

# QUANTITATIVE CONVERGENCE RATES FOR SUBGEOMETRIC MARKOV CHAINS

CHRISTOPHE ANDRIEU,\* *University of Bristol*

GERSENDE FORT,\*\* *CNRS and Télécom ParisTech*

MATTI VIHOLA,\*\*\* *University of Jyväskylä*

## Abstract

We provide explicit expressions for the constants involved in the characterisation of ergodicity of subgeometric Markov chains. The constants are determined in terms of those appearing in the assumed drift and one-step minorisation conditions. The results are fundamental for the study of some algorithms where uniform bounds for these constants are needed for a family of Markov kernels. Our results accommodate also some classes of inhomogeneous chains.

*Keywords:* Markov chain; inhomogeneous; subgeometric ergodicity; polynomial ergodicity

2010 Mathematics Subject Classification: Primary 60J05  
Secondary 60J22

## 1. Introduction

Quantitative convergence rates of Markov chains have been extensively studied in the geometric ergodicity scenario; see, for example, [8] and [9] and the references therein for homogeneous and inhomogeneous Markov chains, respectively. Such results have proved to be very useful in certain applications, such as the analysis of adaptive Markov chain Monte Carlo (MCMC) or stochastic approximation (SA) recursions (see, for example, [3], [6], and [14]), where quantifying the convergence rates of a family of Markov kernels  $\{P_\theta\}_{\theta \in \Theta}$  in terms of  $\theta \in \Theta$  is required. In some cases, delicate control of the constants can also be used to deduce the stability of the algorithms; see, for example, [4] and [19].

In this paper we establish explicit bounds on the rate of convergence of subgeometric Markov chains in terms of the constants involved in standard drift and minorisation conditions. As in the geometric context, such results are important for adaptive MCMC and SA with subgeometric kernels; see, for example, [7]. In Section 4 we discuss in more detail two specific applications prompted by two other recent methodological and theoretical developments in the area of MCMC; see [4] and [5].

---

Received 11 September 2013; revision received 18 March 2014.

\* Postal address: School of Mathematics, University of Bristol, Bristol BS8 1TW, UK.

\*\* Postal address: Télécom ParisTech, 46 rue Barrault, 75634 Paris Cedex 13, France.

\*\*\* Postal address: Department of Mathematics and Statistics, University of Jyväskylä, PO Box 35, Jyväskylä, FI-40014, Finland. Email address: matti.s.vihola@jyu.fi

We now provide a brief discussion of the existing results and see how they relate to our work. Hereafter, we shall use the following standard notation whenever well-defined:

$$Pf(x) := \int P(x, dy)f(y), \quad \mu(f) := \int \mu(dx)f(x),$$

$$PQ(x, A) := \int P(x, dy)Q(y, A), \quad \mu P(A) := \int \mu(dx)P(x, A),$$

where  $P$  and  $Q$  are Markov kernels on a measurable space  $(X, \mathcal{B}(X))$ ,  $f: X \rightarrow \mathbb{R}$  is a measurable function and  $\mu$  is a (signed) measure.

In the literature, the Markov chain ‘convergence rate’ often refers to the rate of convergence of marginal distributions, that is, if  $\pi$  is the invariant measure of  $P$ ,

$$\hat{r}(n)|P^n f(x) - \pi(f)| \leq cV(x) \quad \text{for all } n \in \mathbb{N}, x \in X, \quad (1)$$

where  $(\hat{r}(n))_{n \geq 0}$  is a positive nondecreasing rate sequence,  $f$  belongs to a suitable class of functions integrable with respect to  $\pi$ , the function  $V: X \rightarrow [1, \infty)$  is measurable and  $c$  is a finite constant which is often left unspecified. Here, we focus instead on establishing the stronger property

$$\sum_{n=0}^{\infty} r(n)|P^n f(x) - \pi(f)| \leq cV(x) \quad \text{for all } x \in X, \quad (2)$$

and aim to quantify the constant  $c$  in terms of the constants in Condition 1. The rate  $(r(n))_{n \geq 0}$  is positive nondecreasing as  $(\hat{r}(n))_{n \geq 0}$ , and if  $r(n) = \hat{r}(n)$ , (2) clearly implies (1). While the distinction between (1) and (2) is often not essential in the geometric case, it turns out to be important in some subgeometric scenarios. Indeed, for some applications, using the marginal convergence rate (1) to deduce a property of the type in (2) may be suboptimal for subgeometric Markov chains; an example is briefly discussed below.

The characterisation of subgeometric Markov chains with drift and minorisation conditions has been considered in various earlier works starting with the pioneering work of Tuominen and Tweedie [20]. In the more recent works Fort and Moulines [13] and Jarner and Roberts [15] establish polynomial rates of convergence, but do not provide quantitative results. Douc *et al.* [11] (see also [12]) have extended these results to more general subgeometric ergodicity scenarios. The latter works consider quantities of the type (2), but do not provide a quantitative expression for the constant  $c$ .

Douc *et al.* [10] later provided rates of convergence for subgeometric chains with computable constants, but their approach is restricted to the convergence of the marginals (1) and no result is available concerning (2). Although bounds of the form (1) may imply (2) in some scenarios, such an approach may be suboptimal and lead to a significant loss. This is the case, for example, with certain polynomial kernels yielding (1) with rate  $\hat{r}(n) \propto n^\beta$  with some  $\beta > 0$  [15]. This guarantees the finiteness of the sum in (2) with a constant rate  $r(n) = 1$  only if  $\beta > 1$ , whereas our results imply (2) also with weaker polynomial rates including the cases  $\beta \in (0, 1]$  of [15].

Our main result, Theorem 1 in Section 2, provides an explicit upper bound for the constant  $c$  for a slight generalisation of (2). The approach follows that of Andrieu and Fort [1], but we complement it by providing explicit and relatively simple expressions, valid under a slightly stronger but more easily applicable one-step minorisation condition. In Section 3 we then establish a set of corollaries of Theorem 1 for the important special case of polynomially ergodic

chains, and continue with discussion on two specific applications in Section 4. The proof of Theorem 1 is given in Section 6, after describing the notation and definitions in Section 5. Our proof is nearly self-contained, using only two auxiliary results which are restated in Appendix A for the reader’s convenience.

### 2. Explicit rate of convergence for subgeometric Markov chains

We start with the generic main assumption, a subgeometric drift condition towards a small set, and recall the definition of Young’s functions.

