



**HAL**  
open science

## Subgeometric ergodicity of Markov chains

Randal Douc, Eric Moulines, Philippe Soulier

► **To cite this version:**

Randal Douc, Eric Moulines, Philippe Soulier. Subgeometric ergodicity of Markov chains. Patrice Bertail, Paul Doukhan, Philippe Soulier. Dependence in probability and statistics, Springer, pp.55–64, 2006, Lecture notes in statistics. hal-00154180

**HAL Id: hal-00154180**

**<https://hal.science/hal-00154180>**

Submitted on 12 Jun 2007

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

---

## Subgeometric ergodicity of Markov chains

Randal Douc<sup>1</sup>, Eric Moulines<sup>2</sup>, and Philippe Soulier<sup>3</sup>

<sup>1</sup> CMAP, Ecole Polytechnique, 91128 Palaiseau Cedex, France  
 douc@cmap.polytechnique.fr

<sup>2</sup> Département TSI, Ecole nationale supérieure des Télécommunications, 46 rue  
 Barrault, 75013 Paris, France moulines@tsi.enst.fr

<sup>3</sup> Equipe MODAL'X, Université de Paris X Nanterre, 92000 Nanterre, France  
 philippe.soulier@u-paris10.fr

### 1 Introduction

Let  $P$  be a Markov transition kernel on a state space  $X$  equipped with a countably generated  $\sigma$ -field  $\mathcal{X}$ . For a control function  $f : X \rightarrow [1, \infty)$ , the  $f$ -total variation or  $f$ -norm of a signed measure  $\mu$  on  $\mathcal{X}$  is defined as

$$\|\mu\|_f := \sup_{|g| \leq f} |\mu(g)| .$$

When  $f \equiv 1$ , the  $f$ -norm is the total variation norm, which is denoted  $\|\mu\|_{TV}$ . Assume that  $P$  is aperiodic positive Harris recurrent with stationary distribution  $\pi$ . Then the iterated kernels  $P^n(x, \cdot)$  converge to  $\pi$ . The rate of convergence of  $P^n(x, \cdot)$  to  $\pi$  does not depend on the starting state  $x$ , but exact bounds may depend on  $x$ . Hence, it is of interest to obtain non uniform or quantitative bounds of the following form

$$\sum_{n=1}^{\infty} r(n) \|P^n(x, \cdot) - \pi\|_f \leq g(x) , \quad \text{for all } x \in X \quad (1)$$

where  $f$  is a control function,  $\{r(n)\}_{n \geq 0}$  is a non-decreasing sequence, and  $g$  is a nonnegative function which can be computed explicitly.

As emphasized in [RR04, section 3.5], quantitative bounds have a substantial history in Markov chain theory. Applications are numerous including convergence analysis of Markov Chain Monte Carlo (MCMC) methods, transient analysis of queueing systems or storage models, etc. With few exception however, these quantitative bounds were derived under conditions which imply geometric convergence, *i.e.*  $r(n) = \beta^n$ , for some  $\beta > 1$  (see for instance [MT94], [Ros95], [RT99], [RR04], and [Bax05]).

Geometric convergence does not hold for many chains of practical interest. Hence it is necessary to derive bounds for chains which converge to the

stationary distribution at a rate  $r$  which grows to infinity slower than a geometric sequence. These sequences are called subgeometric sequences and are defined in [NT83] as non decreasing sequences  $r$  such that  $\log r(n)/n \downarrow 0$  as  $n \rightarrow \infty$ . These sequences include among other examples the polynomial sequences  $r(n) = n^\gamma$  with  $\gamma > 0$  and subgeometric sequences  $r(n)e^{cn^\delta}$  with  $c > 0$  and  $\delta \in (0, 1)$ .

The first general results proving subgeometric rates of convergence were obtained by [NT83] and later extended by [TT94], but do not provide computable expressions for the bound in the rhs of (1). A direct route to quantitative bounds for subgeometric sequences has been opened by [Ver97, Ver99], based on coupling techniques. Such techniques were later used in specific contexts by many authors, among others, [FM00] [JR01] [For01] [FM03b].

The goal of this paper is to give a short and self contained proof of general bounds for subgeometric rates of convergence, under practical conditions. This is done in two steps. The first one is Theorem 1 whose proof, based on coupling, provides an intuitive understanding of the results of [NT83] and [TT94]. The second step is the use of a very general drift condition, recently introduced in [DFMS04]. This condition is recalled in Section 2.1 and the bounds it implied are stated in Proposition 1.

