SUPPLEMENT TO PAPER "CONVERGENCE OF ADAPTIVE AND INTERACTING MARKOV CHAIN MONTE CARLO ALGORITHMS"

By G. Fort[*,†], E. Moulines[‡] and P. Priouret[§]

LTCI, CNRS - TELECOM ParisTech [†‡], *LPMA, Université Pierre et Marie Curie*[§]

This supplement provides a detailed proof of Lemma 4.2 and Propositions 3.1, 4.3 and 5.2 of Fort, Moulines and Priouret (2010). It also contains a discussion on the setwise convergence of transition kernels (see Section 1).

For completeness and ease of references, we repeat the assumptions and the main notations.

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

$$D_V(\theta, \theta') \stackrel{\text{def}}{=} \sup_{x \in \mathsf{X}} \frac{\|P_\theta(x, \cdot) - P_{\theta'}(x, \cdot)\|_V}{V(x)} . \tag{1}$$

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 non-decreasing positive sequence $\{r_\varepsilon(n), n \geq 0\}$ such that $\limsup_{n\to\infty} r_\varepsilon(n)/n = 0$ and

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

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

A3 $\sum_{k=1}^{\infty} k^{-1} \left(L_{\theta_k} \vee L_{\theta_{k-1}}\right)^6 \; D_V(\theta_k, \theta_{k-1}) \; V(X_k) < +\infty$ $\mathbb{P}$-a.s. , where D_V and L_θ are defined in (1) and (3).

A4 (a) $\limsup_n \pi_{\theta_n}(V) < +\infty$, $\mathbb{P}$-a.s.

(b) For some $\alpha > 1$, $\sum_{k=0}^{\infty} (k+1)^{-\alpha} \; L^{2\alpha}_{\theta_k} \quad P_{\theta_k} V^\alpha(X_k) < +\infty$, $\mathbb{P}$-a.s. .

*Corresponding author

A5 For all $\theta \in \Theta$, P_θ is phi-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

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

For any $\varepsilon > 0$, $x \in \mathsf{X}$, $\theta \in \Theta$, set

$$M_\varepsilon(x, \theta) \stackrel{\text{def}}{=} \inf\{n \ge 0, \|P_\theta^n(x, \cdot) - \pi_\theta\|_{\mathrm{TV}} \le \varepsilon\} \;. \tag{2}$$

For any $\theta \in \Theta$, set

$$L_\theta \stackrel{\text{def}}{=} C_\theta \vee (1-\rho_\theta)^{-1} \le C \left\{ b_\theta \vee \delta_\theta^{-1} \vee (1-\lambda_\theta)^{-1} \right\}^\gamma \;, \tag{3}$$

where C_θ and $\rho_\theta \in (0,1)$ are finite constants such that

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

1. Setwise convergence of kernels. In many situations (see e.g. (Fort, Moulines and Priouret, 2010, Section 3)), we are able to prove that

> for (a fixed) $x \in \mathsf{X}$ and any $A \in \mathcal{X}$, there exists Ω_A such that $\mathbb{P}(\Omega_A) = 1$ and for any $\omega \in \Omega_A$,
>
> $$\lim_n P_{\theta_n(\omega)}(x, A) = P_{\theta_\star}(x, A) \;.$$

This implies that if $\mathcal{B}_0$ is a countable algebra generating the σ-algebra $\mathcal{X}$, there exists Ω_0 such that $\mathbb{P}(\Omega_0) = 1$ and for any $\omega \in \Omega_0$ and $A \in \mathcal{B}_0$,

$$\lim_n P_{\theta_n(\omega)}(x, A) = P_{\theta_\star}(x, A) \;.$$

Therefore, we are faced to the question: does it imply that there exists $\Omega_\star$ such that $\mathbb{P}(\Omega_\star) = 1$ and for any $\omega \in \Omega_\star$ and $A \in \mathcal{X}$,

$$\lim_n P_{\theta_n(\omega)}(x, A) = P_{\theta_\star}(x, A) \;.$$

The answer is no, in general, as illustrated by the following counter-example.

