Practical drift conditions for subgeometric rates of convergence

Randal Douc* Gersende Fort† Eric Moulines‡ Philippe Soulier§

Abstract

We present a new drift condition which implies rates of convergence to the stationary distribution of the iterates of a ψ-irreducible aperiodic and positive recurrent transition kernel. This condition, extending a condition introduced by Jarner and Roberts (2002) for polynomial convergence rates, turns out to be very convenient to prove subgeometric rates of convergence. Several applications are presented including nonlinear autoregressive models, stochastic unit root models and multidimensional random walk Hastings Metropolis algorithms.

Abbreviated title Subgeometric rates of convergence.

MSC 2000 subject classification 60J10.

Key words and phrases Markov chains, stationary distribution, rate of convergence.

1 Introduction

Let $\mathbf{\Phi} := (\Phi_n, n \geq 0)$ be a discrete time Markov chain on a general measurable state space $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ with transition kernel P. Assume that $\mathbf{\Phi}$ is ψ irreducible, aperiodic and positive recurrent. This paper considers the use of drift conditions to establish the convergence in f norm of the iterates P^n of the kernel to the stationary distribution π at rate $r := (r(n), n \geq 0)$, *i.e.*

$$\lim_{n\to\infty} r(n) \|P^n(x,\cdot) - \pi\|_f = 0, \quad \pi - \text{a.e.} \tag{1.1}$$

where $f : \mathsf{X} \to [1, \infty]$ is an extended real valued function and for any signed measure μ, the f norm $\|\mu\|_f$ is defined as $\sup_{|g|\leq f} |\mu(g)|$ (cf. (Meyn and Tweedie, 1993, Chapter 14), hereafter MT).

*CMAP, Ecole Polytechnique, Palaiseau, France

†LMC-IMAG, 51, rue des Mathématiques, BP53, 38041 Grenoble Cedex 9, France

‡Département TSI, Ecole nationale supérieure des Télécommunications, 46 rue Barrault, 75013 Paris, France

§Département de mathématiques, Université d'Evry, 91025 Evry Cedex, France. Corresponding author. Email: phliippe.soulier@maths.univ-evry.fr

For geometric rate functions, *i.e.* functions r that satisfy

$$0 < \liminf \frac{\log r(n)}{n} \quad \text{and} \quad \limsup \frac{\log r(n)}{n} < \infty \tag{1.2}$$

it is known (MT, Theorem 16.0.1) that (1.1) holds if and only if the Foster Lyapunov drift condition is verified *i.e.* there exist an extended real valued function $V : \mathsf{X} \to [1, \infty]$ finite at some $x_0 \in \mathsf{X}$, a petite set C, $\lambda \in (0,1)$, $b > 0$ and $c > 0$ such that $c^{-1} f \leq V \leq cf$ and

$$PV \leq \lambda V + b\mathbf{1}_C. \tag{1.3}$$

In that case, the convergence (1.1) holds for all x in the set $\{V < \infty\}$ which is of π measure one.

For rates of convergence slower than geometric, no such definitive result exist. An important family of such rates is the class of subgeometric rate functions, defined in Nummelin and Tuominen (1983) as follows. Let Λ_0 be the set of positive non decreasing functions r_0 such that $r_0(0) \geq 1$ and $\log\{r_0(n)\}/n$ decreases to 0. The class of subgeometric rate functions is the set Λ of positive functions r such that there exists a sequence $r_0 \in \Lambda_0$ and

$$\liminf r(n)/r_0(n) > 0 \quad \text{and} \quad \limsup r(n)/r_0(n) < \infty. \tag{1.4}$$

This class includes for example polynomial rate functions, *i.e.* rate functions r such that (1.4) holds with $r_0(n) := (1+n)^\beta$ and $\beta \geq 0$. It also includes rate functions which increase faster than polynomially, *e.g.* rate functions r satisfying (1.4) with

$$r_0(n) := (n+1)^\beta e^{cn^\gamma}, \quad \text{for} \ \ \beta \in \mathbb{R}, \ \gamma \in (0,1) \ \ \text{and} \ c > 0. \tag{1.5}$$

We will refer to these rates as subexponential in order to distinguish them in the broad class of subgeometric rates.

Tuominen and Tweedie (1994) (see also Nummelin and Tuominen (1983)) have given a set of necessary and sufficient conditions for the convergence (1.1) to hold with a subgeometric rate function $r \in \Lambda$. To state this result, we first recall some notations and definitions.

A measurable set C is $\psi_a$petite (or petite) if there exist a distribution $a := (a(n), n \geq 0)$, a constant $\epsilon > 0$ and a non trivial measure ψ_a on $\mathcal{B}(\mathsf{X})$ such that for all $x \in C$, $B \in \mathcal{B}(\mathsf{X})$,

$$K_a(x,B) := \sum_{n \geq 0} a(n) P^n(x,B) \geq \psi_a(B).$$

The return time to a measurable set A, denoted by τ_A is defined as $\tau_A := \inf\{n \geq 1, \Phi_n \in A\}$ (with the convention $\inf \emptyset = +\infty$). Let ψ be a maximal irreducibility measure and let $\mathcal{B}^+(\mathsf{X})$ be the set of accessible sets, *i.e.* sets $B \in \mathcal{B}(\mathsf{X})$ such that $\psi(B) > 0$. A set $A \in \mathcal{B}(\mathsf{X})$ is called full if $\psi(A^c) = 0$, absorbing if $P(x,A) = 1$ for all $x \in A$ and, for a measurable positive function f and a rate function r, A is said (f,r) *regular* if, for every $B \in \mathcal{B}^+(\mathsf{X})$,

$$\sup_{x \in A} \mathbb{E}_x \left[\sum_{k=0}^{\tau_B - 1} r(k) f(\Phi_k) \right] < \infty.$$

A finite positive measure λ on $\mathcal{B}(\mathsf{X})$ is said (f,r) regular if $\mathbb{E}_\lambda[\sum_{k=0}^{\tau_B-1} r(k)\ f(\Phi_k)] < \infty$ for all set $B \in \mathcal{B}^+(\mathsf{X})$. The set of all (f,r) regular points (*i.e.* the points $x \in \mathsf{X}$ such that δ_x is (f,r) regular) is denoted by $S(f,r)$.

We can now recall (part of) (Tuominen and Tweedie, 1994, Theorem 2.1).

Theorem 1.1 (Tuominen and Tweedie (1994)). *Assume that P is ψ irreducible and aperiodic. Let $f : \mathsf{X} \to [1,\infty]$ be a measurable function, and let $r \in \Lambda$ be given. The following conditions are equivalent.*

(i) There exists a petite set $C \in \mathcal{B}(\mathsf{X})$ such that

$$\sup_{x \in C} \mathbb{E}_x \left[\sum_{k=0}^{\tau_C - 1} r(k) f(\Phi_k) \right] < \infty.$$

(ii) There exist a sequence of extended real valued functions $(V_n, n \geq 0)$, $V_n : \mathsf{X} \to [1,\infty]$, a petite set $C \in \mathcal{B}(\mathsf{X})$ and a constant $b < \infty$ such that V_0 is bounded on C,

$$V_0(x) = +\infty \quad \Rightarrow \quad V_1(x) = +\infty$$

and

$$PV_{n+1} + r(n) f \leq V_n + b r(n) \mathbf{1}_C. \tag{1.6}$$

(iii) There exists a (f,r) regular set $A \in \mathcal{B}^+(\mathsf{X})$.

Any of these conditions implies that, for all $x \in S(f,r)$,

$$r(n) \| P^n(x,\cdot) - \pi(\cdot) \|_f = 0, \quad n \to \infty,$$

and the set $S(f,r)$ is full, absorbing and contains the set $\{V_0 < \infty\}$. Moreover, for all (f,r) regular initial distributions λ, μ, there exists a constant c such that

$$\sum_{n=0}^{\infty} r(n) \int\int \lambda(dx)\mu(dy) \| P^n(x,\cdot) - P^n(y,\cdot) \|_f \leq c\left(\lambda(V_0) + \mu(V_0)\right).$$

This theorem cannot be improved since it provides a necessary and sufficient condition, but the sequence of drift conditions (1.6) is notoriously difficult to check in practice and one has very little insight on the way to choose the family of drift function $(V_n, n \geq 0)$. This is why these drift conditions, up to the best of our knowledge, have seldom been used directly.

A first step towards finding a more practical drift condition was taken by Jarner and Roberts (2002) who, simplifying and generalising an argument in Fort and Moulines (2000), have shown that if there exist a function $V : \mathsf{X} \to [1,\infty]$ finite at some $x_0 \in \mathsf{X}$, positive constants b and c, a petite set C and $\alpha \in [0,1)$ such that

$$PV + cV^\alpha \leq V + b\mathbf{1}_C,$$

then the chain is positive recurrent and for each $\beta \in [1, 1/(1-\alpha)]$, the convergence (1.1) holds for all $x \in \{V < \infty\}$ which is of π measure one, with $r(n) := n^{\beta-1}$ and $f := V^{1-\beta(1-\alpha)}$. It is noteworthy that there is a balance between the rate of convergence and the norm: the larger the latter, the slower the former. In particular, the fastest rate of convergence ($r(n) \equiv n^{\alpha/(1-\alpha)}$) corresponds to the total variation norm, and the slowest rate ($r(n) \equiv 1$) corresponds to the V^α norm.

In this paper, we consider the following drift condition which generalizes the Foster Lyapunov and the Jarner Roberts drift conditions.

