2011), Proposition 1].

Setting $r(n) = n^{1/3}$

$$\limsup_{n \to \infty} \mathbb{E} \| P_{\theta_{n-r(n)}}^{r(n)} (X_{n-r(n)}, \cdot) - \pi \|_{\mathrm{TV}} = \limsup_{n \to \infty} |2\theta_n - 1|^{r(n)} = 0$$

shows that A2(a) holds. Furthermore, we have

$$D(\theta, \theta') = \sup_{x \in \{0,1\}} \| P_\theta(x, \cdot) - P_{\theta'}(x, \cdot) \|_{\mathrm{TV}} = 2|\theta - \theta'|.$$

Therefore, with $\theta_n = n^{-1/4}$, $D(\theta_n, \theta_{n-1}) = O(n^{-1})$, and A2(b) is satisfied with $r(n) = n^{1/3}$. Corollary 2.2 therefore applies, and the marginal distribution converges.

To check A2(a), it is often easier to use drift conditions. To simplify the discussion below, this paper only covers the case of drift inequalities for geometric ergodicity. Extensions to subgeometric rates of convergence can be obtained following the same lines [see, e.g., Bai, Roberts and Rosenthal (2011) and Atchadé and Fort (2010)] and are left to future work. In the geometric setting, one commonly assumes the following simultaneous geometric drift and minorization conditions:

A3 For all $\theta \in \Theta$, P_θ is π-irreducible, aperiodic and there exist a function $V : \mathsf{X} \to [1, +\infty)$, and for any $\theta \in \Theta$ there exist some constants $b_\theta < \infty$, $\delta_\theta \in (0, 1)$, $\lambda_\theta \in (0, 1)$ and a probability measure ν_θ on X such that

$$P_\theta V \le \lambda_\theta V + b_\theta,$$

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

A3 implies geometric ergodicity [see, e.g., Meyn and Tweedie (2009), Chapter 15]. The following proposition can be obtained from Roberts and Rosenthal (2004), Fort and Moulines (2003), Douc, Moulines and Rosenthal [(2004), Proposition 3] or Baxendale (2005) [see also the proof of Lemma 3 in Saksman and Vihola (2010) for a similar result].

LEMMA 2.3.   *Assume* A3. *Then for any* θ, *there exists a probability distribution* π_θ *such that* $\pi_\theta P_\theta = \pi_\theta$, $\pi_\theta(V) \le b_\theta(1 - \lambda_\theta)^{-1}$ *and*

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

for some finite constants C_θ *and* $\rho_\theta \in (0, 1)$. *Furthermore, there exist positive constants* C *and* γ *such that for any* $\theta \in \Theta$,

$$(5) \qquad L_\theta \stackrel{\text{def}}{=} C_\theta \vee (1 - \rho_\theta)^{-1} \le C\{b_\theta \vee \delta_\theta^{-1} \vee (1 - \lambda_\theta)^{-1}\}^\gamma.$$

EXAMPLE 2 [The adaptive Metropolis (AM) algorithm].   We establish the ergodicity of the AM algorithm. 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 $|\theta|^2 \stackrel{\text{def}}{=} |\mu|^2 + \text{Tr}(\Gamma^T \Gamma)$, where $\theta = (\mu, \Gamma)$.

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

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

(6)

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

$$(7) \qquad \mu_{n+1} = \mu_n + \frac{1}{n+1}(X_{n+1} - \mu_n), \qquad \mu_0 = 0,$$

$$(8) \qquad \Gamma_{n+1} = \frac{n}{n+1}\Gamma_n + \frac{1}{n+1}\{(X_{n+1} - \mu_n)(X_{n+1} - \mu_n)^T + \kappa I_d\},$$

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 A1 holds with $\pi_\theta = \pi$ for any θ. As in Saksman and Vihola (2010), 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 [(2010), Proposition 15] establishes a drift inequality and a minorization condition on the kernel as in A3, with a drift function $V \propto \pi^{-s}$ with $s = 1/2$. Nevertheless, the generalization to an arbitrary $s \in (0, 1)$ is straightforward. Note that the function

$$(9) \qquad W(x) \overset{\text{def}}{=} \pi^{-s}(x)\|\pi^s\|_\infty$$

grows faster than an exponential under M1 [see, e.g., Saksman and Vihola (2010), Lemma 8]. Hence, Lemma 2.3 and Proposition 15 of Saksman and Vihola (2010) both imply:

LEMMA 2.4. *Assume* M1. *For any* $a \in (0, 1]$ *and* $\theta \in \Theta$, *there exist* $C_{a,\theta} < \infty$ *and* $\rho_{a,\theta} \in (0, 1)$, *such that*

$$\|P_\theta^k(x, \cdot) - \pi\|_{W^a} \leq C_{a,\theta}\rho_{a,\theta}^k W^a(x) \qquad \text{for any } x \in \mathbb{R}^d,$$

where W *is defined by* (9). *In addition, there exist finite constants* c_a, b_a *such that*

$$C_{a,\theta} \vee (1 - \rho_{a,\theta})^{-1} \leq c_a|\theta|^{d\gamma/2} + b_a,$$

where the constant γ *is defined in Lemma* 2.3.

In Saksman and Vihola [(2010), Lemma 12] it is proved that under M1, the rate of growth of the parameters $\{\theta_n, n \geq 0\}$ is controlled. Namely, for any $\tau > 0$,

$$(10) \qquad \sup_{n \geq 1} n^{-\tau}|\theta_n| < +\infty, \qquad \mathbb{P}\text{-a.s.}$$

In the following lemma, we establish a control of the rate of growth of the state of the chain $\{X_n, n \geq 0\}$.

LEMMA 2.5. *Assume* M1. *Then*:

(i) $\mathbb{E}[W(X_n)] \leq \mathbb{E}[W(X_0)] + nb$.

(ii) *For any $t > 0$ and any $\tau > 0$, there exists a constant $C_{t,\tau}$ such that for any $n \geq 0$,*

$$\mathbb{E}\big[W(X_n)\mathbb{1}_{\sup_{k \leq n-1} k^{-\tau}|\theta_k| \leq t}\big] \leq \mathbb{E}[W(X_0)] + C_{t,\tau}n^{\tau d\gamma/2},$$

where γ is defined in Lemma 2.3.

(iii) *If $\mathbb{E}[W(X_0)] < +\infty$, for any $\tau > 0$, $\sup_{n \geq 1} n^{-1-\tau}W(X_n) < +\infty$, $\mathbb{P}$-a.s.*

The proof of this lemma is given in Section 4.2. By combining Lemma 2.4 and Lemma 2.5, we prove A2(a): as a consequence of Lemma 2.4, it holds for any $\tau > 0$ such that $r > \tau d\gamma/2$ and for any $t > 0$

$$(11) \qquad \limsup_{n \to \infty} \sup_{\theta \in \Theta, |\theta| \leq tn^\tau} \sup_{x \in \mathbb{R}^d, W(x) \leq tn^{1+\tau}} \big\| P_\theta^{\lfloor n^r \rfloor}(x, \cdot) - \pi \big\|_{\mathrm{TV}} = 0,$$

where $\lfloor \cdot \rfloor$ denotes the lower integer part. For $t > 0$, set

$$\Omega_t \stackrel{\mathrm{def}}{=} \Big\{\omega : \sup_{n \geq 1} n^{-\tau}|\theta_n| \leq t, \sup_{n \geq 1} n^{-1-\tau}W(X_n) \leq t\Big\}.$$

Equation (10) and Lemma 2.5(iii) show that $\lim_{t \to \infty} \mathbb{P}(\Omega_t) = 1$. Set $r(n) = \lfloor n^r \rfloor$. The Fatou lemma and the monotone convergence theorem show that

$$\limsup_{n \to \infty} \mathbb{E}\big[\big\| P_{\theta_{n-r(n)}}^{r(n)}(X_{n-r(n)}, \cdot) - \pi \big\|_{\mathrm{TV}}\big]$$

$$\leq \mathbb{E}\Big[\limsup_{n \to \infty} \big\| P_{\theta_{n-r(n)}}^{r(n)}(X_{n-r(n)}, \cdot) - \pi \big\|_{\mathrm{TV}}\Big]$$

$$\leq \lim_{t \to \infty} \mathbb{E}\Big[\limsup_{n \to \infty} \big\| P_{\theta_{n-r(n)}}^{r(n)}(X_{n-r(n)}, \cdot) - \pi \big\|_{\mathrm{TV}}\mathbb{1}_{\Omega_t}\Big] = 0.$$

Therefore, A2(a) is satisfied whereas clearly the uniform containment condition [see (3)] seems to be very challenging to check.

Consider now A2(b). It is proved in Andrieu and Moulines [(2006), Lemma 13] that for any $(\theta, \tilde{\theta}) \in \Theta^2$ and $a \in [0, 1]$, $D_{W^a}(\theta, \tilde{\theta}) \leq 2d\kappa^{-1}|\Gamma - \tilde{\Gamma}|$. By definition of Γ_n [see (8)], we have for any $m < n$,

$$D_{W^a}(\theta_n, \theta_{n-m}) \leq \frac{2d\kappa^{-1}}{n}\Bigg(2\kappa md + \frac{m}{n-m}\sum_{j=0}^{n-m-1}|X_{j+1} - \mu_j|^2$$

$$+ \sum_{j=n-m}^{n-1}|X_{j+1} - \mu_j|^2\Bigg).$$

By definition of the empirical mean μ_k [see (7)] there exists a constant C' such that $|\mu_k| \leq C'\{k^{-1}\sum_{j=1}^{k}|X_j|^2\}^{1/2}$; under M1, $\liminf_{|x|\to\infty}\ln W(x)/|x| > 0$ [see the proof of Lemma 8 in Saksman and Vihola (2010)]. Therefore, there exists a constant C such that

(12)
$$D_{W^a}(\theta_n, \theta_{n-m})$$
$$\leq C\frac{m}{n}\left\{1 + \frac{(1+\ln(n-m))}{n-m}\sum_{j=1}^{n-m}\ln^2 W(X_j)\right.$$
$$\left. + \frac{(1+\ln(n))}{m}\sum_{j=n-m}^{n}\ln^2 W(X_j)\right\}.$$

The proof of A2(b) now relies on the control of moments for the r.v. $\{\ln^2 W(X_j), j \geq 0\}$. Lemma 2.5(i) and Jensen's inequality show that the moment $\mathbb{E}[\ln^2 W(X_n)]$ increases at most as $\ln^2 n$. Then there exists a constant C such that for any $m \leq n$ and for any $a \in [0,1]$,

$$\mathbb{E}[D_{W^a}(\theta_n, \theta_{n-m})] \leq Cm\frac{\ln^3(n)}{n}\mathbb{E}[W(X_0)].$$

Then, for any $r \in (0, 1/2)$, $\lim_{n\to+\infty}\sum_{j=0}^{\lfloor n^r\rfloor-1}\mathbb{E}[D(\theta_{n-\lfloor n^r\rfloor+j}, \theta_{n-\lfloor n^r\rfloor})] = 0$ and A2(b) holds. Combining the results above yields:

THEOREM 2.6. *Assume* M1 *and* $\mathbb{E}[W(X_0)] < +\infty$. *Then, for any bounded function* f, $\lim_{n\to\infty}\mathbb{E}[f(X_n)] = \pi(f)$.