Counter-Example: Set $P_{\theta_\star}(x, A) = \mu(A)$ and $P_{\theta_n(\omega)}(x, A) = \mu_n(\omega, A)$ where

- μ is a probability distribution such that for any $x \in \mathsf{X}$, $\mu(\{x\}) = 0$.

- $\mu_n(\omega, A) \stackrel{\text{def}}{=} n^{-1} \sum_{k=1}^n \mathbb{1}_A(X_k(\omega))$ with $\{X_k, k \geq 1\}$ i.i.d. r.v. defined on $(\Omega, \mathcal{F}, \mathbb{P})$ taking value in $(\mathsf{X}, \mathcal{X})$, with distribution μ.

Then, by the strong law of large numbers, for any $A \in \mathcal{X}$, $\lim_n \mu_n(\cdot, A) = \mu(A)$ $\mathbb{P}$-a.s. If we take a countable family of measurable sets $\mathcal{B}_0$, then we may find a $\mathbb{P}$-full set $\mathcal{D} \subseteq \Omega$, such that for any $\omega \in \mathcal{D}$, and $A \in \mathcal{B}_0$, $\lim_{n\to\infty} \mu_n(\omega, A) = \mu(A)$. Of course, $\mathcal{B}_0$ can be an algebra, and even an algebra generating $\mathcal{X}$. Nevertheless, it is wrong to assume that this condition implies the setwise convergence, *i.e.* that $\lim_{n\to\infty} \mu_n(\omega, A) = \mu(A)$ for any $\omega \in \mathcal{D}$ and $A \in \mathcal{X}$. To see why this is wrong, choose $\omega_\star \in \mathcal{D}$ and set $A \stackrel{\text{def}}{=} \bigcup_n \{X_n(\omega_\star)\}$. Then $\mu_n(\omega_\star, A) = 1$ for any n, and $\mu(A) = 0 \neq \lim_n \mu_n(\omega_\star, A)$.

2. Proof of (Fort, Moulines and Priouret, 2010, Proposition 3.1).

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

$$PW \leq \lambda W + b \quad \text{with} \quad W(x) \stackrel{\text{def}}{=} (\pi(x)/\|\pi\|_\infty)^{-\tau} \ . \tag{4}$$

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

Proposition (Proposition 3.1 in Fort, Moulines and Priouret (2010)). *Assume I1, I2. There exist $\tilde{\lambda} \in (0,1)$, $\tilde{b} < \infty$, such that, for any $\theta \in \Theta$,*

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

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 (whatever θ) and the minorization constant does not depend upon θ.

Proof. We prove the drift inequality with $a = 1$. The proof for $a < 1$ follows from the Jensen's inequality. The proof is adapted from (Atchadé, 2010, Lemma 4.1.). Under I2b, we have

$$P_\theta W(x) \leq (1-\upsilon)\ \lambda\ W(x) + (1-\upsilon)\ b + \upsilon W(x) \\ + \upsilon \int \alpha(x,y)\ \{W(y) - W(x)\}\, \theta(\mathrm{d}y) \ .$$

By definition of W and of the acceptance ratio α,

$$\begin{aligned}
&\int \alpha(x,y)\ \{W(y)-W(x)\}\,\theta(\mathrm{d}y) \\
&= \int W(y)\ \left(1 \wedge \frac{\pi^\beta(y)}{\pi^\beta(x)}\right)\ \left\{1-\frac{\pi^\tau(y)}{\pi^\tau(x)}\right\}\ \theta(\mathrm{d}y) \\
&\leq \int_{\{y,\pi(y)\leq\pi(x)\}} W(y)\ \frac{\pi^\beta(y)}{\pi^\beta(x)}\ \left\{1-\frac{\pi^\tau(y)}{\pi^\tau(x)}\right\}\ \theta(\mathrm{d}y) \\
&\leq \Psi(\tau/\beta)\ \theta(W)
\end{aligned}$$

where we have used that, for $a>0$,

$$\sup_{z\in[0,1]} z(1-z^a) \leq \Psi(a) \stackrel{\text{def}}{=} a/(a+1)^{(a+1)/a}\ ,$$

as in (Atchadé, 2010, Lemma 1.5.1). Combining the two latter inequalities yield

$$P_\theta W(x) \leq [(1-\upsilon)\ \lambda+\upsilon]\,W(x)+\upsilon\Psi(\tau/\beta)\theta(W)+(1-\upsilon)b\ .$$