**Condition 1.** *Suppose that  $(P_k)_{k \geq 1}$  is a collection of Markov kernels on a measurable space  $(X, \mathcal{B}(X))$ . Assume that there exist a set  $C \in \mathcal{B}(X)$ , a measurable function  $V : X \rightarrow [1, \infty)$  and a concave, nondecreasing and differentiable function  $\phi : [1, \infty) \rightarrow (0, \infty)$  such that  $\lim_{t \rightarrow \infty} \phi'(t) = 0$ . Moreover, there exist probability measures  $(\nu_k)_{k \geq 1}$  on  $(X, \mathcal{B}(X))$  and constants  $\varepsilon_\nu, \varepsilon_b \in (0, 1)$ ,  $b_V, c_V < \infty$  such that for all  $k \geq 1$ ,*

$$P_k V(x) \leq V(x) - \phi \circ V(x) + b_V \mathbf{1}_{\{x \in C\}}, \quad P_k(x, \cdot) \geq \varepsilon_\nu \nu_k(\cdot) \text{ for all } x \in C,$$

$$\inf_{x \notin C} \phi \circ V(x) \geq b_V(1 - \varepsilon_b)^{-1}, \quad \sup_{x \in C} V(x) \leq c_V,$$

where  $\mathbf{1}_{\{\cdot\}}$  is the indicator function.

**Definition 1.** The nondecreasing functions  $\Psi_1, \Psi_2 : [1, \infty) \rightarrow (0, \infty)$  are (a pair of) Young’s functions if  $\Psi_1(x)\Psi_2(y) \leq x + y$  for all  $x, y \geq 1$ .

Theorem 1 when applied with  $P_k = P$  is a refinement of [1, Proposition 3.1 and Theorem 3.6], since it provides an explicit expression for the upper bound.

**Theorem 1.** *Assume that Condition 1 holds. Then there exists a constant  $c \in [0, \infty)$  dependent on  $b_V, c_V, \varepsilon_b, \varepsilon_\nu$ , and  $\phi$  only, such that for any pair of Young’s functions  $\Psi_1, \Psi_2$  and any measurable  $f : X \rightarrow \mathbb{R}$  satisfying  $\|f\|_W := \sup_{x \in X} |f(x)|/W(x) < \infty$  with  $W(x) := \Psi_2(\phi \circ V(x)/\phi(1))$ ,*

$$\sum_{n \geq 0} \Psi_1(r(n)) |P^{(n)} f(x) - P^{(n)} f(x')| \leq c(V(x) + V(x') - 1) \|f\|_W, \tag{3}$$

where  $P^{(n)} := P_1 \cdots P_n$  with the convention that  $P^{(0)}(x, A) := \mathbf{1}_{\{x \in A\}}$ , and where  $r : \mathbb{N} \rightarrow [1, \infty)$  is defined through  $H_\phi : [1, \infty) \rightarrow [0, \infty)$  by

$$H_\phi(t) := \int_1^t \frac{ds}{\phi(s)}, \quad r(n) := \frac{\phi \circ H_\phi^{-1}(\varepsilon_b n)}{\phi(1)}. \tag{4}$$

The constant  $c$  can be given as

$$c := \frac{2}{\varepsilon_b \phi(1)} \left[ 2 + \frac{\bar{b}}{\varepsilon_\nu} + c_* \bar{b} r(1) \left( 1 + \frac{r(1)}{\varepsilon_b \phi(1)} \right) \right],$$

where

$$\bar{b} := 2b_V + \varepsilon_b \phi(1), \quad c_* := \sum_{j=1}^{\infty} (1 - \varepsilon_\nu)^{j-1} \prod_{k=1}^{j-1} (1 + \delta_k M_1),$$

$$\delta_k := \varepsilon_b (\phi' \circ H_\phi^{-1})(\varepsilon_b k), \quad M_1 := r(1) \left[ 1 + \frac{2r(1)}{\varepsilon_b \phi(1)} \left( \frac{b_V + c_V}{1 - \varepsilon_\nu} - 1 \right) \right].$$

The proof of Theorem 1 is postponed until Section 6.

**Remark 1.** We make a number of remarks in relation to Theorem 1.

- (i) It is easy to see that the assumptions imply  $\lim_{t \rightarrow \infty} H_\phi^{-1}(t) = \infty$  so  $\lim_{k \rightarrow \infty} \delta_k = 0$  and, therefore,  $c_* < \infty$ .
- (ii) In the case of a constant drift, that is, if the function  $\phi \equiv \varepsilon_\phi > 0$  then we have  $c_* = \varepsilon_\phi^{-1}$ .
- (iii) The condition  $\inf_{x \notin C} \phi \circ V(x) \geq b_V / (1 - \varepsilon_b)$  is essential for our proof; we need the bivariate drift established in Lemma 1. If  $\lim_{t \rightarrow \infty} \phi(t) = \infty$  then it is often possible to check Condition 1; see Corollary 2 for the polynomial case.
- (iv) If  $\mu_1$  and  $\mu_2$  are probability measures such that  $\mu_1(V) + \mu_2(V) < \infty$ , then (3) implies the following bound:

$$\sum_{n \geq 0} \Psi_1(r(n)) |\mu_1(P^{(n)} f) - \mu_2(P^{(n)} f)| \leq c(\mu_1(V) + \mu_2(V) - 1) \|f\|_W,$$

because for any function  $g$  integrable with respect to  $\mu_1$  and  $\mu_2$ , we have  $|\mu_1(g) - \mu_2(g)| \leq \int \mu_1(dx) \mu_2(dx') |g(x) - g(x')|$ .

- (v) Suppose that  $\pi$  is the invariant probability measure of  $P_k$  for  $k \geq 1$  and  $\pi(V) < \infty$ , then (1) with  $\mu_1 = \mathbf{1}_{\{x \in \cdot\}}$  and  $\mu_2 = \pi$  yields

$$\sum_{n \geq 0} \Psi_1(r(n)) |P^{(n)} f(x) - \pi(f)| \leq c(V(x) + \pi(V) - 1) \|f\|_W.$$

- (vi) It is possible to refine the bound by replacing the term  $c(V(x) + V(x') - 1)$  with  $c_1(V(x) + V(x')) + c_2$ , where the constants  $c_1$  and  $c_2$  are easily accessible from the statements of Lemmas 3 and 5.

### 3. Rate of convergence for polynomially ergodic chains

Here, we state two convenient corollaries of Theorem 1 in the case where  $P$  satisfies a polynomial drift condition. The first corollary characterises the required balance between the class of functions and the rate of convergence.

**Corollary 1.** Assume Condition 1 holds with  $\phi(v) = \beta v^\alpha$  with some constants  $\beta > 0$ ,  $\alpha \in [0, 1)$ , and  $\varepsilon_b \in (0, 1)$ . Then, for any  $\xi \in [0, 1]$  and for any measurable function  $f: X \rightarrow \mathbb{R}$  with

$$\|f\|_{V^{\alpha(1-\xi)}} := \sup_{x \in X} \frac{|f(x)|}{V^{\alpha(1-\xi)}(x)} < \infty,$$

there exists a constant  $c_{\alpha, \beta, \varepsilon_b, \xi} < \infty$  depending on  $\alpha$ ,  $\beta$ ,  $\varepsilon_b$ , and  $\xi$  such that

$$\sum_{n \geq 0} (n + 1)^{\xi\alpha/(1-\alpha)} |P^{(n)} f(x) - P^{(n)} f(x')| \leq c_{\alpha, \beta, \varepsilon_b, \xi} c \|f\|_{V^{\alpha(1-\xi)}} (V(x) + V(x') - 1), \quad (5)$$

where  $c < \infty$  is the constant given in Theorem 1.