This paper complements the works [DFMS04] and [DMS05], to which we refer for applications of the present techniques to practical examples.

## 2 Explicit bounds for the rate of convergence

The only assumption for our main result is the existence of a small set.

- (A1). There exist a set  $C \in \mathcal{X}$ , a constant  $\epsilon > 0$  and a probability measure  $\nu$  such that, for all  $x \in C$ ,  $P(x, \cdot) \geq \epsilon\nu(\cdot)$ .

For simplicity, only one-step minorisation is considered in this paper. Adaptations to  $m$ -step minorisation can be carried out as in [Ros95] (see also [For01] and [FM03b]).

Let  $\check{P}$  be a Markov transition kernel on  $\mathsf{X} \times \mathsf{X}$  such that, for all  $A \in \mathcal{X}$ ,

$$\check{P}(x, x', A \times \mathsf{X}) = P(x, A)\mathbb{1}_{(C \times C)^c}(x, x') + Q(x, A)\mathbb{1}_{C \times C}(x, x') \quad (2)$$

$$\check{P}(x, x', \mathsf{X} \times A) = P(x', A)\mathbb{1}_{(C \times C)^c}(x, x') + Q(x', A)\mathbb{1}_{C \times C}(x, x') \quad (3)$$

where  $A^c$  denotes the complementary of the subset  $A$  and  $Q$  is the so-called residual kernel defined, for  $x \in C$  and  $A \in \mathcal{X}$  by

$$Q(x, A) = \begin{cases} (1 - \epsilon)^{-1} (P(x, A) - \epsilon\nu(A)) & 0 < \epsilon < 1 \\ \nu(A) & \epsilon = 1 \end{cases} \quad (4)$$

One may for example set

$$\begin{aligned} \check{P}(x, x'; A \times A') = \\ P(x, A)P(x', A')\mathbb{1}_{(C \times C)^c}(x, x') + Q(x, A)Q(x', A)\mathbb{1}_{C \times C}(x, x'), \end{aligned} \quad (5)$$

but this choice is not always the most suitable; cf. Section 2.2. For  $(x, x') \in \mathsf{X} \times \mathsf{X}$ , denote by  $\check{\mathbb{P}}_{x, x'}$  and  $\check{\mathbb{E}}_{x, x'}$  the law and the expectation of a Markov chain with initial distribution  $\delta_x \otimes \delta_{x'}$  and transition kernel  $\check{P}$ .

**Theorem 1.** *Assume (A1).*

For any sequence  $r \in \Lambda$ ,  $\delta > 0$  and all  $(x, x') \in \mathsf{X} \times \mathsf{X}$ ,

$$\sum_{n=1}^{\infty} r(n) \|P^n(x, \cdot) - P^n(x', \cdot)\|_{\text{TV}} \leq (1 + \delta) \check{\mathbb{E}}_{x, x'} \left[ \sum_{k=0}^{\sigma} r(k) \right] + \frac{1 - \epsilon}{\epsilon} M, \quad (6)$$

with  $M = (1 + \delta) \sup_{n \geq 0} \{R^* r(n - 1) - \epsilon(1 - \epsilon) \delta R(n) / (1 + \delta)\}_+$  and  $R^* = \sup_{(y, y') \in C \times C} \check{\mathbb{E}}_{y, y'} [\sum_{k=1}^{\tau} r(k)]$ .

Let  $W : \mathsf{X} \times \mathsf{X} \rightarrow [1, \infty)$  and  $f$  be a non-negative function  $f$  such that  $f(x) + f(x') \leq W(x, x')$  for all  $(x, x') \in \mathsf{X} \times \mathsf{X}$ . Then,

$$\sum_{n=1}^{\infty} \|P^n(x, \cdot) - P^n(x', \cdot)\|_f \leq \check{\mathbb{E}}_{x, x'} \left[ \sum_{k=0}^{\sigma} W(X_k, X'_k) \right] + \frac{1 - \epsilon}{\epsilon} W^*. \quad (7)$$

with  $W^* = \sup_{(y, y') \in C \times C} \check{\mathbb{E}}_{y, y'} [\sum_{k=1}^{\tau} W(X_k, X'_k)]$ .

*Remark 1.* Integrating these bounds with respect to  $\pi(dx')$  yields similar bounds for  $\|P^n(x, \cdot) - \pi\|_{\text{TV}}$  and  $\|P^n(x, \cdot) - \pi\|_f$ .

