Available at www.MATHEMATICSWEB.ORG POWERED BY SCIENCE DIRECT

ELSEVIER

Stochastic Processes and their Applications 103 (2003) 57–99

stochastic processes and their applications

www.elsevier.com/locate/spa

Polynomial ergodicity of Markov transition kernels

G. Fort[a,*], E. Moulines[b]

[a] LMC, IMAG, 51 rue des Mathématiques, BP 53, 38041 Grenoble Cedex 09, France

[b] Laboratoire de Traitement et Communication de l'Information, ENST, 46 rue Barrault, 75634 Paris Cedex 13, France

Received 29 May 2000; received in revised form 29 May 2002; accepted 29 May 2002

Abstract

This paper discusses quantitative bounds on the convergence rates of Markov chains, under conditions implying polynomial convergence rates. This paper extends an earlier work by Roberts and Tweedie (Stochastic Process. Appl. 80(2) (1999) 211), which provides quantitative bounds for the total variation norm under conditions implying geometric ergodicity.

Explicit bounds for the total variation norm are obtained by evaluating the moments of an appropriately defined coupling time, using a set of drift conditions, adapted from an earlier work by Tuominen and Tweedie (Adv. Appl. Probab. 26(3) (1994) 775). Applications of this result are then presented to study the convergence of random walk Hastings Metropolis algorithm for super-exponential target functions and of general state-space models. Explicit bounds for f-ergodicity are also given, for an appropriately defined control function f.

MSC: 60J10; 60J22; 65C40

Keywords: Markov chains with discrete parameters; Computational methods in Markov chain; Mixing; Polynomial convergence

1. Introduction

Let P be a transition kernel on a state space $\mathscr{X}$ equipped with a countably generated σ-field $\mathscr{B}(\mathscr{X})$. Let μ be a probability measure on $\mathscr{B}(\mathscr{X})$. Denote by $\Phi := (\mathscr{X}^{\mathbb{Z}_+}, \mathscr{B}(\mathscr{X})^{\otimes \mathbb{Z}_+}, \{\Phi_n\}, P_\mu)$ the time-homogeneous Markov chain with transition

* Corresponding author.

E-mail addresses: gersende.fort@imag.fr (G. Fort), moulines@tsi.enst.fr (E. Moulines).

0304-4149/03/$ - see front matter
PII: S0304-4149(02)00182-5

probability kernel P and initial distribution μ. Assume that the kernel P is ϕ-irreducible, i.e. there exists a measure ϕ on $\mathscr{B}(\mathscr{X})$, such that for all $x \in \mathscr{X}$, $A \in \mathscr{B}(\mathscr{X})$

$$\phi(A) > 0 \Rightarrow P_x(\tau_A < \infty) > 0, \tag{1}$$

where τ_A is the return time to the set A, i.e. $\tau_A := \inf\{n \geqslant 1, \Phi_n \in A\}$ (with the convention $\inf \emptyset = +\infty$). Denote by ψ a maximal irreducibility measure for P (see Meyn and Tweedie, 1993, hereafter MT, Chapter 4, for the definition and the construction of such a measure). We call a set $A \in \mathscr{B}(\mathscr{X})$ *accessible* if $\psi(A) > 0$, *full* if $\psi(A^c) = 0$, and *absorbing* if $P(x, A) = 1$ for all $x \in A$. Recall that when P is ψ-irreducible, then every absorbing set is full. A collection of sets $D_1, D_2, \ldots, D_d$ is called a d-cycle if $P(x, D_k) = 1$ for all $x \in D_{(k-1)[d]}$, $k = 1, \ldots, d$. The kernel is *aperiodic* if there is no 2-cycle.

If P is ψ-irreducible aperiodic and has an invariant probability measure π, it is well known that π is a maximal irreducibility measure and that for all x in a full and absorbing set

$$\lim_n \|P^n(x, .) - \pi(.)\|_{\mathrm{TV}} = 0,$$

where $\|.\|_{\mathrm{TV}}$ is the *total variation distance*, defined for any signed measure μ as $\|\mu\|_{\mathrm{TV}} := \sup_{|g| \leqslant 1} |\mu(g)|$, where $\mu(g) := \int g(x)\mu(\mathrm{d}x)$. In words, for π almost all starting points x the total variation distance of the iterate of the kernel and of the stationary distribution goes to zero as $n \to \infty$. This property is referred to as *ergodicity*.