*Proof.* We may compute

$$H_\phi(t) = \int_1^t \frac{ds}{\beta s^\alpha} = \frac{t^{1-\alpha} - 1}{\beta(1-\alpha)}, \quad H_\phi^{-1}(n) = (n\beta(1-\alpha) + 1)^{1/(1-\alpha)},$$

to obtain

$$r(n) = (\varepsilon_b n \beta (1 - \alpha) + 1)^{\alpha/(1-\alpha)} \geq (n + 1)^{\alpha/(1-\alpha)} c_{\alpha, \beta, \varepsilon_b},$$

where  $c_{\alpha, \beta, \varepsilon_b} := \min\{1, (\varepsilon_b \beta (1 - \alpha))^{\alpha/(1-\alpha)}\}$ . Define the functions

$$\Psi_1(x) := \begin{cases} \xi^{-1} x^\xi & \text{if } \xi \in (0, 1), \\ x & \text{if } \xi = 1, \\ 1 & \text{if } \xi = 0, \end{cases} \quad \Psi_2(y) := \begin{cases} (1 - \xi)^{-1} y^{1-\xi} & \text{if } \xi \in (0, 1), \\ 1 & \text{if } \xi = 1, \\ y & \text{if } \xi = 0, \end{cases} \tag{6}$$

satisfying  $\Psi_1(x)\Psi_2(y) \leq x + y$ , by Young’s inequality for  $\xi \in (0, 1)$ .

Theorem 1 implies that

$$\sum_{n \geq 0} \Psi_1\left(c_{\alpha, \beta, \varepsilon_b} (n + 1)^{\alpha/(1-\alpha)}\right) |P^{(n)} f(x) - P^{(n)} f(x')| \leq c \|f\| \Psi_2(\beta V^\alpha)(V(x) + V(x') - 1),$$

from which we deduce the claim with

$$c_{\alpha, \beta, \varepsilon_b, \xi} = c_{\alpha, \beta, \varepsilon_b}^{-\xi} \left[ (1 - \xi) \xi \beta^{\xi-1} + \mathbf{1}_{\{\xi=1\}} + \mathbf{1}_{\{\xi=0\}} \beta^{-1} \right].$$

We further consider a corollary which allows us to consider different growth rates of the upper bound in (5) in terms of  $x$  and  $x'$ .

**Condition 2.** Suppose that  $\mathcal{P}$  is a collection of Markov kernels on  $(X, \mathcal{B}(X))$ . Assume that there exist a set  $C \in \mathcal{B}(X)$  and a measurable function  $\hat{V} : X \rightarrow [1, \infty)$  with  $c_{\hat{V}} := \sup_C \hat{V} < \infty$  and constants  $\beta > 0$ ,  $\alpha \in (0, 1)$ , and  $b_{\hat{V}} < \infty$  such that for all  $P \in \mathcal{P}$ ,

$$P \hat{V}(x) \leq \begin{cases} \hat{V}(x) - \beta \hat{V}^\alpha(x) & \text{if } x \notin C, \\ b_{\hat{V}} & \text{if } x \in C. \end{cases}$$

Furthermore, suppose that every level set  $A_{\hat{V}}(v) := \{x \in X : \hat{V}(x) \leq v\}$  is uniformly 1-small, that is, there exist  $\varepsilon_v > 0$  and probability measures  $(\nu_P)_{P \in \mathcal{P}}$  on  $(X, \mathcal{B}(X))$  such that for all  $P \in \mathcal{P}$ ,

$$P(x, \cdot) \geq \varepsilon_v \nu_P(\cdot) \quad \text{for all } x \in A_{\hat{V}}(v).$$

We first observe that Condition 2 implies Condition 1 for functions  $V = \hat{V}^\eta$  with any  $\eta \in (1 - \alpha, 1]$ .

**Proposition 1.** Suppose Condition 2 holds. Then, for any  $(P_k)_{k \geq 1} \subset \mathcal{P}$  and  $\lambda \in [0, 1)$ , Condition 1 holds with  $V(x) = \hat{V}(x)^{1-\lambda\alpha}$ ,  $\phi(v) = (1 - \lambda\alpha)\beta v^{\alpha\lambda}$ , where  $\alpha_\lambda := \alpha(1 - \lambda)/(1 - \lambda\alpha)$  with the set  $C := A_{\hat{V}}(c_V)$  and with some constants  $\varepsilon_b, \varepsilon_v \in (0, 1)$ , and  $b_V, c_V < \infty$ , whose values depend only on  $\lambda$  and the constants and the function  $\hat{V}$  in Condition 2.

*Proof.* Let  $P \in \mathcal{P}$ . Following the proof of [15, Lemma 3.5], Jensen’s inequality, and the mean value theorem imply that, with  $\eta = 1 - \lambda\alpha$ ,

$$P \hat{V}^\eta(x) \leq \begin{cases} (\hat{V} - \beta \hat{V}^\alpha(x))^\eta \leq \hat{V}^\eta(x) - \eta \beta \hat{V}^{\eta\alpha_\lambda}(x) & \text{if } x \notin C, \\ b_{\hat{V}}^\eta & \text{if } x \in C, \end{cases}$$

where  $\alpha_\lambda = (\alpha - (1 - \eta))/\eta = \alpha(1 - \lambda)/(1 - \lambda\alpha)$ . Clearly,

$$P V(x) \leq V(x) - \phi \circ V(x) + b_V \mathbf{1}_{\{x \in C\}}, \tag{7}$$

where  $\phi(v) = \eta \beta v^{\alpha_\lambda}$  and  $b_V = b_{\hat{V}}^\eta + \phi(c_{\hat{V}})$ . Let  $\varepsilon_b \in (0, 1)$  and take  $c_V \in [c_{\hat{V}}, \infty)$  sufficiently large so that  $\phi(c_V) \geq b_V(1 - \varepsilon_b)^{-1}$ .

**Corollary 2.** *Suppose that Condition 2 holds. Then, for any  $\xi \in [0, 1]$  and  $\lambda \in [0, 1)$ , there exists a constant  $c_* < \infty$  such that for all  $(P_k)_{k \geq 1} \subset \mathcal{P}$  and  $\|f\|_{\hat{V}^{\alpha_{\lambda, \xi}}} < \infty$  where  $\alpha_{\lambda, \xi} = \alpha(1 - \lambda)(1 - \xi)$ ,*

$$\sum_{n \geq 0} (n + 1)^{\alpha(1-\lambda)\xi/(1-\alpha)} |P^{(n)} f(x) - P^{(n)} f(x')| \leq c_* \|f\|_{\hat{V}^{\alpha_{\lambda, \xi}}} (\hat{V}^{1-\lambda\alpha}(x) + \hat{V}^{1-\lambda\alpha}(x') - 1).$$

*Proof.* Proposition 1 and Corollary 1 imply that the claim holds.