*Remark 2.* The trade off between the size of the coupling set and the constant  $\epsilon$  appears clearly: if the small set is big, then the chain returns more often to the small set and the moments of the hitting times can be expected to be smaller, but the constant  $\epsilon$  will be smaller. This trade-off is illustrated numerically in [DMS05, Section 3].

By interpolation, intermediate rates of convergence can be obtained. Let  $\alpha$  and  $\beta$  be positive and increasing functions such that, for some  $0 \leq \rho \leq 1$ ,

$$\alpha(u)\beta(v) \leq \rho u + (1 - \rho)v, \quad \text{for all } (u, v) \in \mathbb{R}^+ \times \mathbb{R}^+. \quad (8)$$

Functions satisfying this condition can be obtained from Young's inequality. Let  $\psi$  be a real valued, continuous, strictly increasing function on  $\mathbb{R}^+$  such that  $\psi(0) = 0$ ; then for all  $a, b > 0$ ,

$$ab \leq \Psi(a) + \Phi(b), \quad \text{where } \Psi(a) = \int_0^a \psi(x) dx \quad \text{and} \quad \Phi(b) = \int_0^b \psi^{-1}(x) dx,$$

where  $\psi^{-1}$  is the inverse function of  $\psi$ . If we set  $\alpha(u) = \Psi^{-1}(\rho u)$  and  $\beta(v) = \Phi^{-1}((1 - \rho)v)$ , then the pair  $(\alpha, \beta)$  satisfies (8). A trivial example is obtained by taking  $\psi(x) = x^{p-1}$  for some  $p \geq 1$ , which yields  $\alpha(u) = (p\rho u)^{1/p}$  and  $\beta(v) = (p(1 - \rho)v)^{1/p}$ . Other examples are given in Section 2.1.

**Corollary 1.** *Let  $\alpha$  and  $\beta$  be two positive functions satisfying (8) for some  $0 \leq \rho \leq 1$ . Then, for any non-negative function  $f$  such that  $f(x) + f(x') \leq \beta \circ W(x, x')$  and  $\delta > 0$ , for all  $x, x' \in \mathbf{X}$  and  $n \geq 1$ ,*

$$\begin{aligned} \sum_{n=1}^{\infty} \alpha(r(n)) \|P^n(x, \cdot) - P^n(x', \cdot)\|_f &\leq \rho(1 + \delta) \check{\mathbb{E}}_{x, x'} \left[ \sum_{k=0}^{\sigma} r(k) \right] \\ &+ (1 - \rho) \check{\mathbb{E}}_{x, x'} \left[ \sum_{k=0}^{\sigma} W(X_k, X'_k) \right] \frac{1 - \epsilon}{\epsilon} \{ \rho M + (1 - \rho) W^* \}. \end{aligned} \quad (9)$$

## 2.1 Drift Conditions for subgeometric ergodicity

The bounds obtained in Theorem 1 and Corollary 1 are meaningful only if they are finite. Sufficient conditions are given in this section in the form of drift conditions. The most well known drift condition is the so-called Foster-Lyapounov drift condition which not only implies but is actually equivalent to geometric convergence to the stationary distribution, cf. [MT93, Chapter 16]. [JR01], simplifying and generalizing an argument in [FM00], introduced a drift condition which implies polynomial rates of convergence. We consider here the following drift condition, introduced in [DFMS04], which allows to bridge the gap between polynomial and geometric rates of convergence.

**Condition  $\mathbf{D}(\phi, V, C)$ :** There exist a function  $V : \mathbf{X} \rightarrow [1, \infty]$ , a concave monotone non decreasing differentiable function  $\phi : [1, \infty] \mapsto (0, \infty]$ , a measurable set  $C$  and a constant  $b > 0$  such that

$$PV + \phi \circ V \leq V + b \mathbf{1}_C.$$

If the function  $\phi$  is concave, non decreasing and differentiable, define

$$H_\phi(v) := \int_1^v \frac{dx}{\phi(x)}. \quad (10)$$

Then  $H_\phi$  is a non decreasing concave differentiable function on  $[1, \infty)$ . Moreover, since  $\phi$  is concave,  $\phi'$  is non increasing. Hence  $\phi(v) \leq \phi(1) + \phi'(1)(v - 1)$  for all  $v \geq 1$ , which implies that  $H_\phi$  increases to infinity. We can thus define its inverse  $H_\phi^{-1} : [0, \infty) \rightarrow [1, \infty)$ , which is also an increasing and differentiable function, with derivative  $(H_\phi^{-1})'(x) = \phi \circ H_\phi^{-1}(x)$ . For  $k \in \mathbb{N}$ ,  $z \geq 0$  and  $v \geq 1$ , define

$$r_\phi(z) := (H_\phi^{-1})'(z) = \phi \circ H_\phi^{-1}(z). \quad (11)$$

It is readily checked that if  $\lim_{t \rightarrow \infty} \phi'(t) = 0$ , then  $r_\phi \in \Lambda$ , cf [DFMS04, Lemma 2.3].