We will study in this paper a stronger form of ergodicity. Let $r := \{r(n)\}$ be a non-decreasing sequence of positive real numbers and $f \geqslant 1$ be a Borel function. A ψ-irreducible and aperiodic kernel is said (f, r)-*ergodic* if there exists an unique invariant probability measure π such that $\pi(f) < \infty$ and for all x in a full and absorbing set

$$\lim_n r(n) \|P^n(x, .) - \pi(.)\|_f = 0, \tag{2}$$

where $\|.\|_f$ is the f-norm, defined for each signed measure μ as $\|\mu\|_f := \sup_{|g| \leqslant f} |\mu(g)|$. Most of the works in Markov chain theory have been devoted to the case where $r(n) = \beta^n$ for some $\beta > 1$, extending well-known results for irreducible finite state-space chain (a property referred to as f-uniform ergodicity in Meyn and Tweedie, 1993). Tuominen and Tweedie (1994) have developed a set of necessary and sufficient conditions to establish (f, r)-ergodicity for a class of subgeometrical sequences. To state their results, we need some additional definitions. Let Λ_0 be the set of non-decreasing positive sequence $r := \{r(n)\}$ such that $r(0) \geqslant 1$ and $\log\{r(n)\}/n \downarrow 0$. A sequence $r := \{r(n)\}$ is said to be *subgeometrical* if $r(n)$ is strictly positive for all $n \in \mathbb{Z}_+$ and there exists a sequence $r_0 \in \Lambda_0$ such that

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

Denote by Λ the set of subgeometrical sequences. We finally need the notions of sampled chain and petite set. A measurable set C is ν_a-*petite* if there exist a distribution $a := \{a(n)\}$ on $\mathbb{Z}_+$, a constant $\varepsilon > 0$ and a maximal irreducibility measure ν_a on $\mathscr{B}(\mathscr{X})$

such that for all $x \in C$, $B \in \mathscr{B}(\mathscr{X})$,

$$K_a(x,B) := \sum_n a(n)P^n(x,B) \geqslant \varepsilon \nu_a(B). \tag{3}$$

For ψ-irreducible and aperiodic kernels, every petite set is ν_m-small (Meyn and Tweedie, 1993, Theorem 5.5.7); recall that a set $C \in \mathscr{B}(\mathscr{X})$ is ν_m-*small* if there exist an integer $m > 0$, a constant $\varepsilon > 0$ and a probability measure ν_m on $\mathscr{B}(\mathscr{X})$ such that for all $x \in C$, $B \in \mathscr{B}(\mathscr{X})$,

$$P^m(x,B) \geqslant \varepsilon \nu_m(B). \tag{4}$$

For a ψ-irreducible kernel, there is a countable cover of the space with small sets, and every accessible set contains at least one accessible small set (Meyn and Tweedie, 1993, Theorem 5.2.2).

The following proposition is a (partial) statement of Theorem 2.1 in Tuominen and Tweedie (1994) (see also Meyn and Tweedie, 1993, Theorems 10.4.9. and 14.0.1.)

Proposition 1. *Assume that P is ψ-irreducible and aperiodic. Let $f : \mathscr{X} \to [1,\infty)$ be a Borel function, $r := \{r(n)\} \in \Lambda$ be a sub-geometrical sequence, and $C \in \mathscr{B}(\mathscr{X})$ be a petite set such that*

$$\sup_{x \in C} \mathbb{E}_x \left[\sum_{k=0}^{\tau_C - 1} r(k) f(\Phi_k) \right] < \infty. \tag{5}$$

Then,

(i) *there exists an unique invariant probability measure π, which is equivalent to ψ, such that $\pi(f) < \infty$.*
(ii) *for every x in the full and absorbing set $\mathscr{S}(f,r) := \{x,\ \mathbb{E}_x[\sum_{k=0}^{\tau_C-1} r(k)f(\Phi_k)] < \infty\}$,*

$$\lim_n r(n) \|P^n(x,.) - \pi(.)\|_f = 0.$$

(iii) *C is (f,r)-regular, i.e. for any accessible set $B \in \mathscr{B}(\mathscr{X})$,*

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

The (f,r)-ergodicity is thus implied by the control of the so-called (f,r)-modulated moment of the return time to C, $\mathbb{E}_x[\sum_{k=0}^{\tau_C-1} r(k)f(\Phi_k)]$. Tuominen and Tweedie (1994) suggest a set of drift conditions to control the (f,r)-modulated moments. Namely if P is ψ-irreducible and aperiodic, and if there exist a sequence of drift functions $V_n \geqslant 1$, a Borel function $f \geqslant 1$, a petite set C and a finite constant b such that $\{V_1 < \infty\} \subset \{V_0 < \infty\}$, V_0 is bounded on C and

$$PV_{n+1} \leqslant V_n - r(n)f + br(n)\mathbb{1}_C,$$

where $r := \{r(n)\} \in \Lambda$, then Eq. (5) is verified. This condition is unfortunately rather difficult to check in practice necessitating more practical drift conditions.

1.1. Modulated moments and drift conditions

In this contribution, we will focus on the case of *polynomial sequences* $r := \{r(n)\}$, defined as

$$\liminf_n r(n)(n+1)^{-\beta} > 0 \quad \text{and} \quad \limsup_n r(n)(n+1)^{-\beta} < \infty,$$

for some $\beta > 0$. Control of (f,r)-modulated moments for polynomial sequences has been considered (among others) in Tuominen and Tweedie (1994) and Stramer and Tweedie (1999). The key ideas in these contributions consist in using a set of nested drift conditions, an idea formalized in Fort and Moulines (2000), but that can be traced back to an early work by Tweedie (1983) (see also Jarner and Roberts, 2002). We will briefly review the key ideas below. Let $f : \mathscr{X} \to [1,\infty)$ be a Borel function, q be a positive integer and a non-empty set $C \in \mathscr{B}(\mathscr{X})$.