2.2. *Strong law of large numbers for additive functionals.* In this section, a strong law of large numbers (SLLN) is established. The main result of this section is Theorem 2.7 which provides a SLLN for a special class of additive functionals. To that goal, A3 is assumed to hold (which implies A1, see Lemma 2.3), and it is required to strengthen the diminishing adaptation and the stability conditions.

A4 $\sum_{k=1}^{\infty}k^{-1}(L_{\theta_k}\vee L_{\theta_{k-1}})^6 D_V(\theta_k, \theta_{k-1})V(X_k) < +\infty$ $\mathbb{P}$-a.s., where D_V and L_θ are defined in (2) and (5).

A5 (a) $\limsup_n \pi_{\theta_n}(V) < +\infty$, $\mathbb{P}$-a.s.
   (b) For some $\alpha > 1$, $\sum_{k=0}^{\infty}(k+1)^{-\alpha}L_{\theta_k}^{2\alpha}P_{\theta_k}V^\alpha(X_k) < +\infty$, $\mathbb{P}$-a.s.

Here again, these conditions balance the rate at which the transition kernel P_θ converges to stationarity and the adaptation speed. This is reflected in the condition A4: $(L_{\theta_k}\vee L_{\theta_{k-1}})$ is related to the rate of convergence of the kernels P_{θ_k} and $P_{\theta_{k-1}}$ to stationarity and $D_V(\theta_k, \theta_{k-1})$ reflects the adaptation speed.

THEOREM 2.7. *Assume* A3, A4 *and* A5. *Let* $F : \mathsf{X} \times \Theta \to \mathbb{R}$ *be a measurable function such that*:

(i) $\sup_\theta \|F(\cdot, \theta)\|_V < +\infty$,

(ii) $\sum_{k=1}^\infty k^{-1} L_{\theta_{k-1}}^2 \|F(\cdot, \theta_k) - F(\cdot, \theta_{k-1})\|_V V(X_k) < +\infty$ $\mathbb{P}$-a.s.,

(iii) $\lim_{n\to\infty} \int \pi_{\theta_n}(dx) F(x, \theta_n)$ exists $\mathbb{P}$-a.s.

Then,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} F(X_k, \theta_k) = \lim_{n\to\infty} \int \pi_{\theta_n}(dx) F(x, \theta_n), \qquad \mathbb{P}\text{-a.s.}$$

The proof is in Section 4.3. When the function F does not depend upon θ, this theorem becomes the following.

COROLLARY 2.8. *Assume* A3, A4 *and* A5. *Let* $f: \mathsf{X} \to \mathbb{R}$ *be a measurable function such that* $\|f\|_V < +\infty$ *and* $\lim_{n\to\infty} \pi_{\theta_n}(f)$ *exists* $\mathbb{P}$-a.s. *Then,* $n^{-1} \sum_{k=0}^{n-1} f(X_k) \xrightarrow{a.s.} \lim_n \pi_{\theta_n}(f)$.

EXAMPLE 3 (Toy example: law of large numbers). For $\theta \in (0, 1)$, the constants C_θ and ρ_θ (see Lemma 2.3) are, respectively, equal to 1 and $|1 - 2\theta|$ and $V = 1$. This implies that $L_\theta = 1/(2\theta)$ if $\theta \leq 1/2$ and $1/(2(1 - \theta))$ otherwise. Therefore A3 is satisfied since $\sum_{k=1}^\infty k^{-1} \theta_k^{-3} |\theta_{k-1} - \theta_k| < +\infty$ when $\theta_k = k^{-1/4}$. Assumption A4(a) is automatically satisfied because the stationary distribution does not depend on θ. Assumption A4(b) is satisfied for any $\alpha > 4/3$ because in such case $\sum_{k=1}^\infty (k^{-1}\theta_k)^\alpha < \infty$. By Theorem 2.7, the SLLN is satisfied for this nonhomogeneous Markov chain.

The stated assumptions are very general and, when applied to some specific settings, can be simplified. For example, in many interesting examples (see, e.g., Section 3), it is known that $\limsup_{n\to\infty} L_{\theta_n} < \infty$, $\mathbb{P}$-a.s. and for some $\alpha > 1$, $\sup_{n\geq 0} \mathbb{E}[V^\alpha(X_n)] < \infty$. Under these assumptions, it is straightforward to establish the following corollary:

COROLLARY 2.9. *Assume* A3 *and*:

(i) $\limsup_{n\to\infty} L_{\theta_n} < \infty$ *and* $\limsup_{n\to\infty} \pi_{\theta_n}(V) < +\infty$, $\mathbb{P}$-a.s.,

(ii) *there exists* $\alpha > 1$ *such that* $\sup_{k\geq 0} \mathbb{E}[V^\alpha(X_k)] < +\infty$,

(iii) $\sum_{k=1}^\infty k^{-1} D_V(\theta_k, \theta_{k-1}) V(X_k) < +\infty$ $\mathbb{P}$-a.s.

Let $f: \mathsf{X} \to \mathbb{R}$ *be a measurable function such that* $\|f\|_V < +\infty$ *and* $\lim_{n\to\infty} \pi_{\theta_n}(f)$ *exists* $\mathbb{P}$-a.s. *Then,* $n^{-1} \sum_{k=0}^{n-1} f(X_k) \xrightarrow{a.s.} \lim_{n\to\infty} \pi_{\theta_n}(f)$.

EXAMPLE 4 (AM: law of large numbers). Application of the above criteria yields the SLLN for the AM algorithm. This result has recently been obtained by Saksman and Vihola (2010).

Let $a \in (0, 1)$ and set $W(x) \stackrel{\text{def}}{=} \pi^{-s}(x)\|\pi^s\|_\infty$ for $s \in (0, 1)$. We prove that a (strong) LLN holds for any function f in L_{W^a}. We choose $\tau > 0$ small enough so that

$$(13) \qquad (1 - a) > \tau(a + 3d\gamma), \qquad 1/a - 1 > \tau d\gamma(1/a + 1/2),$$

where γ is given by Lemma 2.3. Consider A4. By Lemma 2.4 and (10), there exists a r.v. U_1, $\mathbb{P}$-a.s. finite such that $L_{\theta_k} \vee L_{\theta_{k-1}} \leq U_1 k^{\tau d\gamma/2}$. By (12) and Lemma 2.5(iii), there exists a r.v. U_2, $\mathbb{P}$-a.s. finite such that $D_{W^a}(\theta_k, \theta_{k-1}) \leq U_2 k^{-1} \ln^3 k$. Finally, applying Lemma 2.5(iii) again, there exists a r.v. U_3, $\mathbb{P}$-a.s. finite such that $W^a(X_k) \leq U_3 k^{a(1+\tau)}$. Combining these inequalities show that there exists a r.v. U, $\mathbb{P}$-a.s. finite such that

$$\sum_k k^{-1}(L_{\theta_k} \vee L_{\theta_{k-1}})^6 D_{W^a}(\theta_k, \theta_{k-1})W^a(X_k) \leq U \sum_k k^{2-a-\tau(a+3d\gamma)} \ln^3 k,$$

thus showing A4 [observe that the RHS is finite by definition of τ, equation (13)]. The proof of A5(b) could rely on the same inequalities in the case $a \in (0, 1/2)$. Nevertheless, a SLLN can be established for larger values of a by using the bound on $W(X_n)$ given by Lemma 2.5(ii) which improves on Lemma 2.5(iii). Set $\Omega_t \stackrel{\text{def}}{=} \{\sup_{n \geq 1} n^{-\tau}|\theta_n| \leq t\}$. By Lemma 2.5, $\lim_{t \to +\infty} \mathbb{P}(\Omega_t) \uparrow 1$ and A5(b) holds provided $\sum_{k \geq 1} k^{-1/a} L_{\theta_{k-1}}^{2/a} P_{\theta_{k-1}} W(X_k)\mathbb{1}_{\Omega_t}$ is finite $\mathbb{P}$-a.s. for any $t > 0$. Lemmas 2.4 and 2.5(ii) imply that there exists a constant C_t such that

$$\mathbb{E}\left[\sum_k k^{-1/a} L_{\theta_{k-1}}^{2/a} P_{\theta_{k-1}} W(X_k)\mathbb{1}_{\Omega_t}\right] \leq C_t \sum_k k^{-1/a+\tau d\gamma(1/a+1/2)}.$$

The RHS is finite by definition of τ [see (13)].

The above discussion is summarized in the following theorem.

THEOREM 2.10. *Assume* M1 *and* $\mathbb{E}[W(X_0)] < +\infty$. *Then, for any* $a \in (0, 1)$ *and any function* $f \in \mathcal{L}_{W^a}$, $n^{-1}\sum_{k=1}^n f(X_k) \stackrel{a.s.}{\longrightarrow} \pi(f)$.