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

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

Here ϕ is assumed differentiable for convenience. It can be relaxed since a concave function has non increasing left and right derivatives everywhere. If P is ψ irreducible and aperiodic, and $\mathbf{D}(\phi, V, C)$ holds for some petite set C such that $\sup_C V < \infty$, then the f norm ergodic Theorem for aperiodic chain (see MT, Theorem 14.0.1) states that there exists an unique invariant distribution π and that the limit

$$\lim_n \|P^n(x, \cdot) - \pi\|_{\phi \circ V} = 0,$$

for all x in the set of π measure one $\{V < \infty\}$. The $\phi \circ V$ norm is the maximal norm for which convergence can be proved under condition $\mathbf{D}(\phi, V, C)$, and in that case, the rate of convergence is minimal: $r \equiv 1$. This implies that for any function $1 \leq f \leq \phi \circ V$ convergence in the f norm also holds. In order to determine the rate of convergence in the f norm, we should try to find a sequence of function $(V_n, n \geq 0)$ such that (1.6) holds, but this is precisely what we are trying to avoid doing for all functions f. Instead, having in mind the balance between the rate of convergence and the norm, we will first determine the rate of convergence in the total variation norm by using the criterion (1.6) and then deduce intermediate rates of convergence in f norm using an interpolation technique.

The rest of the paper is organized as follows. Our main result, Theorem 2.8, is stated and proved in the next section. Several typical functions ϕ are then considered, leading to a variety of subgeometric rate functions. In particular, by setting $\phi(v) := v^\alpha$, $\alpha \in [0, 1)$, we retrieve the results in Jarner and Roberts (2002). Several applications are given in section 3. We establish subgeometric rates of convergence in several models: first order nonlinear autoregressive models, stochastic unit root models, and random walk multidimensional Hastings Metropolis algorithm, under conditions which do not imply geometric ergodicity.

2 Main result

2.1 Rate of convergence in the total variation norm

Let $\phi : [1, \infty) \to (0, \infty)$ be a concave non decreasing differentiable function. Define

$$H_\phi(v) := \int_1^v \frac{dx}{\phi(x)}. \tag{2.1}$$

Then H_ϕ is a non decreasing concave differentiable function on $[1, \infty)$. Moreover, since ϕ is concave, ϕ' is non increasing. Hence $\phi(v) \leq \phi(1) + \phi'(1)(v-1)$ for all $v \geq 1$, which implies that H_ϕ increases to infinity. We can thus define its inverse $H_\phi^{-1} : [0, \infty) \to [1, \infty)$, 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

$$\begin{aligned} r_\phi(z) &:= (H_\phi^{-1})'(z) = \phi \circ H_\phi^{-1}(z), \qquad (2.2)\\ H_k(v) &:= \int_0^{H_\phi(v)} r_\phi(z+k)\, dz = H_\phi^{-1}(H_\phi(v)+k) - H_\phi^{-1}(k),\\ V_k &:= H_k \circ V. \end{aligned}$$

We will show that, provided $\mathbf{D}(\phi, C, V)$ holds with C petite and $\sup_{x \in C} V(x) < \infty$, then the chain $(\Phi_k, k \geq 0)$ is $(1, r_\phi)$ regular, *i.e.* r_ϕ is the rate of convergence in total variation norm that can be deduced from the drift condition. To this end, we will use Theorem 1.1 condition (ii), i.e., we will show that (1.6) holds with $(V_k, k \geq 0)$, $f := 1$ and $r := r_\phi$.

Proposition 2.1. *Assume $\mathbf{D}(\phi, V, C)$. Then r_ϕ is log concave and for all $k \geq 0$, H_k is concave and*

$$PV_{k+1} \leq V_k - r_\phi(k) + \frac{b r_\phi(k+1)}{\phi(1)} \mathbf{1}_C.$$

Proof. Note first that $r_\phi'(z)/r_\phi(z) = \phi' \circ H_\phi^{-1}(z)$ for all $z \geq 0$. Since ϕ' is non increasing and H_ϕ^{-1} is increasing, $\phi' \circ H_\phi^{-1}$ is non increasing and $\log(r_\phi)$ is concave. This implies that for any fixed $k \geq 0$, the function $z \mapsto r_\phi(z+k)/r_\phi(z)$ is a decreasing function. The derivative of H_k has the following expression

$$H_k'(v) = r_\phi(H_\phi(v)+k)/\phi(v) = r_\phi(H_\phi(v)+k)/r_\phi(H_\phi(v)). \tag{2.3}$$

Since H_ϕ is increasing, it follows from the discussion above that H_k' is non increasing, hence H_k is concave for all $k \geq 0$. Applying (2.3) and the fact that r_ϕ is increasing, we obtain:

$$\begin{aligned} H_{k+1}(v) - H_k(v) &= \int_0^{H_\phi(v)} \{r_\phi(z+k+1) - r_\phi(z+k)\}\, dz = \int_0^{H_\phi(v)} \int_0^1 r_\phi'(z+k+s)\, ds dz\\ &= \int_0^1 \{r_\phi(H_\phi(v)+k+s) - r_\phi(k+s)\}\, ds\\ &\leq r_\phi(H_\phi(v)+k+1) - r_\phi(k) = \phi(v) H_{k+1}'(v) - r_\phi(k). \end{aligned}$$

We have thus shown the following inequality which is the key tool of the proof.

$$H_{k+1}(v) - \phi(v) H'_{k+1}(v) \le H_k(v) - r_\phi(k). \tag{2.4}$$

Let g be a concave differentiable function on $[1, \infty)$. Since g' is decreasing, for all $v \ge 1$ and $x \in \mathbb{R}$ such that $v + x \ge 1$, it holds that

$$g(v + x) \le g(v) + g'(v)x. \tag{2.5}$$

Applying this property to the concave function H_{k+1}, we obtain for all $k \ge 0$, $x \in \{V < \infty\}$,

$$\begin{aligned} PV_{k+1}(x) &\le H_{k+1}\{V(x) - \phi \circ V(x) + b\mathbf{1}_C(x)\} \\ &\le H_{k+1}(V(x)) - \phi \circ V(x) H'_{k+1}(V(x)) + bH'_{k+1}(V(x))\mathbf{1}_C(x) \\ &\le H_{k+1}(V(x)) - \phi \circ V(x) H'_{k+1}(V(x)) + bH'_{k+1}(1)\mathbf{1}_C(x). \end{aligned}$$

Applying (2.3) and (2.4), we obtain that $H'_{k+1}(1) = r_\phi(k+1)/\phi(1)$ and

$$PV_{k+1}(x) \le V_k(x) - r_\phi(k) + \frac{br_\phi(k+1)}{\phi(1)} \mathbf{1}_C(x).$$

This inequality still holds for $x \in \{V = \infty\}$. Which concludes the proof. □

The drift condition $\mathbf{D}(\phi, V, C)$ and Proposition 2.1 imply the following bounds for the modulated moments of the return time τ_C, by application of Dynkin's inequality (see MT, Theorem 11.3.2).

Proposition 2.2. *Assume $\boldsymbol{D}(\phi, V, C)$. Then, for all $x \in \mathsf{X}$,*

$$\begin{aligned} \mathbb{E}_x \left[\sum_{k=0}^{\tau_C - 1} \phi \circ V(\Phi_k) \right] &\le V(x) + b\mathbf{1}_C(x), \\ \mathbb{E}_x \left[\sum_{k=0}^{\tau_C - 1} r_\phi(k) \right] &\le V(x) + \frac{br_\phi(1)}{\phi(1)} \mathbf{1}_C(x). \end{aligned}$$

In order to apply Theorem 1.1 we must also check the following conditions:

- the rate sequence $r_\phi := (\phi \circ H_\phi^{-1}(k), k \ge 0)$ belongs to Λ,
- the drift function V is bounded on C, $\sup_{x \in C} V(x) < \infty$.

The next Lemma gives a simple criterion to check that $r_\phi \in \Lambda$.

Lemma 2.3. *If $\lim_{t \to \infty} \phi'(t) = 0$, then $r_\phi \in \Lambda$.*

Proof. We have already noted that $r'_\phi(x)/r_\phi(x) = \phi' \circ H_\phi^{-1}(x)$ for all $x \geq 0$. Let r be any differentiable function such that $r(0) = 1$ and $\lim_{x\to\infty} r'(x)/r(x) = 0$. Then, applying Cesaro's Lemma, we obtain:

$$\frac{\log(r(n))}{n} = \frac{1}{n}\int_0^n \frac{r'(s)}{r(s)}\, ds \to 0.$$

If moreover r'/r decreases, then $\log(r(x))/x$ also decreases. Thus $r_\phi \in \Lambda$. $\square$

The condition $\sup_{x\in C} V(x) < \infty$ can easily be avoided, thanks to the following Lemma, adapted from Theorem 14.2.6 of MT.

Lemma 2.4. *Assume that $\boldsymbol{D}(\phi, V, C)$ holds for some petite set C and that $\lim_{v\to\infty}\phi(v) = \infty$. Then for all $M \geq 1$, the sublevel sets $\{x \in \mathsf{X}, V(x) \leq M\}$ are petite. In addition, for any β, $0 < \beta < 1$, there exists a sublevel set C_β such that $\boldsymbol{D}(\beta\phi, V, C_\beta)$ holds.*

Proof. Since ϕ is positive non decreasing and $V \geq 1$, the condition $\mathbf{D}(\phi, V, C)$ implies the drift condition $PV \leq V - \phi(1) + b\mathbf{1}_C$. Theorem 11.3.11 of MT shows that, for all accessible set $B \in \mathcal{B}^+(\mathsf{X})$, there exists a constant $c(B) < \infty$ such that, for all $x \in \mathsf{X}$ we have $\phi(1)\mathbb{E}_x[\tau_B] \leq V(x) + c(B)$. Hence, every set $A \in \mathcal{B}(\mathsf{X})$ such that $\sup_{x\in A} V(x) < \infty$ is regular, and the sublevel sets are all regular. Proposition 11.3.8 of MT shows that if a set A is regular, then it is petite. Hence, all the sublevel sets are petite.