Proposition 2.2 and Theorem 2.3 in [DMS05] show that the drift condition  $\mathbf{D}(\phi, V, C)$  implies that the bounds of Theorem 1 are finite. We gather here these results.

**Proposition 1.** *Assume that Condition  $\mathbf{D}(\phi, V, C)$  holds for some small set  $C$  and that  $\inf_{x \notin C} \phi \circ V(x) > b$ . Fix some arbitrary  $\lambda \in (0, 1 - b/\inf_{x \notin C} \phi \circ V(x))$  and define  $W(x, x') = \lambda\phi(V(x) + V(x') - 1)$ . Define also  $V^* = (1 - \epsilon)^{-1} \sup_{y \in C} \{PV(y) - \epsilon\nu(V)\}$ . Let  $\sigma$  be the hitting time of the set  $C \times C$ . Then*

$$\begin{aligned} \tilde{\mathbb{E}}_{x, x'} \left[ \sum_{k=0}^{\sigma} r_{\phi}(k) \right] &\leq 1 + \frac{r_{\phi}(1)}{\phi(1)} \{V(x) + V(x')\} \mathbb{1}_{(x, x') \notin C \times C}, \\ \tilde{\mathbb{E}}_{x, x'} \left[ \sum_{k=0}^{\sigma} W(X_k, X'_k) \right] &\leq \sup_{(y, y') \in C \times C} W(y, y') + \{V(x) + V(x')\} \mathbb{1}_{(x, x') \notin C \times C}, \\ R^* &\leq 1 + \frac{r_{\phi}(1)}{\phi(1)} \{2V^* - 1\} \\ W^* &\leq \sup_{(y, y') \in C \times C} W(y, y') + 2V^* - 1. \end{aligned}$$

*Remark 3.* The condition  $\inf_{y \notin C} \phi \circ V(y) > b$  may not be fulfilled. If level sets  $\{V \leq d\}$  are small, then the set  $C$  can be enlarged so that this condition holds. This additional condition may appear rather strong, but can be weakened by using small sets associated to some iterate  $P^m$  of the kernel (see e.g. [Ros95], [For01] and [FM03b]).

We now give examples of rates that can be obtained by (11).

#### *Polynomial rates*

Polynomial rates of convergence are obtained when Condition  $\mathbf{D}(\phi, V, C)$  holds with  $\phi(v) = cv^{\alpha}$  for some  $\alpha \in [0, 1)$  and  $c \in (0, 1]$ . The rate of convergence in total variation distance is  $r_{\phi}(n) \propto n^{\alpha/(1-\alpha)}$  and the pairs  $(r, f)$  for which (9) holds are of the form  $(n^{(1-p)\alpha/(1-\alpha)}, V^{\alpha p})$  for  $p \in [0, 1]$ , or in other terms,  $(n^{\kappa-1}, V^{1-\kappa(1-\alpha)})$  for  $1 \leq \kappa \leq 1/(1-\alpha)$ , which is Theorem 3.6 of [JR01].

It is possible to extend this result by using more general interpolation functions. For instance, choosing for  $b > 0$ ,  $\alpha(x) = (1 \vee \log(x))^b$  and  $\beta(x) = x(1 \vee \log(x))^{-b}$  yields the pairs  $(n^{(1-p)\alpha/(1-\alpha)} \log^b(n), V^{\alpha p} (1 + \log V)^{-b})$ , for  $p \in [0, 1]$ .

#### *Logarithmic rates of convergence*

Rates of convergence slower than any polynomial can be obtained when condition  $\mathbf{D}(\phi, V, C)$  holds with a function  $\phi$  that increases to infinity slower than polynomially, for instance  $\phi(v) = c(1 + \log(v))^{\alpha}$  for some  $\alpha \geq 0$  and  $c \in (0, 1]$ . A straightforward calculation shows that

$$r_{\phi}(n) \asymp \log^{\alpha}(n).$$

Pairs for which (9) holds are thus of the form  $((1 + \log(n))^{(1-p)\alpha}, (1 + \log(V))^{p\alpha})$ .