The proof of the smallness condition relies on the inequality $P_\theta(x,A) \geq (1-\upsilon)P(x,A)$. $\square$

3. Proof of (Fort, Moulines and Priouret, 2010, Lemma 4.2) .

LEMMA (Lemma 4.2 Fort, Moulines and Priouret (2010)). *Assume A5. 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*

$$\hat{F}_\theta \stackrel{\text{def}}{=} \sum_{n\geq 0} P^n_\theta\{F_\theta-\pi_\theta(F_\theta)\}\ .$$

For any $\theta,\theta'\in\Theta$,

$$\|\pi_\theta-\pi_{\theta'}\|_V \leq L^2_{\theta'}\left\{\pi_\theta(V)+L^2_\theta\ V(x)\right\}\ D_V(\theta,\theta')\ , \tag{5}$$

and

$$\begin{aligned}
\left|P_\theta\hat{F}_\theta-P_{\theta'}\hat{F}_{\theta'}\right|_V \leq \sup_{\theta\in\Theta}\|F_\theta\|_V\ L^2_{\theta'}\ \left(L_\theta\ D_V(\theta,\theta')+\|\pi_\theta-\pi_{\theta'}\|_V\right) \\
+L^2_{\theta'}\ \|F_\theta-F_{\theta'}\|_V\ .
\end{aligned} \tag{6}$$

where L_θ is given by (3).

PROOF. The proof of this Lemma is closely related to (Andrieu and Moulines, 2006, Proposition 3) and its refinement in Andrieu *et al.* (2011). These types of results have a rather long history: Benveniste, Métivier and Priouret (1990) and Glynn and Meyn (1996) and the references therein for early references.

We first establish (5). For any $k \geq 1$, we decompose $P_\theta^k f - P_{\theta'}^k f$ as follows

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

Under A5, there exist constants C_θ and $\rho_\theta \in (0,1)$ such that $\left\| P_\theta^k(x,\cdot) - \pi_\theta \right\|_V \leq C_\theta \rho_\theta^k V(x)$. Therefore, for any $k \geq 1$ and $x_\star \in \mathsf{X}$,

$$\begin{aligned}
&\|\pi_\theta - \pi_{\theta'}\|_V \\
&\leq \left\|\pi_\theta - P_\theta^k(x_\star,\cdot)\right\|_V + \left\|P_\theta^k(x_\star,\cdot) - P_{\theta'}^k(x_\star,\cdot)\right\|_V + \left\|P_{\theta'}^k(x_\star,\cdot) - \pi_{\theta'}\right\|_V \\
&\leq \left(C_\theta \rho_\theta^k + C_{\theta'} \rho_{\theta'}^k\right) V(x_\star) \\
&\qquad + \sup_{\|f\|_V \leq 1} \left| \sum_{j=0}^{k-1} P_\theta^j \left(P_\theta - P_{\theta'}\right) \left(P_{\theta'}^{k-j-1} f - \pi_{\theta'}(f)\right)(x_\star) \right| .
\end{aligned}$$

The second term on the RHS is upper bounded by

$$\begin{aligned}
&C_{\theta'} \; D_V(\theta,\theta') \sum_{j=0}^{k-1} \rho_{\theta'}^{k-j-1} \; P_\theta^j V(x_\star) \\
&\quad \leq C_{\theta'} \; D_V(\theta,\theta') \sum_{j=0}^{k-1} \rho_{\theta'}^{k-j-1} \; \left\{\pi_\theta(V) + C_\theta \rho_\theta^j V(x_\star)\right\} \\
&\quad \leq \frac{C_{\theta'}}{1-\rho_{\theta'}} \; D_V(\theta,\theta') \left(\pi_\theta(V) + C_\theta V(x_\star)\right) .
\end{aligned}$$

Therefore

$$\begin{aligned}
\|\pi_\theta - \pi_{\theta'}\|_V \leq \left(C_\theta \rho_\theta^k + C_{\theta'} \rho_{\theta'}^k\right) \; V(x_\star) \\
+ \frac{C_{\theta'}}{1-\rho_{\theta'}} \; D_V(\theta,\theta') \left(\pi_\theta(V) + C_\theta V(x_\star)\right)
\end{aligned}$$

which implies the desired result by taking the limit as $k \to +\infty$.