Since $\lim_{v\to\infty}\phi(v) = \infty$, for all $\beta \in (0,1)$, there exists M_β such that $v > M_\beta$ implies $\phi(v) \geq b/(1-\beta)$. For $x \notin C_\beta := \{V \leq M_\beta\}$, we thus have $b \leq (1-\beta)\phi(V(x))$ and

$$PV + \beta\phi(V) \leq V + (\beta - 1)\phi(V) + b\mathbf{1}_C \leq V.$$

For $x \in C_\beta$, since $\beta \in (0,1)$, it trivially holds that

$$PV + \beta\phi(V) \leq V + b.$$

$\square$

Theorem 2.5. *Let P be a ψ irreducible and aperiodic kernel. Assume that $\boldsymbol{D}(\phi, V, C)$ holds for a function ϕ such that $\lim_{t\to\infty}\phi'(t) = 0$ and a petite set C such that $\sup_C V < \infty$. Then, there exists an invariant probability measure π, and for all x in the full and absorbing set $\{V < \infty\}$,*

$$\lim_n \; r_\phi(n) \quad \|P^n(x,\cdot) - \pi(\cdot)\|_{TV} = 0.$$

Any probability measure λ such that $\lambda(V) < \infty$ is $(1, r_\phi)$ regular and for two $(1, r_\phi)$ regular distributions λ, ν, there exists a constant c such that

$$\sum_{n=0}^{\infty} r_\phi(n) \int\!\!\int \lambda(dx)\mu(dy)\|P^n(x,\cdot) - P^n(y,\cdot)\|_{TV} \leq c\,(\lambda(V) + \mu(V))\,.$$

Remark 1. Since ϕ' is non increasing, if we do not assume that $\lim_{v\to\infty} \phi'(v) = 0$, then there exists $c \in (0,1)$ such that $\lim_{v\to\infty} \phi'(v) = c > 0$. This yields $v - \phi(v) \leq (1-c)v + c - \phi(1)$. In this case, condition $\mathbf{D}(\phi, V, C)$ implies the Foster Lyapunov drift condition, and the chain is V geometrically ergodic.

Proof of Theorem 2.5. The only statement which requires a proof is the fact that any probability measure such that $\lambda(V) < \infty$ is $(1, r_\phi)$ regular. This assertion is established in (Tuominen and Tweedie, 1994, Proposition 3.1.(ii)), and relies on (Nummelin and Tuominen, 1983, Lemma 3.1.). We nevertheless propose a proof that drastically shortens the previous one. The proof is adapted from the proof of Theorem 14.2.3 of MT. Proposition 2.1 shows that there exist a sequence of drift functions $(V_k, k \geq 0)$ and a constant b such that $V_0 \leq V$ and

$$PV_{k+1} \leq V_k - r_\phi(k) + b\phi(1)^{-1} r_\phi(k+1)\mathbf{1}_C.$$

Dynkin's formula shows that for all accessible set B,

$$\mathbb{E}_x\left[\sum_{k=0}^{\tau_B - 1} r_\phi(k)\right] \leq V_0(x) + b\ \phi(1)^{-1}\ \mathbb{E}_x\left[\sum_{k=0}^{\tau_B-1} r_\phi(k+1)\mathbf{1}_C(\Phi_k)\right].$$

¿From Propositions 5.5.5 and 5.5.6 of MT, we can assume without loss of generality that C is ψ_a petite, where ψ_a is equivalent to ψ, and that the sampling distribution a has finite mean $m_a := \sum_{j=1}^\infty j a_j < \infty$. By the Comparison Theorem (MT, Theorem 14.2.2), the bound $\mathbf{1}_C(x) \leq \psi_a(B)^{-1} K_a(x, B)$ and the fact that r_ϕ is non decreasing, we have:

$$\begin{aligned}
\mathbb{E}_x\left[\sum_{k=0}^{\tau_B - 1} r_\phi(k)\right] &\leq V_0(x) + b\ \phi(1)^{-1}\ \mathbb{E}_x\left[\sum_{k=0}^{\tau_B-1} r_\phi(k+1)\mathbf{1}_C(\Phi_k)\right] \\
&\leq V_0(x) + b\ \phi(1)^{-1}\ \psi_a(B)^{-1} \sum_{i\geq 0} a_i \mathbb{E}_x\left[\sum_{k=0}^{\tau_B-1} r_\phi(k+1)\mathbf{1}_B(\Phi_{k+i})\right] \\
&\leq V_0(x) + b\ \phi(1)^{-1}\ \psi_a(B)^{-1} m_a \mathbb{E}_x[r_\phi(\tau_B)].
\end{aligned}$$

For $k \geq 1$, define $R_\phi(k) := \sum_{j=0}^{k-1} r_\phi(j)$. Since r_ϕ is subgeometric, it holds that $\lim_{k\to\infty} r_\phi(k)/R_\phi(k) = 0$. Hence, for any $\delta > 0$, there exists a constant $c(\delta)$ such that for all $k \geq 1$, $r_\phi(k) \leq \delta R_\phi(k) + c(\delta)$. This yields:

$$\mathbb{E}_x\left[R_\phi(\tau_B)\right] \leq V_0(x) + b\phi(1)^{-1}\ \psi_a(B)^{-1} m_a \left(\delta \mathbb{E}_x\left[R_\phi(\tau_B)\right] + c(\delta)\right).$$

Thus for small enough δ, we obtain

$$\mathbb{E}_x\left[R_\phi(\tau_B)\right] \leq \frac{V_0(x) + b m_a \psi_a^{-1}(B) c(\delta)\phi(1)^{-1}}{1 - b\delta m_a \psi_a^{-1}(B)\phi(1)^{-1}}. \tag{2.6}$$

□

2.2 Rate of convergence in f norms

As seen in the polynomial case and discussed in Tuominen and Tweedie (1994), in the subgeometric case there is a compromise between the rate of convergence and the control function. In what follows, we will show that it is possible at almost no cost to obtain many intermediate different rates of convergence and control functions. Let $\mathcal{Y}$ be the set of pairs of ultimately non decreasing functions Ψ_1 and Ψ_2 defined on $[1,\infty)$ such that $\lim_{x\to\infty}\Psi_1(x)=\infty$ or $\lim_{x\to\infty}\Psi_2(x)=\infty$ and for all $x,y\in[1,\infty)$,

$$\Psi_1(x)\Psi_2(y)\leq x+y. \tag{2.7}$$

The set $\mathcal{Y}$ includes for example $\Psi_1(x)=x$ and $\Psi_2(x)=1$, but there are of course more interesting examples. For example, it is well known that, for any $x,y\geq 0$, and p and q such that $1/p+1/q=1$ we have

$$xy\leq x^p/p+y^q/q.$$

Hence, the pair of functions $\Psi_1(x)=p^{1/p}x^{1/p}$, $\Psi_2(x)=q^{1/q}x^{1/q}$ satisfies (2.7). These are precisely the interpolating functions used in Jarner and Roberts (2002) to derive polynomial rates of convergence. Young functions provide many useful interpolating functions. We recall their definition. Let $\varrho_1:(0,\infty)\to(0,\infty)$ be an increasing left continuous function such that $\varrho_1(0)=0$ and $\lim_{v\to+\infty}\varrho_1(v)=+\infty$. Let ϱ_2 be the left continuous inverse of ϱ_1, which is increasing and satisfies also $\varrho_2(0)=0$ and $\lim_{v\to+\infty}\varrho_2(v)=+\infty$. Define then $G_i(x):=\int_0^x\varrho_i(t)dt$, $i=1,2$. The well known Young inequality states that, for all $x,y\geq 0$, we have

$$xy\leq G_1(x)+G_2(y). \tag{2.8}$$

Let Ψ_i be the inverse of G_i, $i=1,2$. Then Ψ_1 and Ψ_2 are concave functions and it follows immediately from (2.8) that the pair (Ψ_1,Ψ_2) satisfies (2.7).

We use this full scale of interpolating functions in combination with Proposition 2.2 to derive bounds for the modulated moment of return time to the set C. More precisely, we have

Proposition 2.6. *Assume* $\boldsymbol{D}(\phi,V,C)$ *and let* $(\Psi_1,\Psi_2)\in\mathcal{Y}$. *Then*

$$\mathbb{E}_x\left[\sum_{k=0}^{\tau_C-1}\Psi_1(r_\phi(k))\Psi_2(\phi\circ V(\Phi_k))\right]\leq 2V(x)+b(1+r_\phi(1)/\phi(1))\mathbf{1}_C(x).$$

We need a criterion for a rate function $\Psi_1\circ r_\phi$ to be subgeometric. Note that if the pair (Ψ_1,Ψ_2) belongs to $\mathcal{Y}$, then, for large enough x, it holds that $\Psi_i(x)\leq 2x$ $(i=1,2)$.

Lemma 2.7. *Assume that* $\lim_{t\to\infty}\phi'(t)=0$. *For any non decreasing function* Ψ *such that* $\Psi(x)\leq ax$ *for some constant* a, *then* $\Psi\circ r_\phi\in\Lambda_0$.

The next theorem summarizes all our previous results.