D$[f,q,C]$ There exist some measurable functions on $\mathscr{X}$, $1 \leqslant f =: V_0 \leqslant \cdots \leqslant V_q$, and some finite constants b_k, $k \in \{0,\ldots,(q-1)\}$, such that $\sup_C V_q < \infty$ and for all $k \in \{0,\ldots,(q-1)\}$,

$$PV_{k+1}(x) + V_k(x) \leqslant V_{k+1}(x) + b_k \mathbb{1}_C(x). \tag{6}$$

When $q = 1$, $D[f,1,C]$ is the so-called f-modulated drift towards C (see Meyn and Tweedie, 1993, Chapter 14): under this condition, it can be shown that

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

When P is ψ-irreducible and aperiodic, C petite, this condition implies, by application of Proposition 1, that $\pi(f) < \infty$, the set $\mathscr{S}(f,1) := \{x, \mathbb{E}_x[\sum_{k=0}^{\tau_C-1} f(\Phi_k)] < \infty\}$ is full and absorbing, and $\lim_n \|P^n(x,.) - \pi(.)\|_f = 0$ for all $x \in \mathscr{S}(f,1)$. We will now show how the use of nested drift functions can be used to improve this result. Define the polynomial sequence $\mathbb{1}^{*q} := \{\mathbb{1}^{*q}(n)\}$ as

$$\mathbb{1}^{*0}(n) := 1, \quad n \geqslant 0,$$

$$\mathbb{1}^{*j}(0) := 1, \quad \mathbb{1}^{*j}(n) := \sum_{k=1}^{n} \mathbb{1}^{*(j-1)}(k), \quad j \geqslant 1, n \geqslant 1. \tag{7}$$

Note that $\mathbb{1}^{*1}(n) = n$ for $n \geqslant 1$, and for $j \geqslant 1$, $\mathbb{1}^{*j}(n) = n^j/j! + O(n^{j-1})$ as $n \to \infty$. By convention, we set $\mathbb{1}^{*-1}(n) := 0$, $n \geqslant 0$. Observe that the sequence $\mathbb{1}^{*l} := \{\mathbb{1}^{*l}(n)\}$, $l \geqslant 1$, has the property

$$\forall n,m \geqslant 1, \quad \mathbb{1}^{*l}(n+m) = \sum_{k=0}^{l} \mathbb{1}^{*k}(n)\mathbb{1}^{*(l-k)}(m). \tag{8}$$

Proposition 2. *Let q be a positive integer, $C \in \mathscr{B}(\mathscr{X})$, $C \neq \emptyset$, and $f \geqslant 1$ be a Borel function. Suppose that P is ψ-irreducible and that $D[f, q, C]$ holds. Then C is accessible and*

$$\mathbb{E}_x\left[\sum_{k=0}^{\tau_C - 1} \mathbb{1}^{*(q-1)}(k+1) f(\Phi_k)\right] \leqslant V_q(x) + \left(\sum_{l=0}^{q-1} b_l\right) \mathbb{1}_C(x). \tag{9}$$

If in addition, C is petite, then C is $(f, \mathbb{1}^{*(q-1)})$-regular.

Proof. Eq. (6) implies that $\sup_C \mathbb{E}_x[\tau_C] < \infty$ and thus that the set C is Harris recurrent, which implies that C is accessible by Meyn and Tweedie (1993, Proposition 9.1.1). The key tool to prove Eq. (9) is the so-called Comparison Theorem (Meyn and Tweedie, 1993, Theorem 14.2.2) which shows that, for all $j \in \{0, \ldots, (q-1)\}$ and any stopping time τ, we have

$$\mathbb{E}_x\left[\sum_{k=0}^{\tau-1} V_j(\Phi_k)\right] \leqslant V_{j+1}(x) + b_j \mathbb{E}_x\left[\sum_{k=0}^{\tau-1} \mathbb{1}_C(\Phi_k)\right]. \tag{10}$$

Following an idea initially proposed by Tweedie (1983), we may iterate the Comparison Theorem, which shows that, for $j \in \{1, \ldots, (q-1)\}$,

$$\mathbb{E}_x\left[\sum_{k=0}^{\tau-1} \mathbb{E}_{\Phi_k}\left[\sum_{l=0}^{\tau-1} V_{j-1}(\Phi_l)\right]\right]$$

$$\leqslant \mathbb{E}_x\left[\sum_{k=0}^{\tau-1}\left\{V_j(\Phi_k) + b_{j-1}\mathbb{E}_{\Phi_k}\left[\sum_{l=0}^{\tau-1}\mathbb{1}_C(\Phi_l)\right]\right\}\right] \tag{11}$$