We then establish (6). Under A5, $\hat{F}_\theta$ exists (see (20) of Fort, Moulines and Priouret (2010)) and

$$P_\theta \hat{F}_\theta(x) - P_{\theta'} \hat{F}_{\theta'}(x) = \sum_{n \geq 1} P_\theta^n \{F_\theta - \pi_\theta(F_\theta)\} - \sum_{n \geq 1} P_{\theta'}^n \{F_{\theta'} - \pi_{\theta'}(F_{\theta'})\} .$$

We first show

$$
\begin{aligned}
(7)\quad P_\theta \hat{F}_\theta - P_{\theta'} \hat{F}_{\theta'} = \sum_{n \geq 1} \sum_{j=0}^{n-1} \left(P_\theta^j - \pi_\theta \right) \left(P_\theta - P_{\theta'} \right) \left(P_{\theta'}^{n-j-1} F_\theta - \pi_{\theta'}(F_\theta) \right) \\
- \sum_{n \geq 1} \{ P_{\theta'}^n (F_{\theta'} - F_\theta) - \pi_{\theta'}(F_{\theta'} - F_\theta) \} - \sum_{n \geq 1} \pi_\theta \{ P_{\theta'}^n F_\theta - \pi_{\theta'}(F_\theta) \} \;.
\end{aligned}
$$

Following the same lines as in (Andrieu and Moulines, 2006, Proposition 3), we have for any $n \geq 1$,

$$
\begin{aligned}
& P_\theta^n f - P_{\theta'}^n f \\
& = \sum_{j=0}^{n-1} \left(P_\theta^j - \pi_\theta \right) \left(P_\theta - P_{\theta'} \right) \left(P_{\theta'}^{n-j-1} f - \pi_{\theta'}(f) \right) \\
& \quad + \sum_{j=0}^{n-1} \left\{ \pi_\theta P_{\theta'}^{n-j-1} f - \pi_\theta P_{\theta'}^{n-j} f \right\} \\
& = \sum_{j=0}^{n-1} \left(P_\theta^j - \pi_\theta \right) \left(P_\theta - P_{\theta'} \right) \left(P_{\theta'}^{n-j-1} f - \pi_{\theta'}(f) \right) + \pi_\theta(f) - \pi_\theta P_{\theta'}^n f \;,
\end{aligned}
$$

where we used that $\pi_\theta P_\theta = \pi_\theta$. Hence, for any $n \geq 1$,

$$
\begin{aligned}
P_\theta^n F_\theta - P_{\theta'}^n F_{\theta'} = \sum_{j=0}^{n-1} \left(P_\theta^j - \pi_\theta \right) \left(P_\theta - P_{\theta'} \right) \left(P_{\theta'}^{n-j-1} F_\theta - \pi_{\theta'}(F_\theta) \right) \\
+ \pi_\theta(F_\theta) - \pi_\theta P_{\theta'}^n F_\theta + P_{\theta'}^n \left(F_\theta - F_{\theta'} \right)
\end{aligned}
$$

and

$$
\begin{aligned}
& P_\theta^n \{ F_\theta - \pi_\theta(F_\theta) \} - P_{\theta'}^n \{ F_{\theta'} - \pi_{\theta'}(F_{\theta'}) \} \\
& = \sum_{j=0}^{n-1} \left(P_\theta^j - \pi_\theta \right) \left(P_\theta - P_{\theta'} \right) \left(P_{\theta'}^{n-j-1} F_\theta - \pi_{\theta'}(F_\theta) \right) - \{ P_{\theta'}^n F_{\theta'} - \pi_{\theta'}(F_{\theta'}) \} \\
& \quad + \{ P_{\theta'}^n F_\theta - \pi_{\theta'}(F_\theta) \} - \pi_\theta \{ P_{\theta'}^n F_\theta - \pi_{\theta'}(F_\theta) \} \;.
\end{aligned}
$$