*Subexponential rates of convergence*

Subexponential rates of convergence faster than any polynomial are obtained when the condition  $\mathbf{D}(\phi, V, C)$  holds with  $\phi$  such that  $v/\phi(v)$  goes to infinity slower than polynomially. Assume for instance that  $\phi$  is concave and differentiable on  $[1, +\infty)$  and that for large  $v$ ,  $\phi(v) = cv/\log^\alpha(v)$  for some  $\alpha > 0$  and  $c > 0$ . A simple calculation yields

$$r_\phi(n) \asymp n^{-\alpha/(1+\alpha)} \exp\left(\{c(1+\alpha)n\}^{1/(1+\alpha)}\right).$$

Choosing  $\alpha(x) = x^{1-p}(1 \vee \log(x))^{-b}$  and  $\beta(x) = x^p(1 \vee \log(x))^b$  for  $p \in (0, 1)$  and  $b \in \mathbb{R}$ ; or  $p = 0$  and  $b > 0$ ; or  $p = 1$  and  $b < -\alpha$  yields the pairs

$$n^{-(\alpha+b)/(1+\alpha)} \exp\left((1-p)\{c(1+\alpha)n\}^{1/(1+\alpha)}\right), V^p(1 + \log V)^b.$$

**2.2 Stochastically monotone chains**

Let  $\mathbf{X}$  be a totally ordered set and let the order relation be denoted by  $\preceq$  and for  $a \in \mathbf{X}$ , let  $(-\infty, a]$  denote the set of all  $x \in \mathbf{X}$  such that  $x \preceq a$ . A transition kernel on  $\mathbf{X}$  is said to be stochastically monotone if  $x \preceq y$  implies  $P(x, (-\infty, a]) \geq P(y, (-\infty, a])$  for all  $a \in \mathbf{X}$ . If Assumption (A1) holds, for a small set  $C = (-\infty, a_0]$ , then instead of defining the kernel  $\check{P}$  as in (5), it is convenient to define it, for  $x, x' \in \mathbf{X}$  and  $A \in \mathcal{X} \otimes \mathcal{X}$ , by

$$\begin{aligned} \check{P}(x, x'; A) &= \mathbb{1}_{(x, x') \notin C \times C} \int_0^1 \mathbb{1}_A(P^\leftarrow(x, u), P^\leftarrow(x', u)) du \\ &\quad + \mathbb{1}_{C \times C}(x, x') \int_0^1 \mathbb{1}_A(Q^\leftarrow(x, u), Q^\leftarrow(x', u)) du, \end{aligned}$$

where, for any transition kernel  $K$  on  $\mathbf{X}$ ,  $K^\leftarrow(x, \cdot)$  is the quantile function of the probability measure  $K(x, \cdot)$ , and  $Q$  is the residual kernel defined in (4). This construction makes the set  $\{(x, x') \in \mathbf{X} \times \mathbf{X} : x \preceq x'\}$  absorbing for  $\check{P}$ . This means that if the chain  $(X_n, X'_n)$  starts at  $(x_0, x'_0)$  with  $x_0 \preceq x'_0$ , then almost surely,  $X_n \preceq X'_n$  for all  $n$ . Let now  $\sigma_C$  and  $\sigma_{C \times C}$  denote the hitting times of the sets  $C$  and  $C \times C$ , respectively. Then, we have the following very simple relations between the moments of the hitting times of the one dimensional chain and that of the bidimensional chain with transition kernel  $\check{P}$ . For any sequence  $r$  and any non negative function  $V$  all  $x \preceq x'$

$$\check{\mathbb{E}}_{x, x'} \left[ \sum_{k=0}^{\sigma_{C \times C}} r(k) V(X_k, X'_k) \right] \leq \mathbb{E}_{x'} \left[ \sum_{k=0}^{\sigma_C} r(k) V(X'_k) \right].$$

A similar bound obviously holds for the return times. Thus, there only remain to obtain bounds for this quantities, which is very straightforward if moreover condition  $\mathbf{D}(\phi, \mathbf{V}, \mathbf{C})$  holds. Examples of stochastically monotone chains with applications to queuing and Monte-Carlo simulation that satisfy condition  $\mathbf{D}(\phi, \mathbf{V}, \mathbf{C})$  are given in [DMS05, section 3].