$$\leqslant V_{j+1}(x) + b_j \mathbb{E}_x\left[\sum_{k=0}^{\tau-1}\mathbb{1}_C(\Phi_k)\right] + b_{j-1}\mathbb{E}_x\left[\sum_{k=0}^{\tau-1}\mathbb{E}_{\Phi_k}\left[\sum_{l=0}^{\tau-1}\mathbb{1}_C(\Phi_l)\right]\right], \tag{12}$$

and by the Markov property,

$$\mathbb{E}_x\left[\sum_{k=0}^{\tau-1}\mathbb{1}^{*1}(k+1)V_{j-1}(\Phi_k)\right] = \mathbb{E}_x\left[\sum_{k=0}^{\tau-1}(k+1)V_{j-1}(\Phi_k)\right]$$

$$\leqslant V_{j+1}(x) + \sum_{l=0}^{1} b_{j-l}\mathbb{E}_x\left[\sum_{k=0}^{\tau-1}\mathbb{1}^{*l}(k+1)\mathbb{1}_C(\Phi_k)\right].$$

A straightforward backward recursion finally yields

$$\mathbb{E}_x\left[\sum_{k=0}^{\tau-1}\mathbb{1}^{*j}(k+1)f(\Phi_k)\right] \leqslant V_{j+1}(x) + \sum_{l=0}^{j} b_{j-l}\mathbb{E}_x\left[\sum_{k=0}^{\tau-1}\mathbb{1}^{*l}(k+1)\mathbb{1}_C(\Phi_k)\right]. \tag{13}$$

Applying Eq. (13) with $\tau = \tau_C$ and $j = q-1$ proves that

$$\mathbb{E}_x\left[\sum_{k=0}^{\tau_C-1}\mathbb{1}^{*(q-1)}(k+1)f(\Phi_k)\right] \leqslant V_q(x) + \sum_{l=0}^{q-1} b_l \mathbb{1}_C(x), \tag{14}$$

and Eq. (9) follows. The second assertion is along the same lines as in Meyn and Tweedie (1993, Proposition 14.2.3). Assume that C is v_a-petite (where v_a is a maximal irreducibility measure) and that the sampling distribution (see Eq. (3)) $a := \{a(n)\}$ verifies $\sum_n na(n) < \infty$ (a can always be chosen to verify this latter property). For any accessible set B, we have, $\mathbb{1}_C(x) \leqslant v_a(B)^{-1}K_a(x,B)$. Thus,

$$\mathbb{E}_x\left[\sum_{k=0}^{\tau_B-1} \mathbb{1}^{*l}(k+1)\mathbb{1}_C(\Phi_k)\right] \leqslant v_a(B)^{-1}\sum_{n=0}^{\infty} a(n)\mathbb{E}_x\left[\sum_{k=0}^{\tau_B-1} \mathbb{1}^{*l}(k+1)\mathbb{1}_B(\Phi_{k+n})\right],$$

$$\leqslant v_a(B)^{-1}\sum_{n=0}^{\infty} a(n)\mathbb{E}_x\left[\sum_{k=n\vee\tau_B}^{\tau_B+n-1} \mathbb{1}^{*l}(k+1-n)\mathbb{1}_B(\Phi_k)\right],$$

and as $\mathbb{1}^{*l}$ is an increasing sequence,

$$\mathbb{E}_x\left[\sum_{k=0}^{\tau_B-1} \mathbb{1}^{*l}(k+1)\mathbb{1}_C(\Phi_k)\right] \leqslant v_a(B)^{-1}\left(\sum_{n=0}^{\infty} na(n)\right)\mathbb{E}_x[\mathbb{1}^{*l}(\tau_B)].$$

Apply now Eq. (13) with $\tau = \tau_B$. We have,

$$\mathbb{E}_x[\mathbb{1}^{*(j+1)}(\tau_B)] \leqslant \mathbb{E}_x\left[\sum_{k=0}^{\tau_B-1} \mathbb{1}^{*j}(k+1)f(\Phi_k)\right]$$

$$\leqslant V_{j+1}(x) + v_a(B)^{-1}\left(\sum_n na(n)\right)\sum_{l=0}^{j} b_{j-l}\mathbb{E}_x[\mathbb{1}^{*l}(\tau_B)],$$

and the proof is concluded by an obvious induction. □

Remark 1. Eq. (9) implies that

$$\{V_q < \infty\} \subset \mathscr{S}(f, \mathbb{1}^{*(q-1)}) := \left\{x \in \mathscr{X}, \mathbb{E}_x\left[\sum_{k=0}^{\tau_C-1} \mathbb{1}^{*(q-1)}(k+1)f(\Phi_k)\right] < \infty\right\}.$$

Since, under $D[f,q,C]$, $PV_q \leqslant V_q + b_{q-1}\mathbb{1}_C$, then the set $\{V_q < \infty\}$ is absorbing, and since $C \subset \{V_q < \infty\}$, it is non-empty and $\{V_q < \infty\}$ is full when P is ψ-irreducible. Hence, $\mathscr{S}(f, \mathbb{1}^{*(q-1)})$ is full and absorbing. Note also that under $D[f,q,C]$, $\sup_C V_q < \infty$ and then

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

We will establish the converse assertion in the following proposition.