Theorem 2.8. *Let* P *be a* ψ *irreducible and aperiodic kernel. Assume that* $\boldsymbol{D}(\phi,V,C)$ *holds for a function* ϕ *such that* $\lim_{t\to\infty}\phi'(t)=0$ *and a petite set* C *such that* $\sup_C V<\infty$. *Let*

$(\Psi_1, \Psi_2) \in \mathcal{Y}$. *Then, there exists an invariant probability measure* π, *and for all* x *in the full set* $\{V < \infty\}$,

$$\lim_n \Psi_1(r_\phi(n)) \quad \|P^n(x,\cdot) - \pi(\cdot)\|_{\Psi_2(\phi\circ V)} = 0.$$

Any probability measure λ *such that* $\lambda(V) < \infty$ *is* $(\Psi_2(\phi \circ V), \Psi_1(r_\phi))$ *regular and for two such distributions* λ, μ, *there exists a constant* c *such that*

$$\sum_{n=0}^{\infty} \Psi_1(r_\phi(n)) \int\int \lambda(dx)\mu(dy) \|P^n(x,\cdot) - P^n(y,\cdot)\|_{\Psi_2(\phi\circ V)} \leq c\,(\lambda(V) + \mu(V)).$$

Proof. From Proposition 2.6 we have

$$\sup_{x\in C} \mathbb{E}_x \left[\sum_{k=0}^{\tau_C-1} \Psi_1(r_\phi(k)) \; \Psi_2(\phi\circ V(\Phi_k)) \right] < \infty.$$

Theorem 1.1 shows that Φ is $(\Psi_2(\phi\circ V), \Psi_1(r_\phi))$ regular. As in the proof of Theorem 2.5, and using again the Comparison Theorem, for any set $B \in \mathcal{B}^+(\mathsf{X})$, there exist constants $c_1(B)$ and $c_2(B)$ such that

$$\mathbb{E}_x \left[\sum_{k=0}^{\tau_B-1} \phi\circ V(\Phi_k) \right] + \mathbb{E}_x \left[\sum_{k=0}^{\tau_B-1} r_\phi(k) \right] \leq c_1(B)V(x) + c_2(B).$$

Hence, for any $(\Psi_1, \Psi_2) \in \mathcal{Y}$, we have

$$\mathbb{E}_x \left[\sum_{k=0}^{\tau_B-1} \Psi_1(r_\phi(k)) \; \Psi_2(\phi\circ V(\Phi_k)) \right] \leq c_1(B)V(x) + c_2(B),$$

which shows that any probability measure such that $\lambda(V) < \infty$ is $(\Psi_2(\phi\circ V), \Psi_1(r_\phi))$ regular. $\square$

2.3 Some usual rate functions

In this section, we provide examples of rates of convergence obtained by Theorem 2.8. For two sequences u_n and v_n, we write $u_n \asymp v_n$ if there exists positive constants c_1 and c_2 such that for large n, $c_1 u_n \leq v_n \leq c_2 u_n$.
We assume throughout this section that the condition $\mathbf{D}(\phi, V, C)$ holds for some petite set C and $\sup_C V < \infty$.

Polynomial rates of convergence Polynomial rates of convergence have been widely studied recently under various conditions (see Veretennikov (1997, 1999), Tanikawa (2001), Jarner and Roberts (2002), Fort and Moulines (2002)). As already mentioned, polynomial rates of convergence are associated to the functions $\phi(v) := cv^\alpha$ for some $\alpha \in [0,1)$ and $c \in (0,1]$ and the rate of convergence in total variation distance is $r_\phi(n) \propto n^{\alpha/(1-\alpha)}$. Set $\Psi_1(x) := ((1-p)x)^{(1-p)}$ and $\Psi_2(x) := (px)^p$ for some p, $0 < p < 1$. Applying Theorem 2.8 yields, $x \in \{V < \infty\}$,

$$\lim_n \; n^{(1-p)\alpha/(1-\alpha)} \quad \|P^n(x, \cdot) - \pi(\cdot)\|_{V^{\alpha p}} = 0. \tag{2.9}$$

This convergence remains valid for $p = 0, 1$ by Proposition 2.2. Set $\kappa := 1 + (1-p)\alpha/(1-\alpha)$ so that $1 \leq \kappa \leq 1/(1-\alpha)$. With these notations (2.9) reads

$$\lim_n \; n^{\kappa - 1} \quad \|P^n(x, \cdot) - \pi(\cdot)\|_{V^{1-\kappa(1-\alpha)}} = 0,$$

which is the result given in (Jarner and Roberts, 2002, Theorem 3.6).

It is possible to extend this result by using more general interpolation functions. We can for example obtain non polynomial rates of convergence with control functions which are not simply power of the drift functions. To illustrate this point, set for $b > 0$, $\Psi_1(x) := (1 \vee \log(x))^b$ and $\Psi_2(x) := x(1 \vee \log(x))^{-b}$. It is not difficult to check that we have

$$\sup_{(x,y) \in [1,\infty) \times [1,\infty)} (x+y)^{-1} \Psi_1(x) \Psi_2(y) < \infty,$$

so that, for all $x \in \{V < \infty\}$, we have

$$\lim_n \; \log^b(n) \quad \|P^n(x, \cdot) - \pi(\cdot)\|_{V^\alpha (1+\log(V))^{-b}} = 0, \tag{2.10}$$

$$\lim_n \; n^{\alpha/(1-\alpha)} \log^{-b}(n) \; \|P^n(x, \cdot) - \pi(\cdot)\|_{(1+\log(V))^b} = 0, \tag{2.11}$$

and for all $0 < p < 1$,

$$\lim_n \; n^{(1-p)\alpha/(1-\alpha)} \log^b n \quad \|P^n(x, \cdot) - \pi(\cdot)\|_{V^{\alpha p}(1+\log V)^{-b}} = 0.$$

Logarithmic rates of convergence We now consider drift conditions which imply rates of convergence slower than any polynomial. Such rates are obtained when condition $D(\phi, V, C)$ holds with a function ϕ that increases to infinity slower than polynomially. We only consider here the case $\phi(v) = c(1 + \log(v))^\alpha$ for some $\alpha \geq 0$ and $c \in (0, 1]$. A straightforward calculation shows that

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

Theorem 2.5 shows that the chain is $(1, \log^\alpha(n))$ and $((1 + \log V)^\alpha, 1)$ regular. Applying Theorem 2.8, intermediate rate can be obtained along the same lines as above. Choosing for instance $\Psi_1(x) := ((1-p)x)^{1-p}$ and $\Psi_2(x) := (px)^p$ for $0 \leq p \leq 1$, then the chain is $((1 + \log V)^{p\alpha}, \log(n)^{(1-p)\alpha})$ regular and thus for all $x \in \{V < \infty\}$,

$$\lim_{n\to\infty} (1 + \log(n))^{(1-p)\alpha} \; \|P^n(x, \cdot) - \pi(\cdot)\|_{(1+\log(V))^{p\alpha}} = 0.$$

Subexponential rates of convergence Subexponential rates (as defined in (1.5)) have been considered only recently in the literature. An example (in continuous time) has been studied by Malyshkin (2001); discrete time examples are considered in the recent work by Klokov and Veretennikov (2002). These rates, which increase to infinity faster than polynomially, are obtained when the condition $\mathbf{D}(\phi, V, C)$ holds with ϕ such that $v/\phi(v)$ goes to infinity slower than polynomially. More precisely, assume that ϕ 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

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

and thus the chain is $(1, n^{-\alpha/(1+\alpha)} \exp(\{c(1+\alpha)n\}^{1/(1+\alpha)}))$ and $(V/(1+\log V)^\alpha, 1)$ regular. Applying Theorem 2.8 with $\Psi_1(x) := x^{1-p}(1 \vee \log(x))^{-b}$ and $\Psi_2(x) := x^p(1 \vee \log(x))^b$ for $p \in (0,1)$ and $b \in \mathbb{R}$, $p = 0$ and $b > 0$ or $p = 1$ and $b < -\alpha$ yields, for all $x \in \{V < \infty\}$,

$$\lim_n \; n^{-(\alpha+b)/(1+\alpha)} \exp\left((1-p)\{c(1+\alpha)n\}^{1/(1+\alpha)}\right) \|P^n(x,\cdot) - \pi(\cdot)\|_{V^p(1+\log V)^b} = 0. \quad (2.12)$$

3 Applications

We now illustrate our findings by applying Theorem 2.8 to several models.

In this section, we denote by $\langle \cdot, \cdot \rangle$ the scalar product and by $|\cdot|$ the Euclidean norm. If u is a twice continuously differentiable real valued function on $\mathbb{R}^d$, ∇u (resp. $\nabla^2 u$) denotes its gradient (resp. its Hessian matrix).

3.1 Backward recurrence time chain

The backward recurrence time chain (see MT, Section 3.3.1) is a rich source of simple examples of stable and unstable behavior. We consider it here to provide examples of chains satisfying condition $\mathbf{D}(\phi, V, C)$ and for which the rates of convergence implied by it are optimal.

Let $(p_n, n \geq 0)$ be a sequence of positive real numbers such that $p_0 = 1$, $p_n \in (0,1)$ for all $n \geq 1$ and $\lim_{n\to\infty} \prod_{i=1}^n p_i = 0$. Consider the backward recurrence time chain Φ with transition kernel P defined as $P(n, n+1) = 1 - P(n, 0) = p_n$, for all $n \geq 0$. Then Φ is irreducible and strongly aperiodic and $\{0\}$ is an atom. Let τ_0 be the return time to $\{0\}$. We have for all $n \geq 1$

$$\mathbb{P}_0(\tau_0 = n+1) = (1-p_n) \prod_{j=0}^{n-1} p_j \quad \text{and} \quad \mathbb{P}_0(\tau_0 > n) = \prod_{j=0}^{n-1} p_j,$$

By Kac's theorem (MT, Theorem 10.2.2) since Φ is ψ irreducible and aperiodic, Φ is positive recurrent if and only if $\mathbb{E}_0[\tau_0] < \infty$, *i.e.*

$$\sum_{n=1}^\infty \prod_{j=1}^n p_j < \infty,$$

and the stationary distribution π is given, by $\pi(0) = \pi(1) = 1/\mathbb{E}_0[\tau_0]$ and for $j \geq 2$,

$$\pi(j) = \frac{\mathbb{E}_0\left[\sum_{k=1}^{\tau_0} \mathbf{1}\{\Phi_k = j\}\right]}{\mathbb{E}_0[\tau_0]} = \frac{\mathbb{P}_0(\tau_0 \geq j)}{\mathbb{E}_0[\tau_0]} = \frac{p_0 \dots p_{j-2}}{\sum_{n=1}^{\infty} p_1 \dots p_n}.$$

Because the distribution of the return time to the atom $\{0\}$ has such a simple expression in terms of the transition probability $(p_n, n \geq 0)$, we are able to exhibit the largest possible rate function r such that the $(1, r)$ modulated moment of the return time $\mathbb{E}_0\left[\sum_{k=0}^{\tau_0-1} r(k)\right]$ is finite. We will also prove that the drift condition $\mathbf{D}(\phi, V, C)$ holds for appropriately chosen functions V and ϕ and yields the optimal rate of convergence. Note also that for any function f, it holds that

$$\mathbb{E}_0\left[\sum_{k=0}^{\tau_0-1} f(\Phi_k)\right] = \mathbb{E}_0\left[\sum_{k=0}^{\tau_0-1} f(k)\right].$$

Therefore there is no loss of generality to consider only $(1, r)$ modulated moments of the return time to zero.