### 3 Proof of Theorem 1

Define a transition kernel  $\tilde{P}$  on the space  $\tilde{X} = \mathsf{X} \times \mathsf{X} \times \{0, 1\}$  endowed with the product  $\sigma$ -field  $\tilde{\mathcal{X}}$ , for any  $x, x' \in \mathsf{X}$  and  $A, A' \in \mathcal{X}$ , by

$$\tilde{P}((x, x', 0), A \times A' \times \{0\}) = \{1 - \epsilon \mathbb{1}_{C \times C}(x, x')\} \tilde{P}((x, x'), A \times A'), \quad (12)$$

$$\tilde{P}((x, x', 0), A \times A' \times \{1\}) = \epsilon \mathbb{1}_{C \times C}(x, x') \nu_{x, x'}(A \cap A'), \quad (13)$$

$$\tilde{P}((x, x', 1), A \times A' \times \{1\}) = P(x, A \cap A'). \quad (14)$$

For any probability measure  $\tilde{\mu}$  on  $(\tilde{X}, \tilde{\mathcal{X}})$ , let  $\tilde{\mathbb{P}}_{\tilde{\mu}}$  be the probability measure on the canonical space  $(\tilde{X}^{\mathbb{N}}, \tilde{\mathcal{X}}^{\otimes \mathbb{N}})$  such that the coordinate process  $\{\tilde{X}_k\}$  is a Markov chain with transition kernel  $\tilde{P}$  and initial distribution  $\tilde{\mu}$ . The corresponding expectation operator is denoted by  $\tilde{\mathbb{E}}_{\tilde{\mu}}$ .

The transition kernel  $\tilde{P}$  can be described algorithmically. Given  $\tilde{X}_0 = (X_0, X'_0, d_0) = (x, x', d)$ ,  $\tilde{X}_1 = (X_1, X'_1, d_1)$  is obtained as follows.

- If  $d = 1$  then draw  $X_1$  from  $P(x, \cdot)$  and set  $X'_1 = X_1$ ,  $d_1 = 1$ .
- If  $d = 0$  and  $(x, x') \in C \times C$ , flip a coin with probability of heads  $\epsilon$ .
  - If the coin comes up heads, draw  $X_1$  from  $\nu_{x, x'}$  and set  $X'_1 = X_1$  and  $d_1 = 1$ .
  - If the coin comes up tails, draw  $(X_1, X'_1)$  from  $\tilde{P}(x, x'; \cdot)$  and set  $d_1 = 0$ .
- If  $d = 0$  and  $(x, x') \notin C \times C$ , draw  $(X_1, X'_1)$  from  $\tilde{P}(x, x'; \cdot)$  and set  $d_1 = 0$ .

The variable  $d_n$  is called the *bell variable*; it indicates whether coupling has occurred by time  $n$  ( $d_n = 1$ ) or not ( $d_n = 0$ ). The first index  $n$  at which  $d_n = 1$  is the coupling time;

$$T = \inf\{k \geq 1 : d_k = 1\}.$$

If  $d_n = 1$  then  $X_k = X'_k$  for all  $k \geq n$ . This coupling construction is carried out in such a way that under  $\tilde{\mathbb{P}}_{\xi \otimes \xi' \otimes \delta_0}$ ,  $\{X_k\}$  and  $\{X'_k\}$  are Markov chains with transition kernel  $P$  with initial distributions  $\xi$  and  $\xi'$  respectively.

The main tool of the proof is the following relation between  $\tilde{\mathbb{E}}_{x, x', 0}$  and  $\tilde{\mathbb{E}}_{x, x'}$ , proved in [DMR04, Lemma 1]. For any non-negative adapted process  $(\chi_k)_{k \geq 0}$  and  $(x, x') \in \mathsf{X} \times \mathsf{X}$ ,

$$\tilde{\mathbb{E}}_{x, x', 0}[\chi_n \mathbb{1}_{\{T > n\}}] = \tilde{\mathbb{E}}_{x, x'}[\chi_n (1 - \epsilon)^{N_n - 1}], \quad (15)$$

where  $N_n = \sum_{i=0}^n \mathbb{1}_{C \times C}(X_i, X'_i)$  is the number of visits to  $C \times C$  before time  $n$ .

We now proceed with the proof of Theorem 1.