2.3. *Almost sure convergence of the invariant distributions.* When the stationary distribution π_θ is not explicitly known, convergence of the sequence $\{\pi_{\theta_n}, n \geq 0\}$ has to be obtained from the convergence of the transition kernels $\{P_{\theta_n}, n \geq 0\}$. We propose below a set of sufficient conditions allowing to prove the almost sure convergence of $\{\pi_{\theta_n}(f), n \geq 0\}$ for continuous functions f. The proof of Theorem 2.11 is in Section 4.4.

THEOREM 2.11. *Assume that* X *is a Polish space. Assume* A3 *and*:

(i) $\limsup_{n \to \infty} L_{\theta_n} < \infty$ $\mathbb{P}$-*a.s. where* L_θ *is given by* (5),
(ii) *for any function* f *in* $\mathsf{C}_b(\mathsf{X})$, *the class of functions* $\{P_\theta f, \theta \in \Theta\}$ *is equicontinuous*,

(iii) *there exists $\theta_\star \in \Theta$ and for any $x \in \mathsf{X}$, a $\mathbb{P}$-full set Ω_x such that for any $\omega \in \Omega_x$, $\{P_{\theta_n(\omega)}(x, \cdot), n \geq 0\}$ converges weakly to $P_{\theta_\star}(x, \cdot)$.*

Then, there exists a $\mathbb{P}$-full set Ω_0 such that, for any any $\omega \in \Omega_0$ and $f \in \mathsf{C}_b(\mathsf{X})$, $\pi_{\theta_n(\omega)}(f) \xrightarrow{a.s.} \pi_{\theta_\star}(f)$ (or, equivalently, for any $\omega \in \Omega_0$, $\pi_{\theta_n(\omega)}$ converges weakly to $\pi_{\theta_\star}$).

Note that the weak convergence implies that for any $\omega \in \Omega_0$ and for any set A such that $\pi_{\theta_\star}(\partial A) = 0$ where ∂A denotes the boundary of A, $\lim_n \pi_{\theta_n(\omega)}(A) = \pi_{\theta_\star}(A)$.

Theorem 2.11 might be seen as an extension of the classical results on the continuity of the perturbations of the spectrum and eigenprojections; but it is stated under assumptions that are weaker than what is usually assumed [Kato (1980), Theorem 3.16]. The difference stems from the fact that condition (iii) does not imply the convergence of P_θ to $P_{\theta_\star}$ in operator norm. This is crucial to deal with the interacting tempering algorithm (see Section 3).

Condition (iii) of Theorem 2.11 is certainly the most difficult to check. In the case, it is known that for any function $f \in \mathsf{C}_b(\mathsf{X})$, there exists a $\mathbb{P}$-full set $\Omega_{x,f}$ such that for any $\omega \in \Omega_{x,f}$, $\lim_n P_{\theta_n(\omega)} f(x) = P_{\theta_\star} f(x)$, then the existence of a $\mathbb{P}$-full set, uniform in f for $f \in \mathsf{C}_b(\mathsf{X})$, relies on the characterization of the weak convergence by a separable class of functions [see Dudley (2002), Theorem 11.4.1, and Proposition 3.3 below for an example].

3. Convergence of the interacting tempering (IT) algorithm.
We consider the interacting tempering algorithm, which is a simplified form of the equi-energy sampler by Kou, Zhou and Wong (2006).

Assume that X is a Polish space equipped with its Borel σ-field $\mathcal{X}$. Let π be the target density w.r.t. a measure μ on $(\mathsf{X}, \mathcal{X})$. Denote by K the number of different temperature levels, $T_1 = 1 < T_2 < \cdots < T_K$. For $k \in \{1, \ldots, K-1\}$, let $P^{(k)}$ be a transition kernel on $(\mathsf{X}, \mathcal{X})$ with unique invariant distribution π^{1/T_k}. Fix $\upsilon \in (0, 1)$ the probability of interaction.

We denote by $X^{(k)} = (X_n^{(k)})_n$ the sampled values at each temperatures T_k. The chains are defined by induction on k: given the past of the process $X^{(k+1)}$ up to time n, and the current value $X_n^{(k)}$ of the current process $X^{(k)}$, we define $X_{n+1}^{(k)}$ as follows:

1. with probability $(1 - \upsilon)$, the state $X_{n+1}^{(k)}$ is sampled using the Markov kernel $P^{(k)}(X_n^{(k)}, \cdot)$,
2. with probability υ, a tentative state Z_{n+1} is drawn at random from the past $\{X_\ell^{(k+1)}, \ell \leq n\}$. This move is accepted with probability $1 \wedge (\pi(X_n^{(k)})/\pi(Z_{n+1}))^{T_{k+1}^{-1} - T_k^{-1}}$.

We consider first the case $K = 2$. We will then address the general case (see Theorem 3.6 below). For notational simplicity, we set $T_2 = T > 1$ and $P^{(1)} = P$. Denote by Θ the set of the probability measures on $(\mathsf{X}, \mathcal{X})$. For any distribution $\theta \in \Theta$, define the transition kernel $P_\theta(x, \cdot) \stackrel{\text{def}}{=} (1 - \upsilon) P(x, \cdot) + \upsilon K_\theta(x, \cdot)$, where, for any $A \in \mathcal{X}$,

$$(14) \qquad K_\theta(x, A) \stackrel{\text{def}}{=} \int_A \alpha(x, y)\theta(\mathrm{d}y) + \mathbb{1}_A(x) \int \{1 - \alpha(x, y)\}\theta(\mathrm{d}y)$$

with

$$(15) \quad \alpha(x, y) = 1 \wedge \frac{\pi(y)\pi^{1/T}(x)}{\pi(x)\pi(y)^{1/T}} = 1 \wedge \frac{\pi^\beta(y)}{\pi^\beta(x)}, \qquad \beta \stackrel{\text{def}}{=} 1 - \frac{1}{T} \in (0, 1).$$

Denote by $\{Y_n, n \geq 0\}$ the process run at the temperature T. It is not assumed that $\{Y_n, n \geq 0\}$ is a Markov chain. We simply assume that, for any bounded continuous function f, $n^{-1} \sum_{k=1}^n f(Y_k) \to \theta_\star(f)$ a.s. where $\theta_\star$ is the probability distribution on $(\mathsf{X}, \mathcal{X})$ with density (w.r.t. μ) proportional to $\pi^{1/T}$. We consider the process $\{X_n, n \geq 0\}$ defined, for each $n \geq 0$ and any bounded function $f : \mathsf{X} \to \mathbb{R}$,

$$\mathbb{E}[f(X_{n+1})|\mathcal{F}_n] = P_{\theta_n} f(X_n) \qquad \text{where } \theta_n(f) = (n + 1)^{-1} \sum_{k=0}^n f(Y_k).$$

Since, by construction, $\pi P_{\theta_\star} = \pi$, it is expected that the marginal distribution of X_k as k goes to infinity converges to π. To go further, some additional assumptions are required:

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

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

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

$$(16) \qquad PW \leq \lambda W + b \qquad \text{with } W(x) \stackrel{\text{def}}{=} (\pi(x)/\|\pi\|_\infty)^{-\tau}.$$

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

When $\mathsf{X} \subseteq \mathbb{R}^d$ and P is a symmetric random-walk Metropolis (SRWM) algorithm then $\pi P = \pi$ and P is π-irreducible [Mengersen and Tweedie (1996), Lemma 1.1]. If in addition the proposal density is continuous on X then, since π is positive and continuous on X, any compact set of X is 1-small [Mengersen and Tweedie (1996), Lemma 1.2]. Therefore, the transition kernel of a SRWM algorithm satisfies I2(a) and I2(c).

Drift conditions of the form I2(b) for the SRWM algorithm on $\mathsf{X} \subseteq \mathbb{R}^d$ are discussed in Roberts and Tweedie (1996), Jarner and Hansen (2000) and Saksman and Vihola (2010). Under conditions which imply that the target density π is super-exponential and have regular contours (see M1), Jarner and Hansen (2000)

and Saksman and Vihola (2010) show that any functions proportional to π^{-s} with $s \in (0, 1)$ satisfies a Foster–Lyapunov drift inequality [Jarner and Hansen (2000), Theorems 4.1 and 4.3]. Under this condition, I2(b) is satisfied with any τ in the interval $(0, 1/T)$.

Stability conditions on the auxiliary process $\{Y_n, n \geq 0\}$ are also required.

I3 (a) $\theta_\star(W) < +\infty$ and for any continuous function f in $\mathcal{L}_W$, $\theta_n(f) \xrightarrow{\text{a.s.}} \theta_\star(f)$.
   (b) $\sup_n \mathbb{E}[W(Y_n)] < +\infty$.

The following proposition is the key-ingredient to prove the convergence of the IT sampler. Under the stated assumptions, we prove that the transition kernels $\{P_\theta, \theta \in \Theta\}$ satisfy a Foster–Lyapunov drift inequality and a minorization condition. The proof of Proposition 3.1 is adapted from Atchadé [(2010), Lemma 4.1]; a detailed proof is given in Fort, Moulines and Priouret (2011), Section 2.

PROPOSITION 3.1. *Assume* I1 *and* I2. *Then, there exist* $\tilde{\lambda} \in (0, 1)$, $\tilde{b} < \infty$, *such that, for any* $\theta \in \Theta$,

$$(17) \qquad P_\theta W(x) \leq \tilde{\lambda} W(x) + \tilde{b}\theta(W).$$

In addition, for any $p \in (0, \|\pi\|_\infty)$, *the level sets* $\{\pi \geq p\}$ *are* 1-*small w.r.t. the transition kernels* P_θ *and the minorization constant does not depend upon* θ.

COROLLARY 3.2. *Assume* I1, I2, I3 *and* $\mathbb{E}[W(X_0)] < +\infty$. *Then*:

(i) $\sup_{n \geq 0} \mathbb{E}[W(X_n)] < +\infty$,
(ii) $\limsup_{n \to \infty} L_{\theta_n} < +\infty$ $\mathbb{P}$-*a.s., where* L_θ *is defined by* (5).