### 4. Applications

Next, we discuss two specific applications of our results. Both applications are related to the evaluation of the efficiency of MCMC schemes in terms of asymptotic variance: the first application involves the so-called pseudo-marginal MCMC [2, 5], while the second application is related to a general comparison result of inhomogeneous Markov chains recently established in [16].

In both cases, we are interested in estimating an integral

$$\pi(f) := \int_{\mathbb{R}^d} f(x)\pi(x)dx,$$

where  $\pi(x)$  is a probability density and  $f$  is a  $\pi$ -integrable function. The efficiency criterion is the so-called asymptotic variance

$$\sigma^2(f) := \lim_{n \rightarrow \infty} \mathbb{E} \left\{ \frac{1}{\sqrt{n}} \sum_{k=1}^n [f(X_k) - \pi(f)] \right\}^2,$$

where  $(X_k)_{k \geq 0}$  denotes the Markov chain with initial distribution  $\pi$  and with the same  $\pi$ -invariant transition kernel(s) as the MCMC sampler and  $\mathbb{E}$  is the related expectation.

#### 4.1. Efficiency of the pseudo-marginal MCMC

The pseudo-marginal algorithm is relevant to situations where the density  $\pi$  cannot be evaluated point-wise, which prevents a straightforward implementation of the Metropolis–Hastings algorithm, for example. Such a situation occurs naturally, for instance when  $\pi(x)$  is a marginal density of a higher-dimensional density. As pointed out in [2] and [5] it is however possible to implement a valid (auxiliary variable) Metropolis–Hastings algorithm in this scenario, by using nonnegative unbiased estimators of the density values  $\pi(x)$ . Interestingly, regardless of the accuracy of the related estimator, the corresponding Markov chain will be ergodic with minimal assumptions, and therefore yield ergodic averages convergent to the integral of interest; see [2], [5].

However, the efficiency of the algorithm usually depends heavily on the properties of the estimators of  $\pi(x)$ . If the accuracy is increased, the pseudo-marginal algorithm tends to behave in a way similar to the ideal algorithm for which exact values of  $\pi(x)$  are used instead of estimators. In particular, let  $N \geq 1$  be a parameter controlling the accuracy of the estimator (such as the number of estimators used when using an averaging property to reduce variability), and let  $\sigma_N^2(f)$  be the asymptotic variance of the related pseudo-marginal algorithm. Then, under general conditions,  $\sigma_N^2(f) \rightarrow \sigma^2(f)$  as  $N \rightarrow \infty$ , where  $\sigma^2(f)$  is the asymptotic variance of the ideal algorithm [5, Theorem 21].

The key assumption required for the aforementioned result to hold is that the integrated autocorrelation series converge uniformly, that is,

$$\lim_{n \rightarrow \infty} \sup_{N \geq 1} \left| \sum_{k=n}^{\infty} \mathbb{E}\{\bar{f}(\tilde{X}_0^{(N)})\bar{f}(\tilde{X}_k^{(N)})\} \right| = 0, \tag{8}$$

where  $\bar{f}(x) = f(x) - \pi(f)$ , and where  $(\tilde{X}_k^{(N)})_{k \geq 0}$  corresponds to the Markov chain generated by the pseudo-marginal chain with accuracy parameter  $N$ .

The condition in (8) is relatively straightforward to check whenever the pseudo-marginal algorithms are geometrically ergodic with uniformly bounded drift and minorisation constants; see [8], [17]. However, pseudo-marginal algorithms are subgeometric whenever the density estimators of  $\pi(x)$  can take arbitrarily large values [5, Proposition 13].

This is the situation where Corollary 2 becomes relevant, as it is straightforward to check (8) under simultaneous (in  $N$ ) polynomial drift and minorisation conditions. In particular, we may write for any  $N$  for which the drift and minorisation conditions hold,

$$\left| \sum_{k=n}^{\infty} \mathbb{E}\{\bar{f}(\tilde{X}_0^{(N)})\bar{f}(\tilde{X}_k^{(N)})\} \right| \leq \mathbb{E} \left\{ |\bar{f}(X_0^{(N)})| \sum_{k=n}^{\infty} |\mathbb{E}\{\bar{f}(\tilde{X}_k^{(N)}) \mid \tilde{X}_0^{(N)}\}| \right\}.$$

The latter sum can be bounded by Corollary 2; see [5, Proposition 19]. We point out the importance of having explicit quantitative bounds here in order to ensure that an upper bound independent of  $N$  exists, that is, the constant  $c_*$  in Corollary 2 can be taken independent of  $N$ .

### 4.2. Ordering inhomogeneous Markov chains

In a number of scenarios MCMC algorithms may rely on the composition of several  $\pi$ -reversible MCMC kernels. For example, when two sampling strategies are available, that is two  $\pi$ -reversible Markov kernels  $P_0$  and  $Q_0$  can be implemented, one may consider implementing the algorithm which cycles between these two kernels. The recent result of Maire *et al.* [16, Theorem 4] shows that if  $P_1$  and  $Q_1$  form another pair of  $\pi$ -reversible kernels, and if  $P_0 \preceq P_1$  and  $Q_0 \preceq Q_1$  in the covariance order, then the asymptotic variances related to the two algorithms satisfy  $\sigma_1^2(f) \leq \sigma_0^2(f)$ .

The key assumption required by [16, Theorem 4] is that the integrated autocorrelation series converges absolutely; using notation analogous to (8),

$$\sum_{k=1}^{\infty} \left( |\mathbb{E}\{\bar{f}(X_0^{(i)})\bar{f}(X_k^{(i)})\}| + |\mathbb{E}\{\bar{f}(X_1^{(i)})\bar{f}(X_{k+1}^{(i)})\}| \right) < \infty, \quad i \in \{0, 1\}, \tag{9}$$

where  $(X_k^{(i)})_{k \geq 0}$  is the inhomogeneous Markov chain with initial distribution  $\pi$  and with alternating kernels  $P_i$  and  $Q_i$ .

Under geometric ergodicity, (9) is relatively easy to check [16]. In the subgeometric case, to the best of the authors' knowledge no results exist which would be directly applicable to verify (9). When its assumptions are satisfied one can use Theorem 1 to deduce (9), exploiting the fact that our results hold for inhomogeneous Markov chains. In particular, in the polynomial scenario, Corollary 2 may be applied following the arguments in [5, Proposition 19].