**Step 1** *Lindvall's inequality* [Lin79, Lin92]

$$\sum_{k=0}^{\infty} r(k) \|P^k(x, \cdot) - P^k(x', \cdot)\|_f \leq \tilde{\mathbb{E}}_{x, x', 0} \left[ \sum_{j=0}^{T-1} r(j) \{f(X_j) + f(X'_j)\} \right]. \quad (16)$$



*Proof.* For any measurable function  $\phi$  such that  $|\phi| \leq f$ , and for any  $(x, x') \in \mathsf{X} \times \mathsf{X}$  it holds that

$$\begin{aligned} |P^k \phi(x) - P^k \phi(x')| &= \left| \tilde{\mathbb{E}}_{x, x', 0}[\{\phi(X_k) - \phi(X'_k)\} \mathbb{1}_{\{d_k=0\}}] \right| \\ &\leq \tilde{\mathbb{E}}_{x, x', 0}[\{f(X_k) + f(X'_k)\} \mathbb{1}_{\{T > k\}}]. \end{aligned}$$

Hence  $\|P^k(x, \dots) - P^k(x', \dots)\|_f \leq \tilde{\mathbb{E}}_{x, x', 0}[\{f(X_k) + f(X'_k)\} \mathbb{1}_{\{T > k\}}]$ . Summing over  $k$  yields (16).  $\square$

**Step 2** Denote  $W_{r, f}(x, x') = \tilde{\mathbb{E}}_{x, x', 0}[\sum_{k=0}^{\sigma} r(k)f(X_k, X'_k)]$  and  $W^*(r, f) = \sup_{(x, x') \in C \times C} [\sum_{k=1}^{\tau} r(k)f(X_k, X'_k)]/r(0)$ . Then

$$\begin{aligned} \tilde{\mathbb{E}}_{x, x', 0} \left[ \sum_{k=0}^{T-1} r(k)f(X_k, X'_k) \right] \\ \leq W_{r, f}(x, x') + \epsilon^{-1}(1 - \epsilon) W_{r, f}^* \tilde{\mathbb{E}}_{x, x', 0}[r(T - 1)]. \end{aligned} \quad (17)$$

*Proof.* Applying (15), we obtain

$$\begin{aligned} \tilde{\mathbb{E}}_{x, x', 0} \left[ \sum_{k=0}^{T-1} r(k)f(X_k, X'_k) \right] &= \sum_{k=0}^{\infty} \tilde{\mathbb{E}}_{x, x', 0} [r(k)f(X_k, X'_k) \mathbb{1}_{\{T > k\}}] \\ &= \sum_{k=0}^{\infty} \tilde{\mathbb{E}}_{x, x'} [r(k)f(X_k, X'_k)(1 - \epsilon)^{N_{k-1}}] \\ &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} (1 - \epsilon)^j \tilde{\mathbb{E}}_{x, x'} [r(k)f(X_k, X'_k) \mathbb{1}_{\{N_{k-1}=j\}}] \\ &= W_{r, f}(x, x') + \sum_{j=1}^{\infty} \sum_{k=0}^{\infty} (1 - \epsilon)^j \tilde{\mathbb{E}}_{x, x'} [r(k)f(X_k, X'_k) \mathbb{1}_{\{N_{k-1}=j\}}] \end{aligned}$$

For  $j \geq 0$ , let  $\sigma_j$  denote the  $(j + 1)$ -th visit to  $C \times C$ . Then  $N_{k-1} = j$  iff  $\sigma_{j-1} < k \leq \sigma_j$ . Since  $r$  is a subgeometric sequence,  $r(n+m) \leq r(n)r(m)/r(0)$ , thus

$$\begin{aligned} \sum_{k=0}^{\infty} r(k)f(X_k, X'_k) \mathbb{1}_{\{N_{k-1}=j\}} &= \sum_{k=\sigma_{j-1}+1}^{\sigma_j} r(k)f(X_k, X'_k) \\ &= \sum_{k=1}^{\tau \circ \theta^{\sigma_{j-1}}} r(\sigma_{j-1} + k)f(X_k, X'_k) \\ &\leq \frac{r(\sigma_{j-1})}{r(0)} \left( \sum_{k=1}^{\tau \circ \theta^{\sigma_{j-1}}} r(k)f(X_k, X'_k) \right) \circ \theta^{\sigma_{j-1}}. \end{aligned}$$

Applying the strong Markov property yields

$$\begin{aligned} \tilde{\mathbb{E}}_{x,x',0} \left[ \sum_{k=0}^{T-1} r(k) f(X_k, X'_k) \right] &\leq W_{r,f}(x, x') \\ &+ (1 - \epsilon) W^*(f, g) \sum_{j=0}^{\infty} (1 - \epsilon)^j \tilde{\mathbb{E}}_{x,x'}[r(\sigma_j)] . \end{aligned}$$