If $\sup_{n\geq 1} p_n \leq \lambda < 1$, then, for all $\lambda < \mu < 1$, $\mathbb{E}_0[\mu^{-\tau_0}] < \infty$ and $\{0\}$ is thus a geometrically ergodic atom (MT, Theorem 15.1.5). Subgeometric rates of convergence in total variation norm are obtained when $\limsup p_n = 1$. Depending on the rate at which p_n approaches 1, different behaviors can be obtained, covering essentially the three typical rates (polynomial, logarithmic and subexponential) discussed above.

Polynomial rates Assume first that for $\theta > 0$ and large n, $p_n = 1 - (1+\theta)n^{-1}$. Then $\prod_{i=1}^{n} p_i \asymp n^{-1-\theta}$. Thus, $\mathbb{E}_0\left[\sum_{k=0}^{\tau_0-1} r(k)\right] < \infty$ if and only if $\sum_{k=1}^{\infty} r(k)k^{-1-\theta} < \infty$. For instance, $r(n) := n^{\beta}$ with $0 \leq \beta < \theta$ is suitable.

Logarithmic rates If for $\theta > 0$ and large n, $p_n = 1 - 1/n - (1+\theta)/(n\log(n))$, then $\prod_{j=1}^{n} p_j \asymp n^{-1}\log^{-1-\theta}(n)$, which is a summable series. Hence if r is non decreasing and $\sum_{k=1}^{\infty} r(k)\prod_{j=1}^{n} p_j < \infty$, then $r(k) = o(\log^{\theta}(k))$. In particular $r(k) := \log^{\beta}(k)$ is suitable for all $0 \leq \beta < \theta$.

Subgeometric rates If for large n, $p_n = 1 - \theta\beta n^{\beta-1}$ for some $\theta > 0$ and $\beta \in (0,1)$, then $\prod_{i=1}^{n} p_i \asymp \mathrm{e}^{-\theta n^{\beta}}$. Thus, $\mathbb{E}_0[\sum_{k=0}^{\tau_0-1} \mathrm{e}^{ak^{\beta}}] < \infty$ if $a < \theta$, and $\mathbb{E}_0[\sum_{k=0}^{\tau_0-1} \mathrm{e}^{ak^{\beta}}] = \infty$ if $a \geq \theta$.

Checking condition $\mathbf{D}(\phi, V, C)$ In order to prove that Theorem 2.5 provides the optimal rates of convergence, we now compute in each of the previous examples the rates of convergence it yields.

For the polynomial and subexponential cases, the same technique can be used. For $\gamma \in (0,1)$ and $x \in \mathbb{N}^*$, define $V(0) := 1$ and $V(x) := \prod_{j=0}^{x-1} p_j^{-\gamma}$. Then, for all $x \geq 0$, we have:

$$\begin{aligned} PV(x) &= p_x V(x+1) + (1-p_x)V(0) = p_x^{1-\gamma} V(x) + 1 - p_x \\ &\leq V(x) - (1 - p_x^{1-\gamma})V(x) + 1 - p_x \end{aligned}$$

Thus, for $0 < \delta < 1 - \gamma$ and large enough x, it holds that

$$PV(x) \leq V(x) - \delta(1-p_x)V(x). \tag{3.1}$$

• Case $p_n = 1 - (1+\theta)n^{-1}$, $\theta > 0$. Then $V(x) \asymp x^{\gamma(1+\theta)}$ and $(1-p_x)V(x) \asymp V(x)^{1-1/(\gamma(1+\theta)}$. Thus condition $\mathbf{D}(\phi, V, C)$ holds with $\phi(v) = cv^\alpha$ for $\alpha = 1 - 1/(\gamma(1+\theta))$ for any $\gamma \in (0,1)$. Theorem 2.8 yields the rate of convergence $n^{\alpha/(1-\alpha)} = n^{\gamma(1+\theta)-1}$, *i.e.* n^β for any $0 \leq \beta < \theta$.

• Case $p_n = 1 - \theta\beta n^{\beta-1}$. Then, for large enough x, (3.1) yields:

$$PV(x) \leq V(x) - \theta\beta\delta x^{\beta-1} V(x) \leq cV(x)\{\log(V(x))\}^{1-1/\beta},$$

for $c < \theta^{1/\beta}\beta\delta$. Defining $\alpha := 1/\beta - 1$, Proposition 2.1 yields the following rate of convergence in total variation norm:

$$n^{-\alpha/(1+\alpha)} \exp\left(\{c(1+\alpha)n\}^{1/(1+\alpha)}\right) = n^{\beta-1}\exp\left(\theta\delta^\beta n^\beta\right).$$

Since δ is arbitrarily close to 1, we recover the fact that $\mathbb{E}_0[\sum_{k=0}^{\tau_0-1} \mathrm{e}^{ak^\beta}] < \infty$ for any $a < \theta$.

• Case $p_n = 1 - n^{-1} - (1+\theta)n^{-1}\log^{-1}(n)$, $\theta > 0$. Choose $V(x) := \left(\prod_{j=0}^{x-1} p_j\right) / \log^\epsilon(x)$ for $\epsilon > 0$ arbitrarily small. Then, for constants $c < c' < c'' < 1$ and large x, we obtain:

$$\begin{aligned} PV(x) &= \frac{\log^\epsilon(x)}{\log^\epsilon(x+1)} V(x) + 1 - p_x = V(x) - c''\epsilon \frac{V(x)}{x\log(x)} + 1 - p_x \\ &\leq V(x) - c'\epsilon \log^{\theta-\epsilon}(x) \leq V(x) - c\epsilon\log^{\theta-\epsilon}(V(x)). \end{aligned}$$

Here again Theorem 2.8 yields the optimal rate of convergence.

3.2 Symmetric random walk Hastings Metropolis algorithm

We consider the symmetric random walk Hastings Metropolis algorithm. The purpose of this algorithm is to simulate from a probability distribution π which is known only up to a scale factor. At each iteration, a move is proposed according to a random walk whose increment distribution has a symmetric density q with respect to the Lebesgue measure μ_d on $\mathbb{R}^d$. The move is accepted with probability $\alpha(x,y)$ defined by

$$\alpha(x,y) := \begin{cases} \min\left\{\frac{\pi(y)}{\pi(x)}, 1\right\} & \text{if } \pi(x) > 0 \\ 1 & \text{if } \pi(x) = 0. \end{cases} \tag{3.2}$$

The transition kernel of the Metropolis algorithm is then given by

$$P(x, A) = \int_A \alpha(x, x+y) q(y)\, d\mu_d(y) + \mathbf{1}_A(x) \int \Big(1 - \alpha(x, x+y)\Big) q(y)\, d\mu_d(y).$$

It is known that under Assumption 3.1 below, the chain Φ is ψ irreducible, aperiodic with stationary distribution $\pi\ d\mu_d$ and any non empty compact set is petite (Roberts and Tweedie, 1996, Theorem 2.2.).

Assumption 3.1. *The target density π is continuous and positive on $\mathbb{R}^d$. The proposal density q is symmetric and bounded away from zero in a neighborhood of zero.*

A $\mathbb{R}$ valued Metropolis chain is V geometrically ergodic when (a) the proposal density q satisfies moment conditions and (b) the target density π is continuous, positive and log concave in the tails (Mengersen and Tweedie, 1996, Theorem 3.2.). This condition is necessary in the sense that if the chain is geometrically ergodic then $\int \exp(s|z|)\pi(z) d\mu_d(z) < \infty$ for some $s > 0$. These results have later been extended to the multidimensional case by Roberts and Tweedie (1996) and Jarner and Hansen (2000). Polynomial ergodicity was proved by Fort and Moulines (2000) for target density with regularly varying tails. We now state conditions that imply subexponential rates of convergence

Assumption 3.2. *there exist $m \in (0,1)$, $r \in (0,1)$, positive constants $d_i, D_i, i = 0,1,2$ and $R_0 < \infty$ such that if $|x| \geq R_0$, $x \mapsto \pi(x)$ is twice continuously differentiable and*

$$\langle \frac{\nabla \pi(x)}{|\nabla \pi(x)|}, \frac{x}{|x|} \rangle \ \leq -r, \tag{3.3}$$

$$d_0 |x|^m \leq -\log \pi(x) \leq D_0 |x|^m \tag{3.4}$$

$$d_1 |x|^{m-1} \leq |\nabla \log \pi(x)| \leq D_1 |x|^{m-1} \tag{3.5}$$

$$d_2 |x|^{m-2} \leq |\nabla^2 \log \pi(x)| \leq D_2 |x|^{m-2}. \tag{3.6}$$

The Weibull distribution on $\mathbb{R}$ with density $\pi(x) := \beta\gamma x^{\gamma-1} \exp(-\beta x^\gamma)$, for $x > 0$, $\beta > 0$ and $0 < \gamma < 1$ satisfies assumption 3.2. Multidimensional examples are provided in Fort and Moulines (2000). For the sake of simplicity, we make the following assumption on the proposal density q.

Assumption 3.3. *The proposal density is compactly supported, i.e. there exists c_k such that for all $|y| \geq c_k$, $q(y) = 0$.*

Fort and Moulines (2000) show that under Assumptions 3.1 and 3.2, the chain Φ is (f, r) ergodic with $f(x) := (1 + |x|^\mu)$ and $r(n) := (1+n)^\nu$, for any $\mu > 0$ and $\nu \geq 0$, *i.e.* Φ is (f, r) ergodic at any polynomial rate. We show in the next Theorem 3.1 that a stronger result actually holds: the chain is ergodic (in total variation norm) at a subgeometrical rate

$$r_\phi(n) \asymp \exp(z n^{m/(2-m)}),$$

for some $z > 0$; rate of convergence in norms $\pi^{-s}(x)(-\log \pi)^t(x)$ where $0 < s < z$ and $t \in \mathbb{R}$ (resp. $s = 0$ and $t > 0$; $s = z$ and $t < -2(1-m)/m$) are given by (2.12).