Similar to what is done in Meyn and Tweedie (1993, Chapter 14) for f-ergodic chain, it is possible to determine the minimal solutions of the nested drift conditions. Recall that, in linking the single f-modulated drift towards C, $PV_1 + f \leqslant V_1 + b_0\mathbb{1}_C$, with f-regularity, MT consider the extended real value function

$$U(x) := \mathbb{E}_x\left[\sum_{k=0}^{\sigma_C} f(\Phi_k)\right],$$

where $\sigma_C := \inf\{n \geqslant 0, \Phi_n \in C\}$ is the hitting-time on C (with the convention $\inf \emptyset = +\infty$). It is easily shown that

$$PU(x) = \mathbb{E}_x\left[\sum_{k=1}^{\tau_C} f(\Phi_k)\right], \tag{15}$$

which implies that, for $x \notin C$, $PU(x) + f(x) = U(x)$. Note that $U(x) = f(x)$ for $x \in C$. Let $C \in \mathscr{B}(\mathscr{X})$ be a $(f,1)$-regular and accessible set. Then, the set C is petite by Meyn and Tweedie (1993, Proposition 14.2.4). In addition, $\sup_{x\in C} \mathbb{E}_x[\sum_{k=0}^{\tau_C-1} f(\Phi_k)] < \infty$ and $\sup_C f < \infty$, so Eq. (15) implies that $\sup_C PU < \infty$. Hence,

$$PU(x) + f(x) \leqslant U(x) + \mathbb{1}_C(x) \sup_{x\in C} \mathbb{E}_x\left[\sum_{k=1}^{\tau_C} f(\Phi_k)\right]$$

and $D[f,1,C]$ is satisfied with $V_1 := U$ (and C is petite).

Conversely, assume that $D[f,1,C]$ is verified for some drift function V. A straightforward adaptation of Meyn and Tweedie (1993, Proposition 11.3.2) (see also Tweedie, 1983) implies that

$$\mathbb{E}_x\left[\sum_{k=0}^{\tau_C} f(\Phi_k)\right] \leqslant V_1(x) + b_0 \mathbb{1}_C(x) \tag{16}$$

which shows that $U(x) \leqslant V_1(x)$ for $x \notin C$. For $x \in C$, $U(x) = f(x) \leqslant V_1(x)$, where the latter inequality holds by construction. Also, since $\sup_C V_1 < \infty$, then $\sup_{x\in C} \mathbb{E}_x[\sum_{k=1}^{\tau_C} f(\Phi_k)] < \infty$ and it may be shown as above that U verifies the drift condition $PU + f \leqslant U + b\mathbb{1}_C$. In conclusion, U is the minimal pointwise solution of $D[f,1,C]$. We now generalize this result to the nested drift conditions. Set $U_0 := f$ and for $j \in \{0,\ldots,(q-1)\}$

$$U_{j+1}(x) := \mathbb{E}_x\left[\sum_{k=0}^{\sigma_C} \mathbb{1}^{*j}(k+1) f(\Phi_k)\right]. \tag{17}$$

Note that, by construction, $1 \leqslant f = U_0 \leqslant \cdots \leqslant U_q$.

Proposition 3. *Let $f : \mathscr{X} \to [1,\infty)$ be a Borel function, q be a positive integer, and $C \in \mathscr{B}(\mathscr{X})$, $C \neq \emptyset$. Assume that*

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

Then, $D[f,q,C]$ is verified with $V_k = U_k$, for $k \in \{0,\ldots,q\}$. In addition, the functions $\{U_k\}$, $k \in \{0,\ldots,q\}$, are the minimal pointwise solutions of Eq. (6), in the sense that, for any other solution $\{V_k\}$, $k \in \{0,\ldots,q\}$ of Eq. (6) verifying $f =: V_0 \leqslant V_1 \leqslant \cdots \leqslant V_q$, we have

$$U_k(x) = \mathbb{E}_x\left[\sum_{j=0}^{\sigma_C} \mathbb{1}^{*(k-1)}(j+1) f(\Phi_j)\right] \leqslant V_k(x), \quad k \in \{0,\ldots,q\}. \tag{18}$$

Proof. For all $j \in \{0, \ldots, (q-1)\}$, $PU_{j+1}(x) = \mathbb{E}_x[\sum_{k=1}^{\tau_C} \mathbb{1}^{*j}(k) f(\Phi_k)]$. Noting that on $\{\Phi_0 \notin C\}$, $\tau_C = \sigma_C$, we have for $x \notin C$,

$$PU_{j+1}(x) + U_j(x) = \mathbb{E}_x\left[\sum_{k=1}^{\tau_C}\{\mathbb{1}^{*j}(k) + \mathbb{1}^{*j-1}(k+1)\} f(\Phi_k)\right] + f(x)$$

$$= \mathbb{E}_x\left[\sum_{k=0}^{\tau_C}\mathbb{1}^{*j}(k+1) f(\Phi_k)\right] = U_{j+1}(x).$$