This yields (7). We consider the first term in (7): by Lemma 2.3 in Fort, Moulines and Priouret (2010)

$$
\begin{aligned}
& \left\| \left(P_\theta^j - \pi_\theta \right) \left(P_\theta - P_{\theta'} \right) \left(P_{\theta'}^{n-j-1} F_\theta - \pi_{\theta'}(F_\theta) \right) \right\|_V \\
& \qquad \leq C_\theta C_{\theta'} \rho_{\theta'}^{n-1-j} \rho_\theta^j \sup_\theta \| F_\theta \|_V \;\; D_V(\theta, \theta')
\end{aligned}
$$

thus implying

$$\left\| \sum_{n\geq 1} \sum_{j=0}^{n-1} \left(P_\theta^j - \pi_\theta\right) (P_\theta - P_{\theta'}) \left(P_{\theta'}^{n-j-1} F_\theta - \pi_{\theta'}(F_\theta)\right) \right\|_V \\ \leq L_\theta^2 L_{\theta'}^2 \ \sup_\theta \|F_\theta\|_V \ \ D_V(\theta, \theta').$$

Consider now the second term on the RHS of (7). By Lemma 2.3 in Fort, Moulines and Priouret (2010),

$$|P_{\theta'}^n (F_{\theta'} - F_\theta)(x) - \pi_{\theta'}(F_{\theta'} - F_\theta)| \leq C_{\theta'} \ \rho_{\theta'}^n \ V(x) \ \ \|F_{\theta'} - F_\theta\|_V$$

thus implying

$$\sum_{n\geq 1} |P_{\theta'}^n (F_{\theta'} - F_\theta)(x) - \pi_{\theta'}(F_{\theta'} - F_\theta)| \leq L_{\theta'}^2 V(x) \ \ \|F_{\theta'} - F_\theta\|_V \ .$$

Finally, the third term on the RHS of (7) can be bounded by

$$\|\pi_\theta \{P_{\theta'}^n F_\theta - \pi_{\theta'}(F_\theta)\}\|_V = \|(\pi_\theta - \pi_{\theta'}) \{P_{\theta'}^n F_\theta - \pi_{\theta'}(F_\theta)\}\|_V \\ \leq \|\pi_\theta - \pi_{\theta'}\|_V \ \ C_{\theta'} \ \rho_{\theta'}^n \sup_\theta \|F_\theta\|_V \ ,$$

so that

$$\left\| \sum_n \pi_\theta \{P_{\theta'}^n F_\theta - \pi_{\theta'}(F_\theta)\} \right\|_V \leq \|\pi_\theta - \pi_{\theta'}\|_V \ \ L_{\theta'}^2 \ \sup_\theta \|F_\theta\|_V \ .$$

□

4. Proof of (Fort, Moulines and Priouret, 2010, Proposition 4.3).

PROPOSITION (Proposition 4.3 Fort, Moulines and Priouret (2010)). *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* $\{f_n, n \geq 0\}$ *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 $x \in \mathsf{X}$, $\lim_{n\to\infty} f_n(x)$ *exists, and there exists* $\alpha > 1$ *such that* $\sup_n \mu_n(|f_n|^\alpha) + \mu(|\lim_n f_n|) < +\infty$*.*

Then, $\mu_n(f_n) \to \mu(\lim_n f_n)$*.*

Proof. Set $f \stackrel{\text{def}}{=} \lim_{n\to\infty} f_n$. Under the stated assumptions, f is continuous. We can assume without loss of generality that f_n, f are non-negative functions. Let c, ε be positive constants. We decompose $\mu_n(f_n) - \mu(f)$

$$\mu_n(f_n)-\mu(f) = \mu_n(f_n \wedge c)-\mu(f \wedge c)+\mu_n\Big((f_n-c)\mathbb{1}_{\{f_n>c\}}\Big)-\mu\Big((f-c)\mathbb{1}_{\{f>c\}}\Big) .$$