**5. Definitions: coupling and bivariate drift**

**Definition 2.** (*Coupling construction.*) Assume that Condition 1 holds, denote  $\bar{C} = C \times C$  and define the Markov kernels  $\check{P}_k$  on the product space  $(X \times X, \mathcal{B}(X) \times \mathcal{B}(X))$  by

$$\check{P}_k(x, x'; A, A') := P_k(x, A)P_k(x', A') \mathbf{1}_{\{(x,x') \notin \bar{C}\}} + Q_k(x, A)Q_k(x', A') \mathbf{1}_{\{(x,x') \in \bar{C}\}},$$

where  $Q_k(x, A) := (1 - \varepsilon_\nu)^{-1}(P_k(x, A) - \varepsilon_\nu \nu_k(A))$ .

Define then the Markov kernels  $\bar{P}_k$  on  $(X^2 \times \{0, 1\}, \mathcal{B}(X)^2 \times \mathcal{P}(\{0, 1\}))$  as follows for  $\check{A} \in \mathcal{B}(X) \times \mathcal{B}(X)$ :

$$\begin{aligned} \bar{P}_k(x, x', 0; \check{A} \times \{0\}) &= (1 - \varepsilon_\nu \mathbf{1}_{\{(x,x') \in \bar{C}\}}) \check{P}_k(x, x'; \check{A}), \\ \bar{P}_k(x, x', 0; \check{A} \times \{1\}) &= \varepsilon_\nu \mathbf{1}_{\{(x,x') \in \bar{C}\}} \nu_k(\{x \in X \mid (x, x) \in \check{A}\}), \\ \bar{P}_k(x, x', 1; \check{A} \times \{0\}) &= 0, \\ \bar{P}_k(x, x', 1; \check{A} \times \{1\}) &= \mathbf{1}_{\{x=x'\}} \int P_k(x, dy) \mathbf{1}_{\{(y,y) \in \check{A}\}} + \delta_{(x,x')}(\check{A}). \end{aligned}$$

Suppose that  $(X_n, X'_n, D_n)_{n \geq 0}$  is a Markov chain defined by the kernels  $\bar{P}_1, \bar{P}_2, \dots, \bar{P}_n$  and with  $(X_0, X'_0, D_0) \equiv (x, x', d)$ . We denote the probability and the expectation associated with the chain as  $\mathbb{P}_{x,x',d}$  and  $\mathbb{E}_{x,x',d}$ , respectively, and define the stopping times  $T_1 := \inf\{n \geq 0 : (X_n, X'_n) \in \bar{C}\}$  and  $T_k := \inf\{n > T_{k-1} : (X_n, X'_n) \in \bar{C}\}$  for  $k \geq 2$ , and  $\tau := \inf\{n \geq 0 : D_n = 1\}$ , with the convention  $\inf \emptyset = \infty$ .

Suppose that  $D_0 \equiv d = 0$ , then Definition 2 formalises a coupling with probability  $\varepsilon_\nu$  each time  $(X_n, X'_n) \in \bar{C}$ ; the stopping time  $\tau$  is a coupling time, and  $X_{\tau+k} \mathbf{1}_{\{\tau < \infty\}} = X'_{\tau+k} \mathbf{1}_{\{\tau < \infty\}}$  for all  $k \geq 0$   $\mathbb{P}_{x,x',0}$ -almost surely. If the coupling was not successful, then the chains follow independently  $\check{P}_k$  at time  $k$  until hitting  $\bar{C}$  again.

**Proposition 2.** Consider the Markov chain  $(X_n, X'_n, D_n)$  in Definition 2. Then,  $(X_n)_{n \geq 0}$  and  $(X'_n)_{n \geq 0}$  follow marginally  $P^{(n)}$  and, specifically,

$$\mathbb{P}_{x,x',0}\{X_n \in A\} = P^{(n)}(x, A), \quad \mathbb{P}_{x,x',0}\{X'_n \in A\} = P^{(n)}(x', A)$$

for all  $n \geq 0$ , all  $(x, x') \in X^2$ , and any  $A \in \mathcal{B}(X)$ .

*Proof.* It is easy to see that for any  $(x, x') \in X^2$  and  $A \in \mathcal{B}(X)$ ,

$$\begin{aligned} \bar{P}_k(x, x', 0; A \times X \times \{0, 1\}) &= P_k(x, A), \\ \bar{P}_k(x, x', 0; X \times A \times \{0, 1\}) &= P_k(x', A), \end{aligned}$$

and  $\mathbb{P}_{x,x',0}\{X_n = X'_n \mid D_n = 1\} = 1$ .

**Lemma 1.** Assume Condition 1 holds and denote  $\bar{V}(x, x') := V(x) + V(x') - 1$ , then

$$\bar{P}_k \bar{V}(x, x', 0) \leq \begin{cases} \bar{V}(x, x') - \varepsilon_b \phi \circ \bar{V}(x, x') & \text{if } (x, x') \notin \bar{C}, \quad (10a) \\ 2(b_V + c_V) - 1 & \text{if } (x, x') \in \bar{C}, \quad (10b) \\ \bar{V}(x, x') - \varepsilon_b(\phi \circ \bar{V})(x, x') + \bar{b} \mathbf{1}_{\{(x,x') \in \bar{C}\}} & \text{if } (x, x') \in X^2, \quad (10c) \end{cases}$$

$$\check{P}_k \bar{V}(x, x') \leq 2(1 - \varepsilon_\nu)^{-1}(b_V + c_V) - 1 \quad \text{if } (x, x') \in \bar{C}, \quad (10d)$$

where  $\bar{b} = 2b_V + \varepsilon_b \phi(1)$ .

*Proof.* Condition 1 implies that, for  $(x, x') \notin \bar{C}$ ,

$$\begin{aligned} \bar{P}_k \bar{V}(x, x', 0) &\leq \bar{V}(x, x') - \phi \circ V(x) - \phi \circ V(x') + b_V (\mathbf{1}_{\{x \in C\}} + \mathbf{1}_{\{x' \in C\}}) \\ &\leq \bar{V}(x, x') - \varepsilon_b (\phi \circ V(x) + \phi \circ V(x')) - (1 - \varepsilon_b) \inf_{z \notin C} \phi \circ V(z) + b_V \\ &\leq \bar{V}(x, x') - \varepsilon_b \phi \circ \bar{V}(x, x'), \end{aligned}$$

where the last inequality follows because  $\phi$  is convex and nondecreasing and, thus,

$$\phi \circ \bar{V}(x, x') - \phi \circ V(x) \leq \phi \circ V(x') - \phi(1). \tag{11}$$

This establishes (10a). The bound (10d) follows from

$$\check{P}_k \bar{V}(x, x') = Q_k V(x) + Q_k V(x') - 1 \leq 2(1 - \varepsilon_v)^{-1} (c_V + b_V - \varepsilon_v \nu_k(V)) - 1. \tag{12}$$