Hence, for all $j \in \{0, \ldots, (q-1)\}$, we have

$$PU_{j+1}(x) + U_j(x) = U_{j+1}(x) + \mathbb{E}_x\left[\sum_{k=1}^{\tau_C}\mathbb{1}^{*j}(k) f(\Phi_k)\right]\mathbb{1}_C(x)$$

which concludes the proof of the first statement. We now establish that these solutions are pointwise minimal. Let $\{V_k\}$, $k \in \{0, \ldots, q\}$ be a solution of $D[f, q, C]$. On C, $V_j \leqslant V_{j+1}$ whereas on C^c, it may be proved similar to Eq. (16) that $\mathbb{E}_x[\sum_{k=0}^{\tau_C} V_j(\Phi_k)] \leqslant V_{j+1}(x)$. Then

$$\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C} V_j(\Phi_k)\right] \leqslant V_{j+1}(x).$$

By repeated applications of the latter inequality, we obtain similar to Eq. (11), for $j \in \{0, \ldots, (q-1)\}$,

$$\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C}\mathbb{1}^{*j}(k+1) f(\Phi_k)\right] \leqslant V_{j+1}(x),$$

that is $U_{j+1} \leqslant V_{j+1}$, which concludes the proof. □

By combining the two previous propositions and Remark 1, we have

Corollary 1. *The two assertions are equivalent*

(i) $\sup_{x \in C} \mathbb{E}_x[\sum_{k=0}^{\tau_C - 1} \mathbb{1}^{*(q-1)}(k) f(\Phi_k)] < \infty$.
(ii) *There exists a set of functions* $\{V_k\}$, $k \in \{0, \ldots, q\}$, *such that* $D[f, q, C]$ *holds; in addition the subset* $\{V_q < \infty\}$ *is full and absorbing and*

$$\{V_q < \infty\} \subset \left\{x \in \mathscr{X}, \mathbb{E}_x\left[\sum_{k=0}^{\tau_C - 1}\mathbb{1}^{*(q-1)}(k) f(\Phi_k)\right] < \infty\right\}.$$

By combining Corollary 1 and Proposition 1, it is easily seen that

Theorem 1. *Let* q *be a positive integer,* f *be a Borel function and* $C \in \mathscr{B}(\mathscr{X})$ *be a non-empty petite set. Suppose that* P *is* ψ*-irreducible and aperiodic and that* $D[f, q, C]$ *holds. Then* P *possesses an unique invariant probability measure* π *such that* $\pi(f) < \infty$ *and for all* x *in the full and absorbing set* $\{V_q < \infty\}$,

$$\lim_n (n+1)^{(q-1)} \|P^n(x, .) - \pi(.)\|_f = 0. \tag{19}$$

This result extends Theorem 14.0.1 (Meyn and Tweedie, 1993) to the case where q, the number of nested drift conditions is greater than one. Note that if $D[f,q,C]$ is verified for some functions $\{V_j\}$, then, for any $j \in \{0,\dots,(q-1)\}$, assumption $D[V_j, q-j, C]$ also holds which directly implies that $\pi(V_j) < \infty$ and for all x in the full and absorbing set $\{V_q < \infty\}$

$$\lim_n (n+1)^{(q-j-1)} \|P^n(x,.) - \pi(.)\|_{V_j} = 0, \tag{20}$$

showing that there is a trade-off between the rate of convergence and the set of functions that can be controlled.

1.2. Reduction to a single drift condition

To assess the rate of convergence of the Markov chain, it is sometimes convenient to reformulate the set of nested drift conditions into a single drift condition to obtain something similar to the Foster–Lyapunov drift for geometric ergodicity. The possibility of translating the set of nested drift conditions into a single drift condition has been first outlined by Jarner and Roberts (2002)—hereafter JR. In this contribution, a generalization of the JR drift condition is proposed.

Let q be a positive integer, $\phi : [1,\infty) \to [0,\infty)$ be a q-times differentiable function and $C \in \mathscr{B}(\mathscr{X})$, $C \neq \emptyset$. Consider the following assumption

> $\mathbf{S}[\phi,q,C]$ There exist a measurable function $V : \mathscr{X} \to [1,\infty)$ and a constant $b < \infty$ such that $\sup_C V < \infty$ and
>
> $$PV \leqslant V - \phi \circ V + b\mathbb{1}_C. \tag{21}$$
>
> In addition, there exist some finite constants $c_k \geqslant 1$, $k \in \{0,\dots,q\}$, such that for all $k \in \{0,\dots,q\}$ the functions ϕ_k defined recursively as $\phi_q(x) := c_q x$ and $\phi_k := c_k \phi'_{k+1}\phi$ are non-decreasing concave function on $[1,\infty)$ and $1 \leqslant \phi_0 \leqslant \cdots \leqslant \phi_q$.