The proof of Corollary 3.2 is in Section 5.1. As a consequence of Proposition 3.1, the transition kernel P_θ possesses an (unique) invariant distribution π_θ. Ergodicity and SLLN for additive functionals both require the a.s. convergence of $\pi_{\theta_n}(f)$ (see Theorems 2.1 and 2.7). Nevertheless, in this example, π_θ does not have an explicit expression. The proof of the following proposition is postponed in Section 5.2.

PROPOSITION 3.3. *Assume* I1, I2, I3 *and* $\mathbb{E}[W(X_0)] < +\infty$. *Then, the conditions of Theorem* 2.11 *hold and for any bounded continuous function* f, $\lim_n \pi_{\theta_n}(f) = \pi(f)$ $\mathbb{P}$-*a.s.*

We now address the convergence of the marginals.

THEOREM 3.4. *Assume* I1, I2, I3 *and* $\mathbb{E}[W(X_0)] < +\infty$. *Then, for any bounded continuous function* f, $\lim_n \mathbb{E}[f(X_n)] = \pi(f)$.

PROOF. We check the assumptions of Corollary 2.2. By Corollary 3.2(i), $\{W(X_n), n \geq 0\}$ is bounded in probability. Furthermore, Corollary 3.2(ii) implies that $\limsup_n C_{\theta_n} < +\infty$ $\mathbb{P}$-a.s. and $\limsup_n \rho_{\theta_n} < 1$ $\mathbb{P}$-a.s. This proves A2(a).

The next step is to establish A2(b). Since, for any bounded function f, $\theta_{n+m}(f) = (n+m+1)^{-1} \sum_{k=n+1}^{n+m} f(Y_k) + (n+1)(n+m+1)^{-1}\theta_n(f)$, we have

$$|P_{\theta_{n+m}} f(x) - P_{\theta_n} f(x)| \leq \sup_{y,z \in \mathsf{X}} |f(y) - f(z)| \|\theta_{n+m} - \theta_n\|_{\text{TV}} \leq \frac{2\|f\|_\infty m}{n+m+1}.$$

Consequently, $D(\theta_{n+m}, \theta_n)$ is deterministically bounded by a sequence converging to zero. We have

$$\sum_{j=0}^{r_\varepsilon(n)-1} \mathbb{E}\big[D\big(\theta_{n-r_\varepsilon(n)+j}, \theta_{n-r_\varepsilon(n)}\big)\big] \leq 2 \frac{r_\varepsilon^2(n)}{n - r_\varepsilon(n)}$$

thus proving A2(b) with any sequence of the form $r_\varepsilon(n) = n^r$ with $r < 1/2$.

Finally, Proposition 3.3 proves the convergence of $\pi_{\theta_n}(f)$ for any bounded continuous function f. $\square$

We now state the strong law of large numbers for the IT sampler.

THEOREM 3.5. *Assume* I1, I2, I3 *and* $\mathbb{E}[W(X_0)] < +\infty$. *Then*:

(i) *for any measurable set A such that $\int_{\partial A} \pi \, d\mu = 0$ where ∂A is the boundary of A,*

$$\frac{1}{n} \sum_{k=0}^{n-1} \mathbb{1}_A(X_k) \xrightarrow{a.s.} \int_A \pi \, d\mu;$$

(ii) *for any $a \in (0, 1)$ and any continuous function f in L_{W^a},*

$$\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) \xrightarrow{a.s.} \int f\pi \, d\mu.$$

PROOF. We check conditions (i), (ii) and (iii) of Corollary 2.9 with $V \stackrel{\text{def}}{=} W^a$ for $a \in (0, 1)$, and $\alpha \stackrel{\text{def}}{=} 1/a$. Assumption A3 holds and $\limsup_n L_{\theta_n} < +\infty$ $\mathbb{P}$-a.s. [see Proposition 3.1 and Corollary 3.2(ii)]. The drift condition (17) implies that

$$(18) \qquad \limsup_n \pi_{\theta_n}(W) \leq \frac{\tilde{b}}{1 - \tilde{\lambda}} \limsup_n \theta_n(W).$$

Since W is continuous, the assumption I3(a) implies that $\limsup_n \theta_n(W) < \infty$ $\mathbb{P}$-a.s. Hence, condition (i) of Corollary 2.9 holds. Corollary 3.2(i) implies the condition (ii) of Corollary 2.9. The definition (2) of D_V implies

$$D_V(\theta_k, \theta_{k-1}) \leq 2\upsilon\|\theta_k - \theta_{k-1}\|_V \leq \frac{2}{k+1}\theta_{k-1}(V) + \frac{2}{k+1}V(Y_k).$$

Hence, under I3(a), condition (iii) of Corollary 2.9 holds if $\sum_k k^{-2} V(X_k) < +\infty$ and $\sum_k k^{-2} V(X_k) V(Y_k) < +\infty$ $\mathbb{P}$-a.s. The first series converges since, by Corollary 3.2(i), $\sup_k \mathbb{E}[V(X_k)] < +\infty$. For the second series, it is sufficient to prove that $\sum_k k^{-2/p} V^{1/p}(X_k) V^{1/p}(Y_k) < +\infty$ w.p.1 with $p \stackrel{\text{def}}{=} (2a) \vee 1$. We have by the Cauchy–Schwarz inequality

$$\begin{aligned}
\mathbb{E}[V^{1/p}(Y_k) V^{1/p}(X_k)] &\leq \mathbb{E}[V^{2/p}(Y_k)]^{1/2} \mathbb{E}[V^{2/p}(X_k)]^{1/2} \\
&\leq \mathbb{E}[V^{1/a}(Y_k)]^{1/2} \mathbb{E}[V^{1/a}(X_k)]^{1/2} \\
&= \mathbb{E}[W(Y_k)]^{1/2} \mathbb{E}[W(X_k)]^{1/2}.
\end{aligned}$$

The RHS is finite under I3(b) and Corollary 3.2(i). Then, this concludes the proof of condition (iii) of Corollary 2.9.

It remains to prove that $\lim_n \pi_{\theta_n}(f) = \pi(f)$ $\mathbb{P}$-a.s. By Proposition 3.3, this property holds for any bounded continuous function f and any set A such that $\int_{\partial A} \pi \, d\mu = 0$. We proved that there exists $\alpha > 1$ such that $\limsup_n \pi_{\theta_n}(V^\alpha) + \pi(V^\alpha) < +\infty$ [see (18)]. Classical truncation arguments imply that $\lim_n \pi_{\theta_n}(f)$ exists $\mathbb{P}$-a.s. for any continuous function $f \in \mathcal{L}_V$ [see, e.g., Billingsley (1999), Theorem 3.5, or similar arguments in the proof of Proposition 4.3]. $\quad\square$

To summarize the above discussions, the process $\{X_n, n \geq 0\}$ has uniformly bounded W-moments (see Corollary 3.2), the distribution of X_n converges to π as $n \to +\infty$ (Theorem 3.4) and a strong law of large numbers is satisfied for a wide family of functions (Theorem 3.5). The results are obtained provided the auxiliary process also possesses uniformly bounded W-moments and satisfies a strong law of large numbers (see I3). Repeated applications of this result provides sufficient conditions for the interacting tempering with multiple stages to be ergodic and to satisfy a strong law of large numbers. Recall that IT algorithm defines recursively K random sequences $X^{(i)} = \{X_n^i, n \geq 0\}$ for $i \in \{1, \ldots, K\}$ such that $X^{(i)}$ targets the distribution proportional to π^{1/T_i}. We are interested in $X^{(1)}$ which targets $\pi^{1/T_1} = \pi$. The proof of Theorem 3.6 is in Section 5.3.

THEOREM 3.6. *Let* $(\mathsf{X}, \mathcal{X})$ *be a Polish space, and* π *be a density (w.r.t. a measure* μ*) satisfying* I1. *Choose* $T_\star > 1$ *and* $T_1 = 1 < T_2 < \cdots < T_K < T_\star$. *Assume that for any* $i \in \{1, \ldots, K-1\}$, *there exists a* π-*irreducible Feller transition kernel* $P^{(i)}$ *on* $(\mathsf{X}, \mathcal{X})$ *such that*:

(i) $\pi^{1/T_i} P^{(i)} = \pi^{1/T_i}$,

(ii) *for any* $s \in (0, 1/T_i)$, *there exist* $\lambda^{(i)} \in (0, 1)$ *and* $b^{(i)} < +\infty$ *such that* $P^{(i)} U_s \leq \lambda^{(i)} U_s + b^{(i)}$ *where* $U_s \propto \pi^{-s}$.
Assume in addition that there exists $\tilde{T} \in (T_K, T_\star)$ *such that*:

(iii) $\int \pi^{1/T_K - 1/\tilde{T}} \, d\mu < +\infty$,

(iv) *for any continuous function in $\mathcal{L}_{\pi^{-1/\tilde{T}}}$,*

$$n^{-1} \sum_{k=1}^{n} f(X_k^{(K)}) \overset{a.s.}{\longrightarrow} \int f \frac{\pi^{1/T_K}}{\int \pi^{1/T_K} \, d\mu} \, d\mu,$$

(v) $\sup_n \mathbb{E}[\pi^{-1/\tilde{T}}(X_n^{(K)})] < \infty$.

Finally, assume that for any $i \in \{1, \ldots, K-1\}$, $\mathbb{E}[\pi^{-1/\tilde{T}}(X_0^{(i)})] < +\infty$. Then, for any continuous function f in $\mathcal{L}_{\pi^{-1/T_\star}}$,

$$n^{-1} \sum_{k=1}^{n} f(X_k^{(1)}) \overset{a.s.}{\longrightarrow} \int f \pi \, d\mu.$$

Note that since convergence holds for any continuous function f in $\mathcal{L}_{\pi^{-1/T_\star}}$, it also holds with $f = \mathbb{1}_A$ where A is a measurable set such that $\int_{\partial A} \pi \, d\mu = 0$.

We conclude this section by an example of SRWM-based interacting tempering algorithm, for which the conditions of Theorem 3.6 hold. The proof is in Section 5.4.