Theorem 3.1. *Under Assumptions 3.1 to 3.3, there exist $z > 0$ and $c > 0$ such that the functions $V(x) := \pi(x)^{-z}$ and $\phi(v) := cv(1 + \log v)^{-2(1-m)/m}$ satisfy the drift condition $\mathbf{D}(\phi, V, C)$ where C is a petite set such that $\sup_C V < \infty$.*

Remark 2. The compactness assumption 3.3 could be relaxed and replaced by a moment condition. It may be shown (the computations are not detailed here but are similar to those of Section 3.3) that if there exists $z_0 > 0$ such that

$$\int \mathrm{e}^{z_0 |y|^m} q(y) d\mu_d(y) < \infty, \tag{3.7}$$

then the conclusion of Theorem 3.1 still holds by choosing $V(x) := \pi(x)^{-z}$ for some $0 < z < z_0$.

Proof. Define $\mathcal{R}(x) := \{y \in \mathbb{R}^d, \pi(x+y) \leq \pi(x)\}$ the potential rejection region. Using the definition of the transition kernel P, we have

$$PV(x) - V(x) = \int (V(x+y) - V(x))\, q(y) d\mu_d(y) \\ + \int_{\mathcal{R}(x)} (V(x+y) - V(x)) \left(\frac{\pi(x+y)}{\pi(x)} - 1 \right) q(y) d\mu_d(y).$$

Set $l(x) := -\log \pi(x)$, $R(V, x, y) := V(x+y) - V(x) + zV(x)\langle \nabla l(x), y\rangle$ and $R(\pi, x, y) := \pi(x+y)/\pi(x) - 1 + \langle \nabla l(x), y \rangle$. It is proved in (Fort and Moulines, 2000, Lemma B.4.), that

$$\limsup_{|x| \to \infty} |x|^{2(1-m)} \sup_{|y| \leq c_k} |R(\pi, x, y)| |y|^{-2} < \infty. \tag{3.8}$$

Using a Taylor expansion with integral remainder term of the function $x \mapsto V(x)$, it is easily shown that there exists c independent on z such that for large $|x|$

$$\sup_{|y| \leq c_k} |R(V, x, y)| |y|^{-2} \leq c z^2 V(x) |x|^{2(m-1)} \left(1 + o(1)\right). \tag{3.9}$$

Since $q\ d\mu_d$ is a zero mean distribution, we have

$$PV(x) - V(x) = -zV(x) \int_{\mathcal{R}(x)} \langle \nabla l(x), y\rangle^2 q(y) d\mu_d(y) + \int R(V, x, y) q(y) d\mu_d(y) \\ - \int_{\mathcal{R}(x)} R(V, x, y) \langle \nabla l(x), y \rangle q(y) d\mu_d(y) \\ + zV(x) \int_{\mathcal{R}(x)} \langle \nabla l(x), y \rangle q(y) R(\pi, x, y) q(y) d\mu_d(y) + \int_{\mathcal{R}(x)} R(V, x, y) R(\pi, x, y) q(y) d\mu_d(y)$$

and for large $|x|$, we deduce from (3.8) and (3.9) that

$$\frac{PV(x) - V(x)}{V(x)} = -z \int_{\mathcal{R}(x)} \langle \nabla l(x), y \rangle^2 q(y) d\mu_d(y) + c z^2 |x|^{2(m-1)} + o(|x|^{2(m-1)}),$$

for some positive constant c that does not depend on z. It is shown in (Fort and Moulines, 2000, Lemma B.3.) (see below) that there exists $\eta > 0$ such that for large $|x|$,

$$\int_{\mathcal{R}(x)} \langle \nabla l(x), y\rangle^2 q(y) d\mu_d(y) > \eta |\nabla l(x)|^2 > \eta d_1^2 |x|^{2(m-1)}. \tag{3.10}$$

Hence, upon noting that $d_0|x|^m \leq \log V(x)$, there exists a constant κ which is positive for z small enough, such that for large $|x|$

$$PV(x) - V(x) \leq -\kappa \left[\log V(x)\right]^{-2(1-m)/m} \; V(x)\left(1 + o(1)\right).$$

Since π is bounded on compact sets, $\sup_{|x|\leq M} PV(x) + V(x) < \infty$ and the proof is concluded. □

Proof of (3.10). The proof is similar to that of (Fort and Moulines, 2000, Lemma B.3.). For completeness, we sketch the arguments here. Set $n(x) := x/|x|$ and

$$W(x) := \{z \in \mathbb{R}^d, z = x + a\xi, 0 < a \leq c_k, \xi \in S^{d-1}, |\xi - n(x)| \leq r/3\}.$$

We establish that there exists $c > 0$ such that

$$\begin{aligned}\int_{\mathcal{R}(x)} \langle \nabla l(x), y\rangle^2 q(y) d\mu_d(y) &\geq \int_{W(x)-x} \langle \nabla l(x), y\rangle^2 q(y) d\mu_d(y) \\ &\geq r^2/9 \; |\nabla l(x)|^2 \int_{W(x)-x} |y|^2 q(y) d\mu_d(y) \geq c \; |\nabla l(x)|^2,\end{aligned}$$

since the Lebesgue measure of the domain $W(x) - x$ does not depend on x.
We first prove that $W(x) - x \subset \mathcal{R}(x)$. To that goal, we establish that for $z \in W(x)$, the function $\phi(t) := \pi(x + t(z - x))$ defined on $[0, 1]$ is monotonically decreasing on $[x, z]$ by showing that $\langle n(z - x), n(\nabla\pi(y))\rangle >\leq 0$ for any $y \in [x, z]$. We write

$$\begin{aligned}&\langle n(z-x), n(\nabla\pi(y))\rangle \\ &\quad = \langle n(z-x) - n(x), n(\nabla\pi(y))\rangle + \langle n(x) - n(y), n(\nabla\pi(y))\rangle + \langle n(y), n(\nabla\pi(y))\rangle.\end{aligned}$$

By definition of $W(x)$, we have

$$|\langle n(z-x) - n(x), n(\nabla\pi(y))\rangle| \leq r/3 \qquad\qquad |\langle n(x) - n(y), n(\nabla\pi(y))\rangle| \leq r/3$$

so that $\langle n(z - x), n(\nabla\pi(y))\rangle \leq -r/3 < 0$. It remains to prove that for all $y + x \in W(x)$, $|\langle \nabla l(x), y\rangle| \geq r/3 |\nabla l(x)||y|$, which is deduced from the previous calculations applied with $y := z - x$ and $x := y$. □

3.3 Nonlinear autoregressive model

Consider a process $(\Phi_n, n \geq 0)$ that satisfies the following nonlinear autoregressive equation of order 1:

$$\Phi_{n+1} = g(\Phi_n) + \epsilon_{n+1}, \tag{3.11}$$

where the innovation and the function g satisfy the following assumption.

Assumption 3.4. $(\epsilon_n, n \geq 0)$ *is a sequence of i.i.d. zero mean, d dimensional random vectors that satisfy*

$$\mathbb{E}[\mathrm{e}^{z_0 |\epsilon_0|^{\gamma_0}}] < \infty, \tag{3.12}$$

for some $z_0 > 0$ *and* $\gamma_0 \in (0, 1]$*; g is bounded on the set* $\{x \in \mathbb{R}^d, |x| \leq R_0\}$ *for some* $R_0 > 0$ *and there exists* $\rho \in [0, 2)$ *such that*

$$|g(x)| \leq |x|(1 - r|x|^{-\rho}) \quad \text{if} \quad |x| \geq R_0. \tag{3.13}$$

There already exists a wide literature on on conditions implying a geometric rate of convergence for nonlinear autoregressive models (see *e.g.* Duflo (1997) and Grunwald et al. (2000) and the references therein). Conditions implying a polynomial rate of convergence have been obtained by Tuominen and Tweedie (1994) and AngoNze (1994) and have later been refined by Veretennikov (1997, 1999), AngoNze (2000) and Fort and Moulines (2002). Conditions implying truly subexponential rate of convergence are considered in Klokov and Veretennikov (2002) (see also Malyshkin (2001) for diffusion processes).

Theorem 3.2. *Assume that Assumption 3.4 holds.*

- *If* $\rho > \gamma_0$*, the drift condition* $\mathbf{D}(\phi, V, C)$ *holds with* $\phi(v) := cv(1 + \log(v))^{1-\rho/(\gamma_0 \wedge (2-\rho))}$, $V(x) := \mathrm{e}^{z|x|^{\gamma_0 \wedge (2-\rho)}}$ *and* $C := \{x \in \mathbb{R}^d, |g(x)| \leq M_1\} \cup \{x \in \mathbb{R}^d, |x| \leq M_2\}$ *for some* $z \in (0, z_0)$*,* $c > 0$*,* $M_1 > 0$ *and* $M_2 \geq R_0$*.*
- *If* $\rho = \gamma_0$*, then the Foster Lyapunov condition (1.3) holds with* $V(x) := \mathrm{e}^{z|x|^{\gamma_0}}$ *and* $C := \{x \in \mathbb{R}^d, |g(x)| \leq M_1\} \cup \{x \in \mathbb{R}^d, \ |x| \leq M_2\}$ *for some* $z \in (0, z_0)$*,* $M_1 > 0$ *and* $M_2 \geq R_0$*.*
- *If* $\rho < \gamma_0$*, then the Foster Lyapunov condition (1.3) holds with* $V(x) := \mathrm{e}^{z_0|x|^{\gamma_0}}$ *and* $C := \{x \in \mathbb{R}^d, |x| \leq M\}$ *for some* $M \geq R_0$*.*

If in addition the chain is ψ irreducible, aperiodic and sublevel sets of g and compact sets are petite, then we may apply Theorem 2.8 (resp. Theorem 15.0.1. MT) to prove (f, r) ergodicity or geometrical ergodicity, depending upon the value of ρ and γ_0. Conditions implying irreducibility, and aperiodicity of the kernel, and petiteness of the level sets $\{x, |x| \leq M_1\}$ and $\{x, |g(x)| \leq M_2\}$ may be found in Tuominen and Tweedie (1994).