The condition above is in particular fulfilled by $\phi(x) := cx^{1-\delta}$, for some $0 < \delta < 1$ and $0 < c \leqslant 1$. Then Eq. (21) may be expressed as

$$PV \leqslant V - cV^{1-\delta} + b\mathbb{1}_C, \tag{22}$$

and (22) coincides with the JR drift condition. In such case, $S[\phi,q,C]$ is satisfied by choosing q as any integer such that $q \leqslant 1/\delta$ and for $k \in \{0,\dots,q\}$, $\phi_k(x) = c^{-k} \prod_{l=q-k}^{q-1} (1-l\delta)^{-1}\ x^{1+k\delta-q\delta} \propto x^{1+k\delta-q\delta}$.

Proposition 4. *Assume $S[\phi,q,C]$. Then $D[f,q,C]$ is verified with $f := \phi_0 \circ V$ and functions $V_k \propto \phi_k \circ V$, $k \in \{1,\dots,q\}$.*

Proof. We verify the nested drift conditions with $V_0(x) := \phi_0 \circ V(x) \geqslant 1$ and for $k \in \{1,\dots,q\}$, $V_k(x) := \{\prod_{l=0}^{k-1} c_l\}\phi_k \circ V(x)$. Observe that for a continuously differentiable concave function ψ, ψ' is decreasing and for all $0 \leqslant x \leqslant y$,

$$\psi(y-x) = \psi(y) - \int_{y-x}^{y} \psi'(t)\,\mathrm{d}t \leqslant \psi(y) - x\psi'(y). \tag{23}$$

We have

$$PV_q = \left\{\prod_{l=0}^{q} c_l\right\} PV \leqslant V_q - \left\{\prod_{l=0}^{q-2} c_l\right\} c_{q-1}\ (\phi_q'\phi)\circ V + \left\{\prod_{l=0}^{q} c_l\right\} b\mathbb{1}_C$$

$$\leqslant V_q - \left\{\prod_{l=0}^{q-2} c_l\right\} \phi_{q-1}\circ V + \left\{\prod_{l=0}^{q} c_l\right\} b\mathbb{1}_C$$

$$= V_q - V_{q-1} + \left\{\prod_{l=0}^{q} c_l\right\} b\mathbb{1}_C.$$

In addition, since $c_{q-1}\geqslant 1$, $\{\prod_{l=0}^{q-2} c_l\}\phi_{q-1} \leqslant \{\prod_{l=0}^{q-1} c_l\}\phi_q$ and $V_{q-1}\leqslant V_q$. Let $1\leqslant k\leqslant q-1$ and assume that $V_k\leqslant V_{k+1}$ and there exists $b_k<\infty$ such that $PV_{k+1}\leqslant V_{k+1}-V_k+b_k\mathbb{1}_C$. Since ϕ_k is concave, the Jensen's inequality implies that on C^c, $P(\phi_k\circ V)\leqslant \phi_k(PV)$, and as ϕ_k is non-decreasing, $P(\phi_k\circ V)\leqslant\phi_k(V-\phi\circ V)$. Since $0\leqslant\phi\circ V\leqslant V$, we can apply Eq. (23), and on C^c,

$$P(\phi_k\circ V)\leqslant \phi_k\circ V-(\phi_k'\ \phi)\circ V=\phi_k\circ V-c_{k-1}^{-1}\phi_{k-1}\circ V.$$

Then on C^c, $PV_k\leqslant V_k-V_{k-1}$. The inequalities $0<\phi_{k-1}\circ V\leqslant\phi_k\circ V\leqslant c_qV$ show that on C,

$$P(\phi_k\circ V)-\phi_k\circ V+c_{k-1}^{-1}\phi_{k-1}\circ V\leqslant P(\phi_k\circ V)-\phi_k\circ V+\phi_{k-1}\circ V$$

$$\leqslant P(\phi_k\circ V)\leqslant c_qPV\leqslant c_q(V+b).$$

The assumption $\sup_C V<\infty$ proves that the constant $b_{k-1}:=\sup_C\{PV_k-V_k+V_{k-1}\}$ is finite and $PV_k\leqslant V_k-V_{k-1}+b_{k-1}\mathbb{1}_C$. The induction is concluded by noting that $c_{k-1}\geqslant 1$ implies $V_{k-1}\leqslant V_k$. □

Remark 2. Assume that the drift function V is unbounded off petite sets, i.e. that the level sets $\{V\leqslant M\}$ are petite. This situation is typical in the applications considered herein. The situation of interest is when V and PV are unbounded on the state space. If V is bounded then the state space is petite and in such case, a ψ-irreducible kernel P is uniformly ergodic. If PV is bounded while V is unbounded, then the Foster–Lyapunov condition holds and if P is ψ-irreducible aperiodic then the chain is V-geometrically ergodic. When V and PV are both unbounded, Proposition 4 still holds when $S[\phi,q,C]$ is relaxed as follows: for $k\in\{0,\ldots,(q-1)\}$, the function ϕ_k is equivalent to $\phi_{k+1}'\phi$ as $|x|$ goes to infinity, and is non-decreasing and concave for large $|x|$. See Paragraph 2.1.2 for an example involving the function $\phi(x)\propto x\log^b x$, $b<0$. The proof is a straightforward adaptation of Proposition 4 and is omitted.