PROPOSITION 3.7. *Let π be a super-exponential density on $\mathsf{X} = \mathbb{R}^d$ with regular contours (i.e., satisfying M1). Let $T_\star \in (1, +\infty)$ and choose a temperature ladder $1 = T_1 < \cdots < T_K < T_\star$. Consider the K-stages interacting tempering algorithm with:*

- *for $i \in \{1, \ldots, K-1\}$, $P^{(i)}$ is a SRWM transition kernel with invariant distribution proportional to π^{1/T_i} and proposal distribution $\mathcal{N}_d(0, \Sigma^{(i)})$,*
- *$\{X_n^{(K)}, n \geq 0\}$ is a SRWM Markov chain with invariant distribution proportional to π^{1/T_K} and proposal distribution $\mathcal{N}_d(0, \Sigma^{(K)})$.*

Finally, assume that for any $i \in \{1, \ldots, K\}$, $\mathbb{E}[\pi^{-1/T_\star}(X_0^{(i)})] < +\infty$. Then, for any continuous function $f \in \mathcal{L}_{\pi^{-1/T_\star}}$, $n^{-1} \sum_{k=1}^{n} f(X_k^{(1)}) \overset{a.s.}{\longrightarrow} \pi(f)$ as $n \to +\infty$.

4. Proofs of Section 2.

4.1. *Proof of Theorem 2.1.* We preface the proof by a lemma, which is proved in Atchadé et al. (2011), Proposition 1.7.1.

LEMMA 4.1. *For any integers $n, N > 0$,*

$$\sup_{\|f\|_\infty \leq 1} \left| \mathbb{E}[f(X_{n+N})|\mathcal{F}_n] - P_{\theta_n}^N f(X_n) \right| \leq \sum_{j=1}^{N-1} \mathbb{E}[D(\theta_{n+j}, \theta_n)|\mathcal{F}_n], \qquad \mathbb{P}\text{-}a.s.$$

PROOF OF THEOREM 2.1. Let f be a bounded nonnegative function. Without loss of generality, assume that $\|f\|_\infty \leq 1$. For any $N \leq n$,

$$(19) \quad |\mathbb{E}[f(X_n)] - \alpha| \leq |\mathbb{E}[f(X_n) - P^N_{\theta_{n-N}} f(X_{n-N})]| + |\mathbb{E}[\pi_{\theta_{n-N}}(f) - \alpha]| + |\mathbb{E}[P^N_{\theta_{n-N}} f(X_{n-N}) - \pi_{\theta_{n-N}}(f)]|.$$

Let $\varepsilon > 0$. By setting $N = r_\varepsilon(n)$ where the sequence $\{r_\varepsilon(n), n \geq 0\}$ is as in A2(a), the third term on the RHS in (19) is bounded by

$$\mathbb{E}\big[\big\| P^{r_\varepsilon(n)}_{\theta_{n-r_\varepsilon(n)}}(X_{n-r_\varepsilon(n)}, \cdot) - \pi_{\theta_{n-r_\varepsilon(n)}}\big\|_{\text{TV}}\big].$$

Under A2(a), for any large n this expectation is upper bounded by ε. Lemma 4.1 shows that

$$\big|\mathbb{E}[f(X_n) - P^{r_\varepsilon(n)}_{\theta_{n-r\varepsilon(n)}} f(X_{n-r_\varepsilon(n)})]\big| \leq \sum_{j=1}^{r_\varepsilon(n)-1} \mathbb{E}\big[D\big(\theta_{n-r_\varepsilon(n)+j}, \theta_{n-r_\varepsilon(n)}\big)\big].$$

Under A2(b), the RHS tends to zero as $n \to +\infty$. Finally, the remaining term in (19) converges to zero, as a consequence of the a.s. convergence of $\{\pi_{\theta_n}(f), n \geq 0\}$ to α, and of the property $\lim_n n - r_\varepsilon(n) = +\infty$. $\square$

4.2. *Proof of Lemma 2.5.* The proof of (i) follows by iterating the drift inequality in Saksman and Vihola (2010), Proposition 15. We now prove (ii). Saksman and Vihola [(2010), Proposition 15] implies that there exists a constant c such that on the set $\{\sup_{k \leq n-1} k^{-\tau} |\theta_k| \leq t\}$,

$$\sup_{k \leq n-1} \lambda_{\theta_k} \leq 1 - (ct^{d\gamma/2} k^{\tau d\gamma/2})^{-1} \leq 1 - (ct^{d\gamma/2} n^{\tau d\gamma/2})^{-1}, \qquad \mathbb{P}\text{-a.s.}$$

Then by iterating the drift inequality in Saksman and Vihola [(2010), Proposition 15] this yields

$$\mathbb{E}\big[W(X_n)\mathbb{1}_{\sup_{k \leq n-1} k^{-\tau} |\theta_k| \leq t}\big]$$

$$\leq \mathbb{E}[W(X_0)] + b \sum_{k=0}^{n-1} (1 - (ct^{d\gamma/2} n^{\tau d\gamma/2})^{-1})^k$$

$$\leq \mathbb{E}[W(X_0)] + b(ct^{d\gamma/2} n^{\tau d\gamma/2}).$$

The last assertion follows from (10), (ii), and the Markov inequality: let $\varepsilon, \tau > 0$; choose t_ε and $\tau' > 0$ such that $\tau - \tau' d\gamma/2 > 0$ and $\mathbb{P}(\sup_{n \geq 1} |\theta_n| n^{-\tau'} \geq t_\varepsilon) \leq \varepsilon/2$. Then

$$\mathbb{P}\Big[\sup_n n^{-1-\tau} W(X_n) \geq M\Big]$$

$$\leq \varepsilon/2 + \mathbb{P}\Big[\sup_n n^{-1-\tau} W(X_n) \geq M, \sup_{n \geq 1} |\theta_n| n^{-\tau'} \leq t_\varepsilon\Big]$$

$$\leq \varepsilon/2 + \frac{1}{M}\mathbb{E}\Big[\sup_{n\geq 1} n^{-1-\tau} W(X_n)\mathbb{1}_{\sup_{n\geq 1}|\theta_n|n^{-\tau'}\leq t_\varepsilon}\Big]$$

$$\leq \varepsilon/2 + \frac{C}{M}\sum_{n\geq 1}\frac{1}{n^{1+\tau}}n^{\tau' d\gamma/2}$$

for some constant C, and the RHS is upper bounded by ε for large enough M.

4.3. *Proof of Theorem* 2.7. The proof of Theorem 2.7 is prefaced by lemmas on the regularity in θ of the invariant distribution π_θ and on the function $\hat{F}_\theta$ solution of the Poisson equation $\hat{F}_\theta - P_\theta \hat{F}_\theta = F(\cdot, \theta) - \pi_\theta(F(\cdot, \theta))$.

Under A3, $\hat{F}_\theta(x) \stackrel{\text{def}}{=} \sum_n P_\theta^n\{F(\cdot, \theta) - \pi_\theta(F(\cdot, \theta))\}(x)$ exists for all $x \in \mathsf{X}$, solves the Poisson equation, and by Lemma 2.3

(20) $$|\hat{F}_\theta(x)| \leq \|F(\cdot, \theta)\|_V L_\theta^2 V(x),$$

where L_θ is defined in (5).

The following lemma is adapted from Andrieu et al. (2011). A detailed proof is given in Section 3 of the supplemental paper [Fort, Moulines and Priouret (2011)].

LEMMA 4.2. *Assume* A3. *For any* $\theta \in \Theta$, *let* $F_\theta : \mathsf{X} \to \mathbb{R}^+$ *be a measurable function such that* $\sup_\theta \|F_\theta\|_V < +\infty$ *and define* $\hat{F}_\theta \stackrel{\text{def}}{=} \sum_{n\geq 0} P_\theta^n\{F_\theta - \pi_\theta(F_\theta)\}$. *For any* $\theta, \theta' \in \Theta$,

$$\|\pi_\theta - \pi_{\theta'}\|_V \leq L_{\theta'}^2\{\pi_\theta(V) + L_\theta^2 V(x)\}D_V(\theta, \theta')$$

and

$$|P_\theta \hat{F}_\theta - P_{\theta'}\hat{F}_{\theta'}|_V \leq \sup_{\theta\in\Theta}\|F_\theta\|_V L_{\theta'}^2(L_\theta D_V(\theta, \theta') + \|\pi_\theta - \pi_{\theta'}\|_V)$$

$$+ L_{\theta'}^2\|F_\theta - F_{\theta'}\|_V,$$

where L_θ *is given by* (5).

PROOF OF THEOREM 2.7. We denote by L the limit $\lim_n \int \pi_{\theta_n}(\mathrm{d}x)F(\theta_n, x)$. We write $\frac{1}{n}\sum_{k=0}^{n-1} F(X_k, \theta_k) - L = \sum_{i=1}^4 T_{i,n}$ with

$$T_{1,n} \stackrel{\text{def}}{=} \frac{1}{n}F(X_0, \theta_0) - \frac{L}{n},$$

$$T_{2,n} \stackrel{\text{def}}{=} \frac{1}{n}\sum_{k=1}^{n-1}\{F(X_k, \theta_k) - F(X_k, \theta_{k-1})\},$$

$$T_{3,n} \stackrel{\text{def}}{=} \frac{1}{n}\sum_{k=1}^{n-1}\Big\{F(X_k, \theta_{k-1}) - \int \pi_{\theta_{k-1}}(\mathrm{d}x)F(x, \theta_{k-1})\Big\},$$

$$T_{4,n} \stackrel{\text{def}}{=} \frac{1}{n}\sum_{k=0}^{n-2}\Big\{\int \pi_{\theta_k}(\mathrm{d}x)F(x, \theta_k) - L\Big\}.$$