For (10b), let us write for  $(x, x') \in \bar{C}$ ,

$$\bar{P}_k \bar{V}(x, x', 0) = P_k V(x) + P_k V(x') - 1 \leq \bar{V}(x, x') - (\phi \circ V(x) + \phi \circ V(x')) + 2b_V.$$

Finally, we turn to (10c) and observe that the above inequality with (11) and (10a) imply that

$$\begin{aligned} \bar{P}_k \bar{V}(x, x', 0) &\leq \bar{V}(x, x') - \varepsilon_b \phi \circ \bar{V}(x, x') \mathbf{1}_{\{(x, x') \notin \bar{C}\}} \\ &\quad - (\phi \circ \bar{V}(x, x') + \phi(1) + 2b_V) \mathbf{1}_{\{(x, x') \in \bar{C}\}} \\ &\leq \bar{V}(x, x') - \varepsilon_b \phi \circ \bar{V}(x, x') \\ &\quad + \sup_{(x, x') \in \bar{C}} [2b_V + \phi(1) - (1 - \varepsilon_b) \phi \circ \bar{V}(x, x')] \mathbf{1}_{\{(x, x') \in \bar{C}\}}. \end{aligned}$$

The claim follows noting that  $\phi \circ \bar{V}(x, x') \geq \phi(1)$ .

### 6. Proof of Theorem 1

We provide the skeleton of the proof of Theorem 1 next, and postpone bounding the involved terms to lemmas.

*Proof of Theorem 1.* It is sufficient to prove the claim assuming that  $\|f\|_W = 1$ . Consider the coupling construction in Definition 2. We may write

$$\begin{aligned} &\sum_{n \geq 0} \Psi_1(r(n)) |P^{(n)} f(x) - P^{(n)} f(x')| \\ &= \sum_{n \geq 0} \Psi_1(r(n)) |\mathbb{E}_{x, x', 0} \{ (f(X_n) - f(X'_n)) \mathbf{1}_{\{\tau > n\}} \}| \\ &\leq \mathbb{E}_{x, x', 0} \left\{ \sum_{n=0}^{\tau-1} \Psi_1(r(n)) W(X_n) \right\} + \mathbb{E}_{x, x', 0} \left\{ \sum_{n=0}^{\tau-1} \Psi_1(r(n)) W(X_n) \right\}. \end{aligned}$$

Because  $\Psi_1(x) \Psi_2(y) \leq x + y$ , we obtain the bound

$$\sum_{n \geq 0} \Psi_1(r(n)) |P^{(n)} f(x) - P^{(n)} f(x')| \leq 2 \left[ E_1(x, x') + \frac{E_2(x, x')}{\phi(1)} \right],$$

where the terms on the right-hand side are defined as

$$E_1(x, x') := \mathbb{E}_{x, x', 0} \left\{ \sum_{n=0}^{\tau-1} r(n) \right\}, \quad E_2(x, x') := \mathbb{E}_{x, x', 0} \left\{ \sum_{n=0}^{\tau-1} \phi \circ \bar{V}(X_n, X'_n) \right\},$$

and these terms are bounded by Lemmas 3 and 5 below.

**Lemma 2.** *Let  $\phi: [1, \infty) \rightarrow (0, \infty)$  be concave, nondecreasing and differentiable, and let  $r(n)$  be as defined in (4). Then,  $r(n)$  is nondecreasing and for all  $n, m \geq 0$ ,*

$$r(n + m) \leq r(n)r(m), \tag{13a}$$

$$r(n + m) - r(n) \leq \varepsilon_b(\phi' \circ H_\phi^{-1})(\varepsilon_b n)r(n) \sum_{k=1}^m r(k). \tag{13b}$$

*Proof.* Denote  $r(t) := (\phi \circ H_\phi^{-1})(\varepsilon_b t)/\phi(1)$  for  $t \in \mathbb{R}_+$ , and compute

$$r'(t) = \varepsilon_b(\phi' \circ H_\phi^{-1})(\varepsilon_b t)r(t) \geq 0, \quad (\log r)'(t) = \varepsilon_b(\phi' \circ H_\phi^{-1})(\varepsilon_b t).$$

The latter is nonincreasing, therefore (13a) follows from

$$\log r(n + m) - \log r(n) = \int_n^{n+m} (\log r)'(t)dt \leq \int_0^m (\log r)'(t)dt = \log r(m),$$

because  $r(0) = 1$ . By the mean value theorem

$$r(n + m) - r(n) = \sum_{k=n}^{n+m-1} (r(k + 1) - r(k)) = \sum_{k=n}^{n+m-1} r'(k + \xi_k)$$

for some  $\xi_n, \dots, \xi_{n+m-1} \in [0, 1]$ . Observe that  $r'(k + \xi_k) \leq \varepsilon_b(\phi' \circ H_\phi^{-1})(\varepsilon_b n)r(k + 1)$ , so

$$r(n + m) - r(n) \leq \varepsilon_b(\phi' \circ H_\phi^{-1})(\varepsilon_b n) \sum_{k=n}^{n+m-1} r(k + 1).$$

We deduce (13b) by applying (13a).

**Lemma 3.** *Assume Condition 1 holds and consider the coupling construction in Definition 2. Then,*

$$\mathbb{E}_{x, x', 0} \left\{ \sum_{n=0}^{\tau-1} \phi \circ \bar{V}(X_n, X'_n) \right\} \leq \frac{1}{\varepsilon_b} \bar{V}(x, x') + \frac{\bar{b}}{\varepsilon_b \varepsilon_v}, \tag{14}$$

where  $\bar{b}$  is defined in Theorem 1.

*Proof.* By (10c) and Proposition 3,

$$\mathbb{E}_{x, x', 0} \left\{ \sum_{n=0}^{\tau-1} \varepsilon_b(\phi \circ \bar{V})(X_n, X'_n) \right\} \leq \bar{V}(x, x') + \bar{b} \mathbb{E}_{x, x', 0} \left\{ \sum_{n=0}^{\tau-1} \mathbf{1}_{\{(X_n, X'_n) \in \bar{C}\}} \right\},$$

and we can write

$$\mathbb{E}_{x,x',0} \left\{ \sum_{n=0}^{\tau-1} \mathbf{1}_{\{(X_n, X'_n) \in \bar{C}\}} \right\} = \sum_{j=1}^{\infty} \mathbb{P}_{x,x',0} \{\tau > T_j\} = \sum_{j=1}^{\infty} (1 - \varepsilon_\nu)^{j-1},$$

because  $\mathbb{P}_{x,x',0} \{\tau > T_j\} = \mathbb{P}_{x,x',0} \{D_{T_1+1} = 0, \dots, D_{T_{j-1}+1} = 0\}$  and  $\mathbb{P}_{x,x',0} \{D_{T_i+1} = 0 \mid D_{T_1+1} = 0, \dots, D_{T_{i-1}+1} = 0\} = 1 - \varepsilon_\nu$ .