Proof of Theorem 3.2. Throughout the proof, c is a generic constant that can change upon each appearance.
(i) We start by examining the case $\rho > \gamma_0$. Set $\beta := \gamma_0 \wedge (2 - \rho)$. We write

$$\frac{PV(x)}{V(x)} - 1 = \frac{PV(x) - V(g(x))}{V(x)} + \frac{V(g(x))}{V(x)} - 1. \tag{3.14}$$

Using the inequality $(1 - u)^{\gamma_0} \leq 1 - \gamma_0 u$ for all $0 \leq u \leq 1$, we have for $|x| \geq R_0$, $|g(x)|^\beta \leq |x|^\beta - \beta r|x|^{\beta-\rho}$ and since $\mathrm{e}^x - 1 \leq x + x^2/2$ for all $x \leq 0$,

$$\frac{V(g(x))}{V(x)} - 1 = \mathrm{e}^{z|g(x)|^\beta - z|x|^\beta} - 1 \leq -zr\beta|x|^{\beta-\rho} + \frac{1}{2}z^2r^2\beta^2|x|^{2(\beta-\rho)}. \tag{3.15}$$

Let $0<\eta<1$. We establish that for large $|x|$ and large $|g(x)|$,

$$PV(x) - V(g(x)) \leq \frac{1}{2} z^2 \beta^2 \ \mathbb{E}[|\epsilon_0|^2 V(\epsilon_0)] \ |x|^{2\beta-2} V(x) \ \ (1+o(1)) \,. \tag{3.16}$$

To that goal, set $R(u,w) := V(u+w) - V(u) - \langle \nabla V(u), w\rangle$. Since $\mathbb{E}[\epsilon_0]=0$, this yields

$$PV(x) - V(g(x)) = \mathbb{E}[V(g(x)+\epsilon_0)] - V(g(x)) = \mathbb{E}[R(g(x),\epsilon_0)] \,, \tag{3.17}$$

and we have to upper bound the remainder term $\mathbb{E}[R(g(x),\epsilon_0)]$. If $|w| \leq \eta |u|$, then by using a Taylor expansion with integral remainder term, one has,

$$\begin{aligned} |R(u,w)| &\leq \int_0^1 (1-t) \left| w' \nabla^2 V(u+tw) w \right| \, dt \\ &\qquad \leq \frac{1}{2} |w|^2 \ z\beta \sup_{t\in[0,1]} \left(1 + z\beta |u+tw|^\beta \right) |u+tw|^{\beta-2} \ V(u+tw) \end{aligned}$$

Since $y \mapsto |y|^{2\beta-2} \mathrm{e}^{z|y|^\beta}$ and $y \mapsto |y|^{\beta-2} \mathrm{e}^{z|y|^\beta}$ are ultimately nondecreasing, then for large $|x|$, we have:

$$\begin{aligned} |R(u,w)| &\leq \frac{1}{2} |w|^2 \ z\beta \ \left(1 + z\beta (|u|+|w|)^\beta \right) (|u|+|w|)^{\beta-2} \ V(u)V(w) \\ &\leq \frac{1}{2} \ z^2\beta^2 \ |w|^2 V(w) \ |u|^{2\beta-2} V(u) + +c \ |w|^2 V(w) \ |u|^{\beta-2} V(u). \qquad \text{if } |w| \leq \eta |u| \end{aligned} \tag{3.18}$$

If $|w| \geq \eta |u|$, using again the inequality $V(u+w) \leq V(u)V(w)$

$$\begin{aligned} |R(u,w)| &\leq V(u+w) + V(u) + |\nabla V(u)| \ |w| \leq c\,|w|\,V(w) \ |u|^{\beta-1} \, V(u) \\ &\leq c|w|^2 V(w) |u|^{\beta-2} \ V(u) \qquad \text{if } |w| \geq \eta |u|. \end{aligned} \tag{3.19}$$

We now apply (3.18) and (3.19) with $u := g(x)$ and $w := \epsilon_0$; since $y \mapsto |y|^{2\beta-2} \mathrm{e}^{z|y|^\beta}$ and $y \mapsto |y|^{\beta-2} \mathrm{e}^{z|y|^\beta}$ are ultimately nondecreasing, then for large $|g(x)|$, we have:

$$|R(g(x),\epsilon_0)| \leq \frac{1}{2} \ z^2\beta^2 \ |\epsilon_0|^2 V(\epsilon_0) \ |x|^{2\beta-2} V(x) + c \ |\epsilon_0|^2 V(\epsilon_0) \ |x|^{\beta-2} V(x). \tag{3.20}$$

Eq. (3.16) now follows from (3.20). Gathering (3.15) and (3.16), as $\beta \leq 2-\rho$, we obtain that for large $|x|$ and large $|g(x)|$

$$PV(x) - V(x) = -z\kappa\beta \ |x|^{\beta-\rho} V(x) \ \ (1+o(1)) = -z^{\rho/\beta} \kappa\beta \ [\log V(x)]^{1-\rho/\beta} \, V(x) \ \ (1+o(1)) \,,$$

where

$$\begin{aligned} &\kappa := r && \text{if } \beta < 2-\rho, \ \ i.e. \ \ \gamma_0 < 2-\rho \\ &\kappa := r - 1/2 \ \beta z \ \mathbb{E}\left[\epsilon_0^2 \mathrm{e}^{z|\epsilon_0|^\beta} \right] && \text{if } \beta = 2-\rho \ \ i.e. \ \ \gamma_0 \geq 2-\rho, \end{aligned}$$

and z is chosen small enough such that $\kappa > 0$.
(ii) We now consider the case $\rho = \gamma_0$ (observe that $\beta := \gamma_0 \wedge (2-\rho) = \gamma_0$ and that many results above remain valid). By (3.14), (3.15), (3.17) and (3.20), we have for large $|x|$ and large $|g(x)|$,

$$\frac{PV(x) - V(x)}{V(x)} \leq -zr\gamma_0 + \frac{1}{2}z^2r^2{\gamma_0}^2 + \frac{1}{2}z^2{\gamma_0}^2 \, |x|^{2\gamma_0 - 2}\mathbb{E}\left[\epsilon_0^2 V(\epsilon_0)\right](1 + o(1)).$$

For z small enough, the term on the right hand side is in the interval $(-1, 0)$ and this shows that the Foster Lyapunov drift condition (1.3) holds with C on the form $\{x, g(x) \leq M_1\} \cup \{x, |x| \leq M_2\}$ for large enough M_1, M_2.
(iii) We finally consider the case $\rho < \gamma_0$. Using the inequality $(1-u)^{\gamma_0} \leq 1 - \gamma_0 u$ for all $0 \leq u \leq 1$, we have for $|x| \geq R_0$, $|g(x)|^{\gamma_0} \leq |x|^{\gamma_0} - \gamma_0 r|x|^{\gamma_0 - \rho}$. Hence, since $V(u+w) \leq V(u)V(w)$, this yields, for $|x| \geq R_0$,

$$PV(x) = \mathbb{E}[V(g(x) + \epsilon_0)] \leq V(g(x))\mathbb{E}[V(\epsilon_0)] \leq \mathrm{e}^{-r\gamma_0 z_0 |x|^{\gamma_0 - \rho}}\mathbb{E}\left[\mathrm{e}^{z_0|\epsilon_0|^{\gamma_0}}\right] V(x).$$

Hence $\lim_{|x|\to\infty} PV(x)/V(x) = 0$, which implies that the Foster Lyapunov drift condition (1.3) holds with $C := \{V(x) \leq M\}$ for large enough M.
(iv) In all cases, to conclude the proof, we must bound $PV(x)$ on sets on the form $\{|g(x)| \leq M_1\}$ and $\{|x| \leq M_2\}$. Applying the inequality $V(u+w) \leq V(u)V(w)$ and the fact that g is bounded on compact sets, we obtain, for all such x,

$$PV(x) = \mathbb{E}[V(g(x) + \epsilon_0)] \leq \left\{\sup_{|x| \leq R_0} V(g(x)) + V(M_1)\right\}\mathbb{E}[V(\epsilon_0)] < \infty.$$

□

3.4 Stochastic unit root

We now consider a process which belongs to the wide family of stochastic unit root models. See for example Granger and Sawnson (1997) for many examples. The model we consider is one of the simplest. It has been considered in Gourieroux and Robert (2001) with main focus on its extremal behavior.

$$\Phi_{n+1} = \mathbf{1}_{\{U_{n+1} \leq g(\Phi_n)\}}\Phi_n + \epsilon_{n+1}, \tag{3.21}$$

where $(\epsilon_n, n \in \mathbb{N})$ is a sequence of i.i.d. random variables that satisfies (3.12) and $(U_n, n \geq 1)$ is a sequence of i.i.d random variables, uniformly distributed on $[0, 1]$ and independent from the sequence $(\epsilon_n, n \in \mathbb{N})$. Moreover, we make the following assumption on g.

Assumption 3.5. *g is a continuous function with values in $[0, 1)$ and there exist $\kappa \in (0, 1)$, $c_+(g) > 0$, $c_-(g) < 1$ and $R_0 > 0$ such that*

$$\forall x \geq R_0, \; 1 - g(x) \geq c_+(g)x^{-\kappa}, \tag{3.22}$$

$$\forall x \leq R_0, \; g(x) \leq c_-(g). \tag{3.23}$$

Let P be the transition kernel of the chain. For all $x \in \mathbb{R}$ and all Borel sets A, it can be expressed as:

$$P(x, A) = g(x)\mathbb{P}(x + \epsilon_0 \in A) + (1 - g(x))\mathbb{P}(\epsilon_0 \in A). \tag{3.24}$$

Under Assumption 3.5, for all $M > 0$, there exists a constant $\eta(M)$ such that for all $x \le M$, and all Borel set A,

$$P(x, A) \ge \eta(M)\mathbb{P}(\epsilon_0 \in A). \tag{3.25}$$

This means that every set of the form $(-\infty, M]$ is 1 small, hence petite. Define $x_+ = \max(x, 0)$.