Consider first $T_{1,n}$. Since $|F(X_0, \theta_0)| < +\infty$ $\mathbb{P}$-a.s., $\lim_{n\to\infty} T_{1,n} = 0$ $\mathbb{P}$-a.s. Under conditions (ii) [resp., (iii)], $T_{2,n}$ (resp., $T_{4,n}$) converges to zero a.s. (for $T_{2,n}$, note that $L_\theta \geq 1$ by definition). Consider finally $T_{3,n}$:

$$\frac{1}{n} \sum_{k=1}^{n-1} \left\{ F(X_k, \theta_{k-1}) - \int \pi_{\theta_{k-1}}(\mathrm{d}x) F(x, \theta_{k-1}) \right\} = M_n + R_n + \tilde{R}_n$$

with $\hat{F}_\theta(x) \stackrel{\text{def}}{=} \sum_{n\geq 0} P_\theta^n \{F(\cdot, \theta) - \pi_\theta(F(\cdot, \theta))\}(x)$ and

$$M_n \stackrel{\text{def}}{=} \frac{1}{n} \sum_{k=1}^{n-1} \{\hat{F}_{\theta_{k-1}}(X_k) - P_{\theta_{k-1}} \hat{F}_{\theta_{k-1}}(X_{k-1})\},$$

$$R_n \stackrel{\text{def}}{=} \frac{1}{n} \sum_{k=1}^{n-1} \{P_{\theta_k} \hat{F}_{\theta_k}(X_k) - P_{\theta_{k-1}} \hat{F}_{\theta_{k-1}}(X_k)\},$$

$$\tilde{R}_n \stackrel{\text{def}}{=} \frac{1}{n} P_{\theta_0} \hat{F}_{\theta_0}(X_0) - \frac{1}{n} P_{\theta_{n-1}} \hat{F}_{\theta_{n-1}}(X_{n-1}).$$

By construction, $\{\hat{F}_{\theta_{k-1}}(X_k) - P_{\theta_{k-1}} \hat{F}_{\theta_{k-1}}(X_{k-1}), k \geq 1\}$ is a martingale-increment sequence. Therefore, by Hall and Heyde [(1980), Theorem 2.18], $M_n \xrightarrow{\text{a.s.}} 0$ provided that

$$(21) \qquad \sum_{k\geq 1} \frac{1}{k^\alpha} \mathbb{E}[|\hat{F}_{\theta_{k-1}}(X_k) - P_{\theta_{k-1}} \hat{F}_{\theta_{k-1}}(X_{k-1})|^\alpha | \mathcal{F}_{k-1}] < +\infty, \qquad \mathbb{P}\text{-a.s.}$$

Equation (20) and Jensen's inequality imply that ($\alpha > 1$)

$$\mathbb{E}[|\hat{F}_{\theta_{k-1}}(X_k) - P_{\theta_{k-1}} \hat{F}_{\theta_{k-1}}(X_{k-1})|^\alpha | \mathcal{F}_{k-1}]$$

$$\leq 2^{\alpha-1} \mathbb{E}[|\hat{F}_{\theta_{k-1}}(X_k)|^\alpha + |P_{\theta_{k-1}} \hat{F}_{\theta_{k-1}}(X_{k-1})|^\alpha | \mathcal{F}_{k-1}]$$

$$\leq 2^\alpha \left(\sup_\theta \|F(\cdot, \theta)\|_V L_{\theta_{k-1}}^2 \right)^\alpha P_{\theta_{k-1}} V^\alpha(X_{k-1}).$$

Under item (i) and A5(b), the series is finite $\mathbb{P}$-a.s. and this concludes the proof of (21). Consider now the remainder term R_n. By Lemma 4.2,

$$|R_n| \leq \frac{\sup_\theta \|F(\cdot, \theta)\|_V}{n}$$

$$\times \sum_{k=1}^{n} L_{\theta_k}^2 L_{\theta_{k-1}}^2 \{1 + \pi_{\theta_k}(V) + L_{\theta_k}^2\} D_V(\theta_k, \theta_{k-1}) V(X_k)$$

$$+ \frac{1}{n} \sum_{k=1}^{n} L_{\theta_k}^2 \|F(\cdot, \theta_k) - F(\cdot, \theta_{k-1})\|_V V(X_k).$$

Assumptions A4, A5(a) and items (i), (ii) imply that $R_n \xrightarrow{\text{a.s.}} 0$. Consider finally $\tilde{R}_n$. By (20),

$$\frac{1}{n}|P_{\theta_0}\hat{F}_{\theta_0}(X_0) - P_{\theta_{n-1}}\hat{F}_{\theta_{n-1}}(X_{n-1})|$$

$$\leq \frac{\sup_\theta \|F(\cdot, \theta)\|_V}{n}(L_{\theta_0}^2 P_{\theta_0} V(X_0) + L_{\theta_{n-1}}^2 P_{\theta_{n-1}} V(X_{n-1}))$$

$$\leq \frac{\sup_\theta \|F(\cdot, \theta)\|_V}{n}(L_{\theta_0}^2 \{V(X_0) + b_{\theta_0}\} + L_{\theta_{n-1}}^2 P_{\theta_{n-1}} V(X_{n-1})).$$

Assumption A5(b), item (i) and the condition $V(X_0) < +\infty$ $\mathbb{P}$-a.s. imply that $\tilde{R}_n \xrightarrow{\text{a.s.}} 0$. $\square$

4.4. *Proof of Theorem 2.11.* We preface the proof of this theorem by a proposition and a lemma. The proof of Proposition 4.3 is postponed to Fort, Moulines and Priouret (2011), Section 4.

PROPOSITION 4.3. *Let X be a Polish space endowed with its Borel σ-field $\mathcal{X}$. Let μ and $\{\mu_n, n \geq 1\}$ be probability distributions on $(\mathsf{X}, \mathcal{X})$. Let $\{h_n, n \geq 0\}$ be an equicontinuous family of functions from X to $\mathbb{R}$. Assume:*

(i) *the sequence $\{\mu_n, n \geq 0\}$ converges weakly to μ,*

(ii) *for any $x \in \mathsf{X}$, $\lim_{n\to\infty} h_n(x)$ exists, and there exists $\alpha > 1$ such that $\sup_n \mu_n(|h_n|^\alpha) + \mu(|\lim_n h_n|) < +\infty$.*

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

LEMMA 4.4. *Let X be a Polish space endowed with its Borel σ-field $\mathcal{X}$. Let $\{P_\theta, \theta \in \Theta\}$ be a family of transition kernels on $(\mathsf{X}, \mathcal{X})$ and $\{\theta_n, n \geq 0\}$ be a Θ-valued random sequence on $(\Omega, \mathcal{A}, \mathbb{P})$. Assume conditions (ii) and (iii) of Theorem 2.11. Then, there exists a $\mathbb{P}$-full set $\Omega_\star$ such that for any $\omega \in \Omega_\star$, $x \in \mathsf{X}$ and $k \geq 1$, the probability distributions $\{P_{\theta_n(\omega)}^k(x, \cdot), n \geq 0\}$ converge weakly to $P_{\theta_\star}^k(x, \cdot)$.*

PROOF. We prove, by induction on k, that there exists a $\mathbb{P}$-full set Ω_k such that for any $\omega \in \Omega_k$ and $x \in \mathsf{X}$, the probability distributions $\{P_{\theta_n(\omega)}^k(x, \cdot), n \geq 0\}$ converge weakly to $P_{\theta_\star}^k(x, \cdot)$. The proof is then concluded by setting $\Omega_\star = \bigcap_k \Omega_k$.

Consider the case $k = 1$. By condition (iii) of Theorem 2.11, for any $x \in \mathsf{X}$ there exists a $\mathbb{P}$-full set Ω_x such that for any $\omega \in \Omega_x$, $\{P_{\theta_n(\omega)}(x, \cdot), n \geq 0\}$ converges weakly to $P_{\theta_\star}(x, \cdot)$. Since X is Polish, it admits a countable dense subset $\mathcal{D}$. Therefore, there exists a $\mathbb{P}$-full set $\Omega_{\mathcal{D}}$ such that for any $\omega \in \Omega_{\mathcal{D}}$ and any $x \in \mathcal{D}$, $\{P_{\theta_n(\omega)}(x, \cdot), n \geq 0\}$ converges weakly to $P_{\theta_\star}(x, \cdot)$. Under condition (ii) of Theorem 2.11, for any bounded continuous function f, the family of functions

$\{\bar{P}_\theta f \overset{\text{def}}{=} P_\theta f - P_{\theta_\star} f, \theta \in \Theta\}$ is equicontinuous. For any $\varepsilon > 0$ and any $x \in \mathsf{X}$, there thus exists $x_\varepsilon \in \mathcal{D}$ such that for any $\theta \in \Theta$, $|\bar{P}_\theta f(x) - \bar{P}_\theta f(x_\varepsilon)| \leq \varepsilon$. Hence, for any $\omega \in \Omega_\mathcal{D}$ and any bounded continuous function f,

$$\left|\bar{P}_{\theta_n(\omega)} f(x)\right| \leq \left|\bar{P}_{\theta_n(\omega)} f(x_\varepsilon)\right| + \left|\bar{P}_{\theta_n(\omega)} f(x) - \bar{P}_{\theta_n(\omega)} f(x_\varepsilon)\right|$$

$$\leq \left|P_{\theta_n(\omega)} f(x_\varepsilon) - P_{\theta_\star} f(x_\varepsilon)\right| + \varepsilon.$$

This implies that $\limsup_n |\bar{P}_{\theta_n(\omega)} f(x)| \leq \varepsilon$. Since ε was arbitrary, it follows $\{P_{\theta_n(\omega)}(x, \cdot), n \geq 0\}$ converges weakly to $P_{\theta_\star}(x, \cdot)$ for any x. Hence, we set $\Omega_1 = \Omega_\mathcal{D}$.