**Lemma 4.** Assume that Condition 1 holds, let  $r$  be defined in (4) and consider the coupling construction in Definition 2. Then,

$$\mathbb{E}_{x,x',0} \left\{ \sum_{n=0}^{T_1} r(n) \right\} \leq 1 + \frac{r(1)}{\varepsilon_b \phi(1)} (\bar{V}(x, x') - 1) \mathbf{1}_{\{(x,x) \notin \bar{C}\}}.$$

*Proof.* The claim holds trivially for  $(x, x') \in \bar{C}$ , so assume that  $(x, x') \in \bar{C}$ . From (10c) we apply Proposition 4 with  $\varphi = \varepsilon_b \phi$  and  $b = \bar{b}$ ; note that  $r_\varphi(n) = r(n)$ . Then, Proposition 3 with  $Z_k := (H_k \circ \bar{V})(X_k, X'_k)$  yields

$$\mathbb{E}_{x,x',0} \left\{ \sum_{n=0}^{T_1-1} \varepsilon_b \phi(1) r(n) \right\} \leq H_0 \circ \bar{V}(x, x') = \bar{V}(x, x') - 1.$$

Equation (13a) implies that  $r(n + 1) \leq r(1)r(n)$  so, we deduce

$$\mathbb{E}_{x,x',0} \left\{ \sum_{n=0}^{T_1} r(n) \right\} = 1 + \mathbb{E}_{x,x',0} \left\{ \sum_{n=1}^{T_1} r(n) \right\} \leq 1 + r(1) \mathbb{E}_{x,x',0} \left\{ \sum_{n=0}^{T_1-1} r(n) \right\}.$$

**Lemma 5.** Assume that Condition 1 holds, let  $r$  be defined in (4) and consider the coupling construction in Definition 2. Then,

$$\mathbb{E}_{x,x',0} \left\{ \sum_{n=0}^{\tau-1} r(n) \right\} \leq \frac{1}{\varepsilon_b \phi(1)} \left[ \left( 1 + \frac{c_* \bar{b} r^2(1)}{\varepsilon_b \phi(1)} \right) \bar{V}(x, x') + \left( c_* \bar{b} r(1) - 1 - \frac{c_* \bar{b} r^2(1)}{\varepsilon_b \phi(1)} \right) \right],$$

where  $\bar{b}$  and  $c_*$  are given in Theorem 1.

*Proof.* Equation (10c) and Propositions 3 and 4 applied as in the proof of Lemma 4 yield

$$\mathbb{E}_{x,x',0} \left\{ \sum_{n=0}^{\tau-1} \varepsilon_b \phi(1) r(n) \right\} \leq \bar{V}(x, x') - 1 + \bar{b} \mathbb{E}_{x,x',0} \left\{ \sum_{n=0}^{\tau-1} r(n+1) \mathbf{1}_{\{(X_n, X'_n) \in \bar{C}\}} \right\}.$$

The latter expectation can be written as

$$\mathbb{E}_{x,x',0} \left\{ \sum_{n=0}^{\tau-1} r(n+1) \mathbf{1}_{\{(X_n, X'_n) \in \bar{C}\}} \right\} = \sum_{j=1}^{\infty} \mathbb{E}_{x,x',0} \{r(T_j + 1) \chi_j\},$$

where  $\chi_j := \mathbf{1}_{\{\tau > T_j\}} = \chi_{j-1} \mathbf{1}_{\{D_{T_j+1} = 0\}}$  for  $j \geq 1$  and  $\chi_0 \equiv 1$ .

Equation (13b) implies that, for  $j \geq 1$ ,

$$r(T_{j+1} + 1) \leq r(T_j + 1) \left( 1 + \varepsilon_b(\phi' \circ H_\phi^{-1})(\varepsilon_b(T_j + 1)) \sum_{k=1}^{T_{j+1}-T_j} r(k) \right),$$

and  $\varepsilon_b(\phi' \circ H_\phi^{-1})(\varepsilon_b(T_j + 1)) \leq \varepsilon_b(\phi' \circ H_\phi^{-1})(\varepsilon_b(j)) = \delta_j$ , because  $T_j + 1 \geq j$  and  $\phi' \circ H_\phi^{-1}$  is nonincreasing. Next, we show that taking conditional expectation with respect to  $\mathcal{F}_{T_j} = \sigma((X_n, X'_n, D_n) : 1 \leq n \leq T_j)$ , we obtain for  $j \geq 1$ ,

$$\mathbb{E}_{x,x',0}\{r(T_{j+1} + 1)\chi_{j+1}\} \leq \mathbb{E}_{x,x',0}\{r(T_j + 1)\chi_j\}(1 + \delta_j M_1)(1 - \varepsilon_v), \tag{15}$$

where  $M_1$  is given below. Namely,  $\mathbb{P}_{x,x',0}\{D_{T_j+1} = 0 \mid \mathcal{F}_{T_j}\} = (1 - \varepsilon_v)$  and

$$\begin{aligned} & \mathbb{E}_{x,x',0} \left\{ \sum_{k=1}^{T_{j+1}-T_j} r(k) \mid \mathcal{F}_{T_j}, D_{T_j+1} = 0 \right\} \\ & \leq \sup_{k \geq 1} \sup_{(x,x') \in \bar{C}} \int \check{P}_k(x, x'; dy, dy') \mathbb{E}_{y,y',0}^{(T_j)} \left\{ \sum_{n=0}^{T_1^{(T_j)}} r(n+1) \right\}, \end{aligned}$$

where  $\mathbb{E}_{x,x',0}^{(j)}$  stands for the expectation over the stopping time  $T_1^{(j)}$  corresponding to the Markov chain  $(X_n^{(j)}, X'_n{}^{(j)}, D_n^{(j)})$  constructed as in Definition 2 but defined using  $(P_{k+j})_{k \geq 1}$  instead of  $(P_k)_{k \geq 1}$ . Lemma 4 still applies for this expectation, because it assumes only that the kernels satisfy Condition 1. Therefore, (10d) and the bound  $r(k + 1) \leq r(k)r(1)$  by (13a) yield

$$\mathbb{E}_{x,x',0} \left\{ \sum_{k=1}^{T_{j+1}-T_j} r(k) \mid \mathcal{F}_{T_j}, D_{T_j+1} = 0 \right\} \leq r(1) \left[ 1 + \frac{2r(1)}{\varepsilon_b \phi(1)} \left( \frac{b_V + c_V}{1 - \varepsilon_v} - 1 \right) \right] = M_1.$$

Applying (15) recursively for  $j + 1, \dots, 2$  yields

$$\mathbb{E}_{x,x',0}\{r(T_{j+1} + 1)\chi_{j+1}\} \leq (1 - \varepsilon_v)^{j-1} \prod_{i=1}^{j-1} (1 + \delta_i M_1) \mathbb{E}_{x,x',0}\{r_\phi(T_1 + 1)\},$$

and from Lemma 4 together with (13a),

$$\mathbb{E}_{x,x',0}\{r(T_1 + 1)\} \leq r(1) \left( 1 + \frac{r(1)}{\varepsilon_b \phi(1)} (\bar{V}(x, x') - 1) \right).$$

Combining these results, we have

$$\mathbb{E}_{x,x',0} \left\{ \sum_{n=0}^{\tau-1} r(n) \right\} \leq \frac{1}{\varepsilon_b \phi(1)} \left\{ \bar{V}(x, x') - 1 + c_* \bar{b} r(1) \left( 1 + \frac{r(1)}{\varepsilon_b \phi(1)} (\bar{V}(x, x') - 1) \right) \right\},$$

which equals the desired bound.