By similar calculations, (15) yields

$$\tilde{\mathbb{E}}[r(T-1)] = \epsilon \sum_{j=0}^{\infty} (1 - \epsilon)^j \tilde{\mathbb{E}}[r(\sigma_j)] ,$$

which concludes the proof of (17).  $\square$

**Step 3** Applying (17) with  $r \equiv 1$  yields (7).

**Step 4** If  $r \in \Lambda$ , then  $\lim_{n \rightarrow \infty} r(n)/R(n) = 0$ , with  $R(0) = 1$  and  $R(n) = \sum_{k=0}^{n-1} r(k)$ ,  $n \geq 1$ . Thus we can define, for  $r \in \Lambda$  and  $\delta > 0$

$$M_\delta = (1 + \delta) \sup_{n \geq 0} \left\{ \epsilon^{-1} (1 - \epsilon) W_{r,1}^* r(n-1) - \delta R(n) / (1 + \delta) \right\}_+ .$$

$M_\delta$  is finite for all  $\delta > 0$ . This yields

$$\tilde{\mathbb{E}}_{x,x',0}[R(T)] \leq (1 + \delta) W_{r,1}(x, x') + M_\delta .$$

Applying this bound with (16) yields (6).  $\square$

## References

- Bax05. Peter H. Baxendale. Renewal theory and computable convergence rates for geometrically ergodic markov chains. *Annals of Applied Probability*, 15(1B):700–738, 2005.
- DFMS04. Randal Douc, Gersende Fort, Eric Moulines, and Philippe Soulier. Practical drift conditions for subgeometric rates of convergence. *Annals of Applied Probability*, 14(3):1353–1377, 2004.
- DMR04. Randal Douc, Eric Moulines, and Jeff Rosenthal. Quantitative bounds for geometric convergence rates of Markov chains. *Annals of Applied Probability*, 14(4):1643–1665, 2004.
- DMS05. Randal Douc, Eric Moulines, and Philippe Soulier. Computable Convergence Rates for Subgeometrically Ergodic Markov Chains Preprint, 2005.
- FM00. Gersende Fort and Eric Moulines.  $V$ -subgeometric ergodicity for a Hastings-Metropolis algorithm. *Statistics and Probability Letters* 49(4):401–410, 2000.
- FM03a. Gersende Fort and Eric Moulines. Convergence of the monte carlo expectation maximization for curved exponential families. *Ann. Statist.*, 31(4):1220–1259, 2003.

- FM03b. Gersende Fort and Eric Moulines. Polynomial ergodicity of Markov transition kernels,. *Stochastic Processes and Their Applications*, 103:57–99, 2003.
- For01. Gersende Fort. *Contrôle explicite d'ergodicité de chaînes de Markov: applications e de convergence de l'algorithme Monte-Carlo EM*. PhD thesis, Université de Paris VI, 2001.
- JR01. Soren Jarner and Gareth Roberts. Polynomial convergence rates of Markov chains. *Annals of Applied Probability*, 12(1):224–247, 2001.
- Lin79. Torgny Lindvall. On coupling of discrete renewal sequences. *Z. Wahrsch. Verw. Gebiete*, 48(1):57–70, 1979.
- Lin92. Torgny Lindvall. *Lectures on the Coupling Method*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New-York, 1992.
- MT93. Sean P. Meyn and Robert L. Tweedie. Markov chains and stochastic stability. Communications and Control Engineering Series. Springer-Verlag London, 1993.
- MT94. Sean P. Meyn and Robert L. Tweedie. Computable bounds for convergence rates of Markov chains. *Annals of Applied Probability*, 4:981–1011, 1994.
- NT83. Esa Nummelin and Pekka Tuominen. The rate of convergence in Orey's theorem for Harris recurrent Markov chains with applications to renewal theory. *Stochastic Processes and Their Applications*, 15:295–311, 1983.
- Ros95. Jeffrey S. Rosenthal. Minorization conditions and convergence rates for Markov chain Monte Carlo. *Journal American Statistical Association*, 90:558–566, 1995.
- RR04. Gareth O. Roberts and Jeffrey S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probability Surveys*, 1:20–71, 2004
- RT99. Gareth O. Roberts and Richard L. Tweedie. Bounds on regeneration times and convergence rates for Markov chains. *Stochastic Processes and Their Applications*, 80:211–229, 1999.
- TT94. Pekka Tuominen and Richard Tweedie. Subgeometric rates of convergence of  $f$ -ergodic Markov Chains. *Advances in Applied Probability*, 26:775–798, 1994.
- Ver97. Alexander Veretennikov. On polynomial mixing bounds for stochastic differential equations. *Stochastic Process. Appl.*, 70:115–127, 1997.
- Ver99. Alexander Veretennikov. On polynomial mixing and the rate of convergence for stochastic differential and difference equations. *Theory of probability and its applications*, pages 361–374, 1999.