Assume that the property holds for $k \geq 1$. We write for any bounded and continuous function f

(22)
$$P_{\theta_n(\omega)}^{k+1} f(x) - P_{\theta_\star}^{k+1} f(x) = \int \left(P_{\theta_n(\omega)}^k(x, \mathrm{d}y) - P_{\theta_\star}^k(x, \mathrm{d}y)\right) P_{\theta_\star} f(y)$$

$$+ \int P_{\theta_n(\omega)}^k(x, \mathrm{d}y) \left(P_{\theta_n(\omega)} f(y) - P_{\theta_\star} f(y)\right).$$

By the induction assumption, there exists a $\mathbb{P}$-full set Ω_k such that for any $\omega \in \Omega_k$, $x \in \mathsf{X}$ and any bounded continuous function h, $\lim_{n \to \infty} P_{\theta_n(\omega)}^k h(x) = P_{\theta_\star}^k h(x)$. Applied with $h = P_{\theta_\star} f$, which is continuous under the assumption (ii), this proves that for any $\omega \in \Omega_k$, the first term on the RHS of (22) goes to zero. For the second term, we use Proposition 4.3. Let $\omega \in \Omega_k \cap \Omega_1$. For any $x \in \mathsf{X}$, $\{P_{\theta_n(\omega)}^k(x, \cdot), n \geq 0\}$ converges weakly to $P_{\theta_\star}^k(x, \cdot)$. Furthermore, the family of bounded functions $\{P_{\theta_n(\omega)} f - P_{\theta_\star} f, n \geq 0\}$ is equicontinuous and, since $\omega \in \Omega_1$, $\lim_{n \to \infty} P_{\theta_n(\omega)} f(y) - P_{\theta_\star} f(y) = 0$ for any $y \in \mathsf{X}$. Proposition 4.3 thus implies that the second term on the RHS of (22) converges to zero, for any bounded continuous function f. The above discussion proves that $\Omega_{k+1} = \Omega_k \cap \Omega_1 = \Omega_1$, and concludes the induction. $\square$

PROOF OF THEOREM 2.11. Fix $x \in \mathsf{X}$. Let f be a bounded continuous function on X. Under A3, we have by Lemma 2.3

$$\limsup_n |\pi_{\theta_n}(f) - P_{\theta_n}^k f(x) + P_{\theta_\star}^k f(x) - \pi_{\theta_\star}(f)|$$

$$\leq \left(\limsup_n C_{\theta_n} [\limsup_n \rho_{\theta_n}]^k + C_{\theta_\star} \rho_{\theta_\star}^k\right) V(x).$$

By Lemma 2.3 and condition (i), $\limsup_n C_{\theta_n} < +\infty$ and $\limsup_n \rho_{\theta_n} < 1$ $\mathbb{P}$-a.s.; then, there exists a $\mathbb{P}$-full set $\Omega_\star''$ such that for any $\omega \in \Omega_\star''$, there exists $k(\omega)$ such that

$$\limsup_n \left|\pi_{\theta_n(\omega)}(f) - P_{\theta_n(\omega)}^{k(\omega)} f(x) + P_{\theta_\star}^{k(\omega)} f(x) - \pi_{\theta_\star}(f)\right| \leq \varepsilon.$$

Note that $\Omega_\star''$ does not depend upon x and f. By Lemma 4.4, there exists a $\mathbb{P}$-full set $\Omega_\star$ such that $\lim_{n \to \infty} P_{\theta_n(\omega)}^k f(x) = P_{\theta_\star}^k f(x)$ for any $\omega \in \Omega_\star$, any $x \in \mathsf{X}$, any $k \geq 1$ and any bounded continuous function f. The proof is concluded by setting $\Omega_\star' = \Omega_\star'' \cap \Omega_\star$. $\square$

5. Proofs of Section 3.

5.1. *Proof of Corollary* 3.2. (i) By iterating the drift inequality (17), we obtain

$$\mathbb{E}[W(X_n)] \leq \tilde{\lambda}^n \mathbb{E}[W(X_0)] + \tilde{b} \sum_{k=0}^{n-1} \tilde{\lambda}^k \mathbb{E}[\theta_{n-k}(W)].$$

Under I3(b), $\sup_{k\geq 0} \mathbb{E}[\theta_k(W)] < +\infty$ so that

$$(23) \qquad \mathbb{E}[W(X_n)] \leq \tilde{\lambda}^n \mathbb{E}[W(X_0)] + \frac{\tilde{b}}{1-\tilde{\lambda}} \sup_{k\geq 0} \mathbb{E}[\theta_k(W)].$$

(ii) Since W is a continuous function, I3(a) implies that $\limsup_n \theta_n(W) < +\infty$, $\mathbb{P}$-a.s. Consequently, $\limsup_n L_{\theta_n} < +\infty$, $\mathbb{P}$-a.s. by Lemma 2.3 and Proposition 3.1.

5.2. *Proof of Proposition* 3.3. We check the conditions of Theorem 2.11. Condition (i) of Theorem 2.11 holds by Corollary 3.2.

The proof of condition (ii) of Theorem 2.11 is a consequence of the following lemma.

LEMMA 5.1. *Let f be a function on* X *such that* $\|f\pi^\beta\|_\infty < +\infty$. *For any* $x, x' \in \mathsf{X}$ *such that* $\pi(x) > 0, \pi(x') > 0$,

$$\sup_{\theta\in\Theta} |P_\theta f(x) - P_\theta f(x')| \leq |Pf(x) - Pf(x')| + |f(x) - f(x')|$$

$$+ 2\|f\pi^\beta\|_\infty |\pi^{-\beta}(x) - \pi^{-\beta}(x')|.$$

PROOF. By definition of the transition kernel P_θ, it is easily checked that

$$(24) \qquad \begin{aligned} P_\theta f&(x) - P_\theta f(x') \\ &= \upsilon \int \{\alpha(x, y) - \alpha(x', y)\}(f(y) - f(x'))\theta(dy) \\ &\quad + (1-\upsilon)(Pf(x) - Pf(x')) + \upsilon(f(x) - f(x'))A(\theta, x), \end{aligned}$$

where $A(\theta, x) \overset{\text{def}}{=} 1 - \int \alpha(x, y)\theta(dy)$. Since $0 \leq \alpha(x, y) \leq 1$, we have

$$|\upsilon(f(x) - f(x'))A(\theta, x)| \leq |f(x) - f(x')|.$$

We can assume w.l.o.g. that $\pi(x) \leq \pi(x')$. By definition of the ratio α, we have

$$\begin{aligned} \alpha(x, y) - \alpha(x', y) &= \mathbb{1}_{\{\pi(x)\leq\pi(y)\leq\pi(x')\}}(\pi^{-\beta}(y) - \pi^{-\beta}(x'))\pi^\beta(y) \\ &\quad + (\pi^{-\beta}(x) - \pi^{-\beta}(x'))\mathbb{1}_{\{\pi(y)\leq\pi(x)\leq\pi(x')\}}\pi^\beta(y), \end{aligned}$$

showing that $|\alpha(x, y) - \alpha(x', y)| \leq (\pi^{-\beta}(x) - \pi^{-\beta}(x'))\pi^\beta(y)\mathbb{1}_{\{\pi(y)\leq\pi(x')\}}$. The proof is concluded by noting that

$$\int |\alpha(x, y) - \alpha(x', y)||f(y) - f(x')|\theta(\mathrm{d}y)$$

$$\leq 2\Big(\sup_X |f|\pi^\beta\Big)(\pi^{-\beta}(x) - \pi^{-\beta}(x')). \qquad \square$$

The most delicate part consists in establishing condition (iii) of Theorem 2.11. The proof relies on the following result which is an extension of the Varadarajan theorem [Dudley (2002), Theorem 11.4.1]. The proof of Proposition 5.2 is detailed in Section 5 of the supplemental paper [Fort, Moulines and Priouret (2011)].

PROPOSITION 5.2. *Let* $(\mathbb{U}, d)$ *be a metric space equipped with its Borel σ-field $\mathcal{B}(\mathbb{U})$. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, μ be a distribution on $(\mathbb{U}, \mathcal{B}(\mathbb{U}))$ and $\{K_n, n \geq 0\}$ be a family of Markov transition kernels $K_n : \Omega \times \mathcal{B}(\mathbb{U}) \to [0, 1]$. Assume that, for any $f \in \mathsf{C}_b(\mathbb{U}, d)$*

$$\Omega_f \overset{def}{=} \Big\{\omega \in \Omega : \limsup_{n\to\infty} |K_n(\omega, f) - \mu(f)| = 0\Big\}$$

is a $\mathbb{P}$-full set. Then

$$\Big\{\omega \in \Omega : \forall f \in \mathsf{C}_b(\mathbb{U}, d) \ \limsup_{n\to\infty} |K_n(\omega, f) - \mu(f)| = 0\Big\}$$

is a $\mathbb{P}$-full set.

PROOF OF (iii) OF THEOREM 2.11. We check the conditions of Proposition 5.2 with $\mu_n = P_{\theta_n}(x, \cdot)$ and $\mu = P_{\theta_\star}(x, \cdot)$. For any $x \in \mathsf{X}$, and $f \in \mathsf{C}_b(\mathsf{X})$, $y \mapsto \alpha(x, y)$ and $y \mapsto \alpha(x, y)f(y)$ are continuous. Thus, I3(a) implies that $P_{\theta_n} f(x) \overset{\text{a.s.}}{\longrightarrow} P_{\theta_\star} f(x)$ and Ω_f is a $\mathbb{P}$-full set. $\square$

5.3. *Proof of Theorem* 3.6. Set $\alpha_0 = 1$ and choose $\alpha_l > 1$ such that $\tilde{T} \times \prod_{l=0}^K \alpha_l = T_\star$. The proof is by induction on i for $i = K$ down to $i = 2$.

Set $W^{(K-1)} \overset{\text{def}}{=} \pi^{-T_\star^{-1} \prod_{l=0}^{K-1} \alpha_l} = \pi^{-1/(\tilde{T}\alpha_K)}$ and $\pi^{(K-1)}$ be the probability distribution proportional to $\pi^{1/T_{K-1}}$. Under the stated assumptions, Theorem 3.5 applies with $Y \leftarrow X^{(K)}$ and $X \leftarrow X^{(K-1)}$: for any continuous function f in $\mathcal{L}_{W^{(K-1)}}$, $n^{-1} \sum_{k=1}^n f(X_k^{(K-1)}) \overset{\text{a.s.}}{\longrightarrow} \pi^{(K-1)}(f)$.