Choose c large enough so that $\sup_n \mu_n(f_n^\alpha)/c^{\alpha-1} + \mu(f\mathbb{1}_{\{f>c\}}) \leq \varepsilon$. Then upon noting that $\mathbb{1}_{\{f>c\}} \leq (f/c)^{\alpha-1}$,

$$\left|\mu_n\Big((f_n-c)\mathbb{1}_{\{f_n>c\}}\Big) - \mu\Big((f-c)\mathbb{1}_{\{f>c\}}\Big)\right| \leq \frac{\mu_n(f_n^\alpha)}{c^{\alpha-1}} + \mu(f\mathbb{1}_{\{f>c\}}) \leq \varepsilon .$$

Consider now the bounded functions $\{f_n \wedge c, n \geq 0\}$ and $f \wedge c$. We write

$$|\mu_n(f_n \wedge c) - \mu(f \wedge c)| \leq \mu_n(|f_n \wedge c - f \wedge c|) + |\mu_n(f \wedge c) - \mu(f \wedge c)| . \tag{8}$$

The sequence $\{\mu_n, n \geq 0\}$ is relatively compact and thus tight since X is Polish ((Billingsley, 1999, Theorem 5.2.)). Then, for $\varepsilon > 0$, there exists a compact set $\mathcal{K}$ such that $\sup_n \mu_n(\mathcal{K}^c) \leq \varepsilon$. Then

$$\mu_n(|f_n \wedge c - f \wedge c|) \leq \sup_{\mathcal{K}} |f_n\wedge c-f\wedge c|+2c\mu_n(\mathcal{K}^c) \leq \sup_{\mathcal{K}} |f_n\wedge c-f\wedge c|+2c\varepsilon .$$

Under the stated assumptions, the family of functions $\{\bar{f}_n \stackrel{\text{def}}{=} f_n\wedge c-f\wedge c, n \geq 0\}$ are equicontinuous, and $\lim_{n\to\infty} \bar{f}_n = 0$. By (Royden, 1988, Lemma 39, Chapter 7), the convergence is uniform on compact sets. Consider now the second term on the RHS of (8). Since $f\wedge c$ is a bounded continuous function and $\{\mu_n, n \geq 0\}$ converges weakly to μ, $\lim_{n\to\infty} \mu_n(f \wedge c) = \mu(f \wedge c)$. This concludes the proof. $\square$

5. Proof of (Fort, Moulines and Priouret, 2010, Proposition 5.2). Let $(\mathbb{U}, \delta)$ be a metric space; recall that for a real-valued function f on $\mathbb{U}$, the Lipschitz semi-norm is defined by

$$|f|_{\mathrm{Lip}(\mathbb{U},\delta)} \stackrel{\text{def}}{=} \sup_{x\neq y,(x,y)\in\mathbb{U}^2} \frac{|f(x) - f(y)|}{\delta(x, y)} .$$

Denote the supremum norm

$$\|f\|_{\infty,\mathbb{U}} \stackrel{\text{def}}{=} \sup_{x\in\mathbb{U}} |f(x)| .$$

Let $|f|_{\mathrm{BL}(\mathbb{U},\delta)} \stackrel{\text{def}}{=} |f|_{\mathrm{Lip}(\mathbb{U},\delta)} + \|f\|_{\infty,\mathbb{U}}$. Here, "BL" stands for "bounded Lipschitz" and $\mathrm{BL}(\mathbb{U}, \delta)$ is the set of all bounded real valued Lipschitz functions on $(\mathbb{U}, \delta)$.

Recall the following extension Theorem for bounded Lipschitz functions (Dudley, 2002, Proposition 11.2.3.)