## Appendix A. Some results from the literature

Here, we restate some results from the literature for the reader's convenience. We start by stating [18, Proposition 11.3.2] for inhomogeneous Markov chains; the proof of [18] applies without modifications.

**Proposition 3.** *Suppose that  $(X_n)_{n \geq 0}$  is a Markov chain and let  $\mathcal{F}_n := \sigma(X_0, \dots, X_n)$  for  $n \geq 0$ . Assume that  $Z_n$  is nonnegative and  $\mathcal{F}_n$ -adapted,  $f_n$  and  $s_n$  are nonnegative measurable functions and*

$$\mathbb{E}\{Z_{n+1} \mid \mathcal{F}_n\} \leq Z_n - f_n(X_n) + s_n(X_n) \quad \text{for all } n \geq 0.$$

Then for any initial condition  $x$  and any stopping time  $\tau$ ,

$$\mathbb{E}_x \left\{ \sum_{k=0}^{\tau-1} f_k(X_k) \right\} \leq Z_0(x) + \mathbb{E}_x \left\{ \sum_{k=0}^{\tau-1} s_k(X_k) \right\}.$$

**Proposition 4.** (See [11, Proposition 2.1].) *Assume that  $P$  is a Markov kernel satisfying*

$$PV(x) \leq V(x) - \varphi \circ V(x) + b \mathbf{1}_{\{x \in C\}},$$

where  $\varphi: [1, \infty) \rightarrow (0, \infty)$  is a nondecreasing convex function. Then for  $r_\varphi(k) := (\varphi \circ H_\varphi^{-1})(n)/\varphi(1)$ , where  $H_\varphi$  is as defined in (4),

$$PV_{k+1}(x) \leq V_k(x) - \varphi(1)r_\varphi(k) + br_\varphi(k+1) \mathbf{1}_{\{x \in C\}} \quad \text{for all } k \geq 0,$$

where  $V_k := H_k \circ V$  and

$$H_k(v) := \varphi(1) \int_0^{H_\varphi(v)} r_\varphi(z+k) dz = H_\varphi^{-1}(H_\varphi(v)+k) - H_\varphi^{-1}(k).$$

## Acknowledgements

The work of Christophe Andrieu was partially supported by a Winton Capital research award. Matti Vihola was supported by the Academy of Finland (project no. 250575).

## References

- [1] ANDRIEU, C. AND FORT, G. (2005). Explicit control of subgeometric ergodicity. Res. Rep. 05:17, School of Mathematics, University of Bristol.
- [2] ANDRIEU, C. AND ROBERTS, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.* **37**, 697–725.
- [3] ANDRIEU, C. AND THOMS, J. (2008). A tutorial on adaptive MCMC. *Statist. Comput.* **18**, 343–373.
- [4] ANDRIEU, C. AND VIHOLA, M. (2014). Markovian stochastic approximation with expanding projections. *Bernoulli* **20**, 545–585.
- [5] ANDRIEU, C. AND VIHOLA, M. (2015). Convergence properties of pseudo-marginal Markov chain Monte Carlo algorithms. *Ann. Appl. Probab.* **25**, 1030–1077.
- [6] ANDRIEU, C., MOULINES, É. AND PRIOURET, P. (2005). Stability of stochastic approximation under verifiable conditions. *SIAM J. Control Optimization* **44**, 283–312.
- [7] ATCHADÉ, Y. AND FORT, G. (2010). Limit theorems for some adaptive MCMC algorithms with subgeometric kernels. *Bernoulli* **16**, 116–154.
- [8] BAXENDALE, P. H. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.* **15**, 700–738.
- [9] DOUC, R., MOULINES, E. AND ROSENTHAL, J. S. (2004). Quantitative bounds on convergence of time-inhomogeneous Markov chains. *Ann. Appl. Probab.* **14**, 1643–1665.

- [10] DOUC, R., MOULINES, E. AND SOULIER, P. (2007). Computable convergence rates for sub-geometric ergodic Markov chains. *Bernoulli* **13**, 831–848.
- [11] DOUC, R., FORT, G., MOULINES, E. AND SOULIER, P. (2004). Practical drift conditions for subgeometric rates of convergence. *Ann. Appl. Prob.* **14**, 1353–1377.
- [12] FORT, G. (2001). Contrôle explicite d'ergodicité de chaîne de Markov: applications à l'analyse de convergence de l'algorithme Monte-Carlo EM. *Doctoral thesis*, Université Paris VI.
- [13] FORT, G. AND MOULINES, E. (2003). Polynomial ergodicity of Markov transition kernels. *Stoch. Process. Appl.* **103**, 57–99.
- [14] FORT, G., MOULINES, E. AND PRIOURET, P. (2011). Convergence of adaptive and interacting Markov chain Monte Carlo algorithms. *Ann. Statist.* **39**, 3262–3289.
- [15] JARNER, S. F. AND ROBERTS, G. O. (2002). Polynomial convergence rates of Markov chains. *Ann. Appl. Prob.* **12**, 224–247.
- [16] MAIRE, F., DOUC, R. AND OLSSON, J. (2014). Comparison of asymptotic variances of inhomogeneous Markov Chains with application to Markov chain Monte Carlo methods. *Ann. Statist.* **42**, 1484–1510.
- [17] MEYN, S. P. AND TWEEDIE, R. L. (1994). Computable bounds for geometric convergence rates of Markov chains. *Ann. Appl. Prob.* **4**, 981–1011.
- [18] MEYN, S. AND TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*, 2nd edn. Cambridge University Press.
- [19] SAKSMAN, E. AND VIHOLA, M. (2010). On the ergodicity of the adaptive Metropolis algorithm on unbounded domains. *Ann. Appl. Prob.* **20**, 2178–2203.
- [20] TUOMINEN, P. AND TWEEDIE, R. L. (1994). Subgeometric rates of convergence of  $f$ -ergodic Markov chains. *Adv. Appl. Prob.* **26**, 775–798.