Assume Theorem 3.5 holds with $Y \leftarrow X^{(i+1)}$ and $X \leftarrow X^{(i)}$ for some $i \in \{2, \ldots, K - 1\}$: for any continuous function f in $\mathcal{L}_{W^{(i)}}$, $n^{-1} \sum_{k=1}^n f(X_k^{(i)}) \overset{\text{a.s.}}{\longrightarrow} \pi^{(i)}(f)$ where $W^{(i)} \overset{\text{def}}{=} \pi^{-T_\star^{-1} \prod_{l=0}^i \alpha_l}$ and $\pi^{(i)} \propto \pi^{1/T_i}$. We apply the above results with

$$\pi \leftarrow \pi^{1/T_{i-1}}, \qquad \theta_\star \leftarrow \pi^{1/T_i}, \qquad P \leftarrow P^{(i-1)},$$

$$T \leftarrow \frac{T_i}{T_{i-1}}, \qquad W \leftarrow W^{-T_\star^{-1} \prod_{l=0}^i \alpha_l}.$$

We thus have that $n^{-1} \sum_{k=1}^{n} f(X_k^{(i-1)}) \xrightarrow{\text{a.s.}} \pi^{(i-1)}(f)$ for any continuous function f in $\mathcal{L}_{W^{(i-1)}}$, where

$$W^{(i-1)} \stackrel{\text{def}}{=} \pi^{-T_\star^{-1} \prod_{l=0}^{i-1} \alpha_l} = \{W^{(i)}\}^{1/\alpha_i}, \qquad \pi^{(i-1)} \propto \pi^{1/T_{i-1}}.$$

This concludes the induction.

5.4. *Proof of Proposition* 3.7. For any $i \in \{1, \dots, K\}$, the transition kernels $P^{(i)}$ are π-irreducible, aperiodic, and compact sets are 1-small. In addition, they are Feller (the proof is on the same lines as the proof of Lemma 5.1). By Saksman and Vihola [(2010), Proposition 15] conditions (i) and (ii) of Theorem 3.6 are satisfied for $i \in \{1, \dots, K\}$. Note that the proof of Proposition 15 in Saksman and Vihola (2010) is in the case $sT_i = 1/2$ but it can be easily adapted for any $sT_i \in (0, 1)$. In the case $i = K$, this implies that there exist $\lambda \in (0, 1)$ and $b < +\infty$ such that

$$P^{(K)} \tilde{U} \leq \lambda \tilde{U} + b,$$

where $\tilde{U} = (\pi / \sup_\mathsf{X} \pi)^{-1/\tilde{T}}$. Standard results on Markov chains [see, e.g., Meyn and Tweedie (2009)] imply (iv). By iterating the drift inequality, we have

$$\sup_n \mathbb{E}[\tilde{U}(X_n^{(K)})] \leq \mathbb{E}[\tilde{U}(X_0^{(K)})] + \frac{b}{1 - \lambda},$$

thus proving (v). Finally, since π satisfies M1, there exist positive constants c_i such that $\pi(x) \leq c_1 \exp(-c_2|x|)$ [see, e.g., Saksman and Vihola (2010), Lemma 8]. Therefore, for any $\tau > 0$, $\int \pi^\tau(x) \, dx < +\infty$ thus showing (iii).

Acknowledgments. The authors are indebted to the associate editor and the referees for the numerous suggestions made to improve the manuscript.

SUPPLEMENTARY MATERIAL

Supplement to paper "Convergence of adaptive and interacting Markov chain Monte Carlo algorithms" (DOI: 10.1214/11-AOS938SUPP; .pdf). This supplement provides a detailed proof of Lemma 4.2 and Propositions 3.1, 4.3 and 5.2. It also contains a discussion on the setwise convergence of transition kernels.

REFERENCES

ANDRIEU, C. and MOULINES, É. (2006). On the ergodicity properties of some adaptive MCMC algorithms. *Ann. Appl. Probab.* **16** 1462–1505. MR2260070

ANDRIEU, C. and THOMS, J. (2008). A tutorial on adaptive MCMC. *Stat. Comput.* **18** 343–373. MR2461882

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 Proceedings* **19** 79–84. EDP Sci., Les Ulis. MR2405652

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

ANDRIEU, C., JASRA, A., DOUCET, A. and DEL MORAL, P. (2011). On non-linear Markov chain Monte Carlo. *Bernoulli* **17** 987–1014.

ATCHADÉ, Y. F. (2010). A cautionary tale on the efficiency of some adaptive Monte Carlo schemes. *Ann. Appl. Probab.* **20** 841–868. MR2680550

ATCHADÉ, Y. and FORT, G. (2010). Limit theorems for some adaptive MCMC algorithms with subgeometric kernels. *Bernoulli* **16** 116–154. MR2648752

ATCHADÉ, Y. F. and LIU, J. S. (2006). Discussion of "Equi-energy sampler," by Kou, Zhou and Wong. *Ann. Statist.* **34** 1620–1628.

ATCHADE, Y., ROBERTS, G. O. and ROSENTHAL, J. S. (2011). Towards optimal scaling of Metropolis-coupled Markov chain Monte Carlo. *Stat. Comput.* **21** 555–568.

ATCHADÉ, Y. F. and ROSENTHAL, J. S. (2005). On adaptive Markov chain Monte Carlo algorithms. *Bernoulli* **11** 815–828. MR2172842

ATCHADÉ, Y., FORT, G., MOULINES, E. and PRIOURET, P. (2011). Adaptive Markov chain Monte Carlo: Theory and methods. In *Bayesian Time Series Models* (D. Barber, A. T. Cemgil and S. Chiappia, eds.) 33–53. Cambridge Univ. Press, Cambridge.

BAI, Y., ROBERTS, G. O. and ROSENTHAL, J. S. (2011). On the containment condition for adaptive Markov chain Monte Carlo algorithms. *Adv. Appl. Stat.* **21** 1–54.

BAXENDALE, P. H. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.* **15** 700–738. MR2114987

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

BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd ed. Wiley, New York. MR1700749

BROCKWELL, A., DEL MORAL, P. and DOUCET, A. (2010). Sequentially interacting Markov chain Monte Carlo methods. *Ann. Statist.* **38** 3387–3411. MR2766856

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

DUDLEY, R. M. (2002). *Real Analysis and Probability. Cambridge Studies in Advanced Mathematics* **74**. Cambridge Univ. Press, Cambridge. MR1932358

FORT, G. and MOULINES, E. (2003). Polynomial ergodicity of Markov transition kernels. *Stochastic Process. Appl.* **103** 57–99. MR1947960

FORT, G., MOULINES, E. and PRIOURET, P. (2011). Supplement to "Convergence of adaptive and interacting Markov chain Monte Carlo algorithms." DOI:10.1214/11-AOS938SUPP.

GELMAN, A., ROBERTS, G. O. and GILKS, W. R. (1996). Efficient Metropolis jumping rules. In *Bayesian Statistics*, 5 (*Alicante*, 1994) 599–607. Oxford Univ. Press, New York. MR1425429

GEYER, C. J. (1991). Markov chain Monte Carlo maximum likelihood. In *Computing Science and Statistics*: *Proc.* 23*rd Symposium on the Interface*, *Interface Foundation*, *Fairfax Station*, *VA* 156–163. Amer. Statist. Assoc., Alexandria, VA.

HAARIO, H., SAKSMAN, E. and TAMMINEN, J. (1999). Adaptive proposal distribution for random walk Metropolis algorithm. *Comput. Statist.* **14** 375–395.

HAARIO, H., SAKSMAN, E. and TAMMINEN, J. (2001). An adaptive Metropolis algorithm. *Bernoulli* **7** 223–242. MR1828504

HAARIO, H., LAINE, M., LEHTINEN, M., SAKSMAN, E. and TAMMINEN, J. (2004). Markov chain Monte Carlo methods for high dimensional inversion in remote sensing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **66** 591–607. MR2088292

HAARIO, H., LAINE, M., MIRA, A. and SAKSMAN, E. (2006). DRAM: Efficient adaptive MCMC. *Stat. Comput.* **16** 339–354. MR2297535

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York. MR0624435

JARNER, S. F. and HANSEN, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.* **85** 341–361. MR1731030

KATO, T. (1980). *Perturbation Theory for Linear Operators*. Springer, Berlin.

KOU, S. C., ZHOU, Q. and WONG, W. H. (2006). Equi-energy sampler with applications in statistical inference and statistical mechanics. *Ann. Statist.* **34** 1581–1615.

LAINE, M. and TAMMINEN, J. (2008). Aerosol model selection and uncertainty modelling by adaptive MCMC technique. *Atmospheric and Chemistry Physics* **8** 7697–7707.

MENGERSEN, K. L. and TWEEDIE, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24** 101–121. MR1389882

METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. and TELLER, E. (1953). Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087–1092.

MEYN, S. and TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*. Cambridge Univ. Press, London.

ROBERT, C. P. and CASELLA, G. (2004). *Monte Carlo Statistical Methods*, 2nd ed. Springer, New York. MR2080278

ROBERTS, G. O. and ROSENTHAL, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probab. Surv.* **1** 20–71. MR2095565

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

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

ROSENTHAL, J. S. (2009). Optimal proposal distributions and adaptive MCMC. In *Handbook of MCMC*. Chapman & Hall/CRC Press, Boca Raton, FL.

SAKSMAN, E. and VIHOLA, M. (2010). On the ergodicity of the adaptive Metropolis algorithm on unbounded domains. *Ann. Appl. Probab.* **20** 2178–2203. MR2759732

G. FORT
E. MOULINES
LTCI, TELECOM PARISTECH-CNRS
46 RUE BARRAULT
75634 PARIS CEDEX 13
FRANCE
E-MAIL: gersende.fort@telecom-paristech.fr
         eric.moulines@telecom-paristech.fr

P. PRIOURET
LPMA, UNIVERSITÉ PIERRE
   ET MARIE CURIE (P6)
BOÎTE COURRIER 188
75252 PARIS CEDEX 05
FRANCE
E-MAIL: priouret@ccr.jussieu.fr