THEOREM 5.1. *If* $\mathbb{U} \subset \mathbb{V}$ *and* $f \in \mathrm{BL}(\mathbb{U}, \delta)$ *then* f *can be extended to a function* $h \in \mathrm{BL}(\mathbb{V}, \delta)$ *with* $h = f$ *on* $\mathbb{V}$ *and* $|h|_{\mathrm{BL}(\mathbb{V},\delta)} = |f|_{\mathrm{BL}(\mathbb{U},\delta)}$.

Recall also that if $(\mathbb{U}, \delta)$ is compact, then the set of continuous functions coincides with the set of continuous bounded functions $\mathsf{C}_b(\mathbb{U}, \delta)$.

THEOREM 5.2. *If* $(\mathbb{V}, \delta)$ *is a compact metric space, the space of bounded Lipschitz functions* $\mathrm{BL}(\mathbb{V}, \delta)$ *equipped with the supremum norm* $\| \cdot \|_{\infty,\mathbb{V}}$ *is a separable space.*

PROOF. By (Dudley, 2002, Corollary 11.2.5), the space of bounded continuous function $\mathsf{C}_b(\mathbb{V}, \delta)$ equipped with the supremum norm $\| \cdot \|_{\infty,\mathbb{V}}$ is separable; denote by $\{f_i, i \in \mathbb{N}\} \subset \mathsf{C}_b(\mathbb{V}, \delta)$ a dense family of bounded functions in $\mathsf{C}_b(\mathbb{V}, \delta)$ for the supremum norm. Since, (Dudley, 2002, Proposition 11.2.4), $\mathrm{BL}(\mathbb{V}, \delta)$ is dense in $\mathsf{C}_b(\mathbb{V}, \delta)$ for the supremum norm $\| \cdot \|_{\infty,\mathbb{V}}$, for any $n \in \mathbb{N}$ and then any $p \in \mathbb{N}$, we may choose $f_{n,p} \in \mathrm{BL}(\mathbb{V}, \delta)$ such that $\|f_n - f_{n,p}\|_{\infty,\mathbb{V}} \leq 1/p$.

By construction, the countable family of function $\{f_{n,p}, (n,p) \in \mathbb{N}^2\}$ is dense in $\mathrm{BL}(\mathbb{V}, \delta)$ equipped with the supremum norm. □

Finally, the set of bounded Lipschitz functions is convergence determining (Dudley, 2002, Theorem 11.3.3)

THEOREM 5.3. *Let* μ *and* $\{\mu_n, n \geq 0\}$ *be distributions on a separable metric space* $(\mathbb{U}, \delta)$ *endowed with its Borel* σ*-field. The following properties are equivalent*

(a) $\{\mu_n, n \geq 0\}$ *converges weakly to* μ.
(b) $\{\mu_n(f), n \geq 0\}$ *converges to* $\mu(f)$ *for any function* $f \in \mathrm{BL}(\mathbb{U}, \delta)$.

LEMMA 5.4. *Let* $(\mathbb{U}, d)$ *be a separable metric space. There exists a metric* δ *on* $\mathbb{U}$ *defining the same topology as* d*, such that the space of bounded Lipschitz functions* $\mathrm{BL}(\mathbb{U}, \delta)$ *equipped with the supremum norm* $\| \cdot \|_{\infty,\mathbb{U}}$ *is separable.*

PROOF. By (Dudley, 2002, Theorem 2.8.2), there exist a space $\mathbb{V}$ and a metric δ on $\mathbb{V}$ such that $\mathbb{U} \subseteq \mathbb{V}$, $(\mathbb{V}, \delta)$ is a compact metric space and $\mathbb{U}$ is dense in $\mathbb{V}$ for the metric δ.