Theorem 3.3. *Under Assumption 3.5 and if ϵ_0 satisfies (3.12), there exist $z \in (0, z_0]$, $\delta > 0$ and $M \ge R_0$ such that the drift condition $\mathbf{D}(\phi, V, C)$ holds with $V(x) = \mathrm{e}^{zx_+^\beta}$, $\phi(v) = \delta z^{\tau/\beta} v\{1 \vee \log(v)\}^{-\tau/\beta}$, $C = (-\infty, M]$ and either*
- *$\beta = \gamma_0 \wedge (1 - \kappa)$ and $\tau = \kappa$ if $\mathbb{E}[\epsilon_0] > 0$;*
- *$\beta = \gamma_0 \wedge (1 - \kappa/2)$ $\tau = \kappa$ if $\mathbb{E}[\epsilon_0] = 0$;*
- *$\beta = \gamma_0$ and $\tau = (1 - \gamma_0) \wedge \kappa$ if $\mathbb{E}[\epsilon_0] < 0$.*

Equations (3.24) and (3.25) prove that the chain is ψ irreducible and (strongly) aperiodic. Since morever V is obviously bounded on intervals $(-\infty, M]$, Theorem 2.8 can be applied.

Proof of Theorem 3.3. Let $z < z_0$ and $x > 0$. Using the definition of the transition kernel P, we have:

$$\begin{aligned} PV(x) - V(x) &= g(x)\mathbb{E}[V(x + \epsilon_0)] + (1 - g(x))\mathbb{E}[V(\epsilon_0)] - V(x) \\ &= g(x)\left(\mathbb{E}[V(x + \epsilon_0)] - V(x)\right) - (1 - g(x))\left(V(x) - \mathbb{E}[V(\epsilon_0)]\right) \\ &\le \mathbb{E}[V(x + \epsilon_0)] - V(x) - (1 - g(x))\left(V(x) - \mathbb{E}[V(\epsilon_0)]\right). \end{aligned}$$

Define $R(x, \epsilon_0) = V(x + \epsilon_0) - V(x) - \epsilon_0 \beta z x^{\beta - 1} V(x)$. For any $\eta \in (0, 1)$, we can write:

$$\mathbb{E}[V(x + \epsilon_0)] - V(x) - \beta z \mathbb{E}[\epsilon_0] x^{\beta-1} V(x) = \mathbb{E}[R(x, \epsilon_0)\mathbf{1}\{|\epsilon_0| \le \eta x\}] + \mathbb{E}[R(x, \epsilon_0)\mathbf{1}\{|\epsilon_0| > \eta x\}].$$

By the same arguments as in the proof of Theorem 3.2, we have:

$$\mathbb{E}[R(x, \epsilon_0)\mathbf{1}\{|\epsilon_0| > \eta x\}] \le \mathbb{E}[V((1 + \eta^{-1})|\epsilon_0|) + V(|\epsilon_0|) + \beta z \eta^{1-\beta}|\epsilon_0|^\beta V(|\epsilon_0|)] \tag{3.26}$$

Thus this term is bounded provided that η and z are chosen such that $(1 + \eta^{-1})^\beta z \le z_0$. To bound the second term, note that for large enough x, the function $x \mapsto x^{2\beta-2}V(x)$ is increasing. Thus, for $x \ge M$, for some M depending on η, and $|\epsilon_0| \le \eta x$, there exists $t \in (0, 1)$ such that

$$\begin{aligned} &V(x + \epsilon_0) - V(x) - \beta z \epsilon_0 x^{\beta-1} V(x) \\ &\quad = \frac{1}{2}\beta(\beta - 1) z (x + t\epsilon_0)^{\beta-2} \epsilon_0^2 V(x + t\epsilon_0) + \frac{1}{2}(\beta z (x + t\epsilon_0)^{\beta-1})^2 \epsilon_0^2 V(x + t\epsilon_0) \\ &\qquad \le \frac{1}{2}\beta^2 z^2 (1 + \eta)^{2\beta-2} x^{2\beta-2} \epsilon_0^2 V(x) V(|\epsilon_0|) \le \frac{1}{2}\beta^2 z^2 x^{2\beta-2} \epsilon_0^2 V(x) V(|\epsilon_0|). \end{aligned}$$

For $c < c_+(g)$ and x large enough, say $x \geq M$ for some $M \geq R_0$, we have

$$(1 - g(x)) (V(x) - \mathbb{E}[V(\epsilon_0)]) \geq c x^{-\kappa} V(x).$$

Hence, taking (3.26) into account, there exists a positive real number M such that if $x \geq M$, then

$$PV(x) - V(x) \leq \left(z\beta x^{\beta-1}\mathbb{E}[\epsilon_0] + \frac{1}{2}\beta^2 z^2 x^{2\beta-2}\mathbb{E}[\epsilon_0^2 V(|\epsilon_0|)] - cx^{-\kappa} \right) V(x).$$

If $\mathbb{E}[\epsilon_0] > 0$, set $\beta = \gamma_0 \wedge (1-\kappa)$. Then, for large enough x, we obtain:

$$PV(x) - V(x) \leq -\delta x^{-\kappa} V(x) = -\delta z^{\kappa/\beta} V(x) \{\log(V(x))\}^{-\kappa/\beta},$$

with $\delta = c < c_+(g)$ if $\gamma_0 < 1-\kappa$ or $\delta = c - \beta z \mathbb{E}[\epsilon_0]$, $c < c_+(g)$ and z such that $\delta > 0$ if $\gamma_0 \geq 1-\kappa$.

If $\mathbb{E}[\epsilon_0] < 0$, set $\beta = \gamma_0$ and $\tau = (1-\gamma_0) \wedge \kappa$. Then, for x large enough,

$$PV(x) - V(x) \leq -\delta x^{-\tau} V(x) = -\delta z^{\tau/\gamma_0} V(x) \{\log(V(x))\}^{-\tau/\gamma_0},$$

with $\delta = c < c_+(g)$ if $\gamma_0 < 1-\kappa$ and $\delta = c - z\beta\mathbb{E}[\epsilon_0]$, $c < c_+(g)$ and z such that $\delta > 0$ if $\gamma_0 \geq 1-\kappa$.

If $\mathbb{E}[\epsilon_0] = 0$, then β must satisfy $2\beta - 2 \leq -\kappa$, thus we set $\beta = (1-\kappa/2) \wedge \gamma_0$, and we obtain

$$PV(x) - V(x) \leq -\delta x^{-\kappa} V(x) = -\delta z^{\kappa/\beta} V(x) \{\log(V(x))\}^{-\kappa/\beta},$$

with $\delta = c < c_+(g)$ if $1-\kappa/2 > \gamma_0$ and $\delta = c - \frac{1}{2}\beta^2 z^2 \mathbb{E}[\epsilon_0^2 V(|\epsilon_0|)]$, with $c < c_+(g)$ and z such that $\delta > 0$ if $1-\kappa/2 \leq \gamma_0$.

$\square$

References

Angonze, P. (1994). *Critères d'ergodicité de Modèles Markoviens. Estimation Non-Paramétrique sous des Hypothèses de Dépendance.* Ph.D. thesis, Université Paris 9, Dauphine.

Angonze, P. (2000). Geometric and subgeometric rates for Markovian processes: A robust approach. Tech. rep., Université de Lille III.

Duflo, M. (1997). *Random Iterative Systems.* Springer Verlag.

Fort, G. and Moulines, E. (2000). V-subgeometric ergodicity for a Hastings-Metropolis algorithm. *Statistics and Probability Letters* **49** 401–410.

Fort, G. and Moulines, E. (2002). Computable bounds for polynomial ergodicity. *To appear in Stochastic Processes ans their Applications.*

Gourieroux, C. and Robert, C. (2001). Tails and extremal behaviour of stochastic unit root models. Tech. rep., Centre de Recherche en Economie et Statistique du Travail.

GRANGER, C. and SAWNSON, N. (1997). An introduction to stochastic unit-root processes. *Journal of Econometrics* **80** 35–62.

GRUNWALD, G., HYNDMAN, R., TEDESCO, L. and TWEEDIE, R. (2000). Non-gaussian conditional linear AR(1) models. *ANZ J Statistics* **42** 479–495.

JARNER, S. and HANSEN, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Processes and Their Applications* **85** 341–361.

JARNER, S. and ROBERTS, G. (2002). Polynomial convergence rates of Markov Chains. *Annals of Applied Probability***12** 224–247.

KLOKOV, S. and VERETENNIKOV, A. (2002). Sub-exponential mixing rate for a class of Markov processes. Tech. Rep. 1, School of Mathematics, University of Leeds.

MALYSHKIN, M. (2001). Subexponential estimates of the rate of convergence to the invariant measure for stochastic differential equations. *Theory of probability and applications* **45** 466–479.

MENGERSEN, K. and TWEEDIE, R. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Annals of Statistics* **24** 101–121.

MEYN, S. and TWEEDIE, R. (1993). *Markov Chains and Stochastic Stability.* Springer-Verlag, London.

NUMMELIN, E. and TUOMINEN, P. (1983). 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.

ROBERTS, G. and TWEEDIE, R. (1996). Geometric convergence and central limit theorem for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83** 95–110.

TANIKAWA, A. (2001). Markov chains satisfying simple drift conditions for subgeometric ergodicity. *Stochastic models* **17** 109–120.

TUOMINEN, P. and TWEEDIE, R. (1994). Subgeometric rates of convergence of f-ergodic Markov Chains. *Advances in Applied Probability* **26** 775–798.

VERETENNIKOV, A. (1997). On polynomial mixing bounds for stochastic differential equations. *Stochastic processes and their Applications* **70** 115–127.

VERETENNIKOV, A. (1999). On polynomial mixing and convergence rate for stochastic difference and differential equations. *Theory of Probability Applications* 361–374.