Denote by $\{\psi_k, k \in \mathbb{N}\} \subset \mathrm{BL}(\mathbb{V}, \delta)$ be a countable family of functions which is dense in $\mathrm{BL}(\mathbb{V}, \delta)$ equipped with the supremum norm; see Theorem 5.2. For any $k \in \mathbb{N}$, denote by ϕ_k the restriction of the function ψ_k to $\mathbb{U}$. Clearly, for any $k \in \mathbb{N}$, $\phi_k \in \mathrm{BL}(\mathbb{U}, \delta)$. By Theorem 5.1, for any $f \in \mathrm{BL}(\mathbb{U}, \delta)$, there is a function $h \in \mathrm{BL}(\mathbb{V}, \delta)$ with $f = h$ on $\mathbb{U}$ and $|h|_{\mathrm{BL}(\mathbb{V},\delta)} = |f|_{\mathrm{BL}(\mathbb{U},\delta)}$. Since for any $k \geq 0$,

$$\sup_{x \in \mathbb{U}} |f(x) - \phi_k(x)| \leq \sup_{x \in \mathbb{V}} |h(x) - \psi_k(x)|$$

the set $\{\phi_k, k \geq 0\}$ is dense in $\mathrm{BL}(\mathbb{U}, \delta)$ equipped with the supremum norm $\| \cdot \|_{\infty,\mathbb{U}}$. □

PROPOSITION (Proposition 5.2 in Fort, Moulines and Priouret (2010)). *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 $f \in \mathsf{C}_b(\mathbb{U}, d)$*

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

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

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

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

PROOF. By Lemma 5.4, there is a metric δ on $\mathbb{U}$ defining the same topology as d and a countable family $\{\phi_k, k \geq 0\} \subset \mathrm{BL}(\mathbb{U}, \delta)$ which is dense in the space $\mathrm{BL}(\mathbb{U}, \delta)$ equipped with supremum norm $\| \cdot \|_{\infty,\mathbb{U}}$. For any $\varepsilon > 0$ and any function $f \in \mathrm{BL}(\mathbb{U}, \delta)$, there exists $k \geq 0$ such that $\|f - \phi_k\|_\infty \leq \varepsilon$.

By Theorem 5.3, it suffices to show that there exists a $\mathbb{P}$-full set Ω_1 such that for any $\omega \in \Omega_1$ and any function $f \in \mathrm{BL}(\mathbb{U}, \delta)$, $\limsup_{n\to\infty} |\mu_n(\omega, f) - \mu(f)| = 0$. Set $\Omega_1 \stackrel{\text{def}}{=} \bigcap_k \Omega_{\phi_k}$. Since $\mathbb{P}(\Omega_{\phi_k}) = 1$, then $\mathbb{P}(\Omega_1) = 1$. For any $\omega \in \Omega_1$, we have

$$\begin{aligned}
&|\mu_n(\omega, f) - \mu(f)| \\
&\leq |\mu_n(\omega, f) - \mu_n(\phi_k)| + |\mu_n(\omega, \phi_k) - \mu(\phi_k)| + |\mu(\phi_k) - \mu(f)| \\
&\leq 2\|f - \phi_k\|_\infty + |\mu_n(\omega, \phi_k) - \mu(\phi_k)| \\
&\leq 2\varepsilon + |\mu_n(\omega, \phi_k) - \mu(\phi_k)| \ .
\end{aligned}$$

and this concludes the proof since by definition of Ω_1, $\lim_n \mu_n(\omega, \phi_k) = \mu(\phi_k)$. □

References.

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

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

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

BENVENISTE, A., MÉTIVIER, M. and PRIOURET, P. (1990). *Adaptive Algorithms and Stochastic Approximations* **22**. Springer, Berlin. Translated from the French by Stephen S. S. Wilson.

BILLINGSLEY, P. (1999). *Convergence of probability measures*, Second ed. *Wiley Series in Probability and Statistics: Probability and Statistics*. John Wiley & Sons Inc., New York. A Wiley-Interscience Publication.

DUDLEY, R. M. (2002). *Real Analysis and Probability*. Cambridge University Press.

FORT, G., MOULINES, E. and PRIOURET, P. (2010). Convergence of adaptive MCMC algorithms: ergodicity and law of large numbers Technical Report. submitted to Ann. Statist.

GLYNN, P. W. and MEYN, S. P. (1996). A Liapounov bound for solutions of the Poisson equation. *Ann. Probab.* **24** 916–931.

ROYDEN, H. L. (1988). *Real analysis*, Third ed. Macmillan Publishing Company, New York. MR1013117 (90g:00004)

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