Remark 3. In practice, it is not harder to establish the nested drift conditions $D[f,q,C]$ directly rather than deriving $D[f,q,C]$ from the single drift condition $S[\phi,q,C]$. The direct evaluation of the drift functions V_k yields in general better explicit bounds of ergodicity (see Section 3). This is why in the examples we will in general verify directly the nested drift conditions. On the other hand, single drift conditions are appealing because they parallel the well-established theory of geometric ergodicity. In particular, it is worthwhile to note that the strength of the "drift term" $\phi \circ V$ is sub-linear whereas it is linear in the Foster–Lyapunov drift condition where $\phi \circ V = (1-\lambda)V$, for some $0 < \lambda < 1$. Linear drift yields geometrical ergodicity whereas sub-linear drift yields sub-geometrical rate.

Remark 4. Assume that the JR drift condition Eq. (22) is verified and $\sup_C V < \infty$. Set $q := \lfloor 1/\delta \rfloor$. Proposition 4 shows that $D[V^{1-q\delta}, q, C]$ holds. When P is ψ-irreducible aperiodic and C is petite, an application of Proposition 1 shows that for all $k \in \{1, \dots, q\}$

$$\lim_n (n+1)^{k-1} \|P^n(x,.) - \pi(.)\|_{V^{1-k\delta}} = 0.$$

We will see below how the maximal rate can be improved and how the rates can be interpolated (non-integer exponents).

The JR drift condition $PV \leqslant V - cV^{1-\delta} + b1_C$ yields naturally to a *finite* number of nested drift conditions (equal to $\lfloor 1/\delta \rfloor$) and hence to a *finite* polynomial rate of convergence. We conjecture that a single drift condition on the form $PV \leqslant V - cV \log^b(1 + V) + B1_C$ (where $b < 0$) may yield an infinite number of nested drift conditions and presumably to a rate of convergence of the form $r(n) = \exp(\alpha n^{\beta})$ for some $\alpha > 0$ and $\beta < 1$.

1.3. Interpolated rates

The previous set of assumptions $D[f, q, C]$ only allows us to be able to obtain integer polynomial rates. This obviously raises the question of interpolating between these integer rates. Similarly, in the derivations above the maximal attainable rate also is an integer. The next result shows how it is possible, at the price of an additional assumption, to improve this upper bound.

Proposition 5. *Assume that P is ψ-irreducible and aperiodic. Let $f : \mathscr{X} \to [1, \infty)$ be a Borel function, q be a positive integer and $C \in \mathscr{B}(\mathscr{X})$ be a non-empty petite set such that $D[f, q, C]$ holds. Suppose that*

$$\sup_{x \in C} \mathbb{E}_x \left[\sum_{k=0}^{\tau_C - 1} V_q^{\beta}(\Phi_k) \right] < \infty, \tag{24}$$

for some $0 < \beta \leqslant 1$. Then there exists an unique invariant probability measure π and for all x in the full and absorbing set $\{V_q < \infty\}$, for all $0 \leqslant \alpha \leqslant 1$,

$$\lim_n (n+1)^{q\beta\alpha} \, \|P^n(x,.) - \pi(.)\|_{V_q^{\beta(1-\alpha)}} = 0. \tag{25}$$

Proof. Eq. (25) follows from Proposition 1 with $r(n) := (n+1)^{q\beta\alpha}$ and $f := V_q^{\beta(1-\alpha)}$, provided that we can show

$$\sup_{x \in C} \mathbb{E}_x \left[\sum_{k=0}^{\tau_C - 1} (k+1)^{q\beta\alpha} V_q^{\beta(1-\alpha)}(\Phi_k) \right] < \infty. \tag{26}$$

Eq. (26) holds for $\alpha=0$ by assumption. The Jensen's inequality, the sub-additivity of $x \mapsto |x|^\beta$ and Eq. (14) imply that

$$\mathbb{E}_x[\mathbb{1}^{*q}(\tau_C)^\beta] \leqslant (\mathbb{E}_x[\mathbb{1}^{*q}(\tau_C)])^\beta \leqslant V_q^\beta(x) + \mathbb{1}_C(x)\left(\sum_{k=0}^{q-1} b_k\right)^\beta . \tag{27}$$

By the Markov property and using standard manipulations we have

$$\mathbb{E}_x\left[\sum_{k=0}^{\tau_C-1} \mathbb{E}_{\Phi_k}[\mathbb{1}^{*q}(\tau_C)^\beta]\right] = \mathbb{E}_x\left[\sum_{k=0}^{\tau_C-1} \mathbb{1}^{*q}(\tau_C-k)^\beta\right] = \mathbb{E}_x\left[\sum_{k=1}^{\tau_C} \mathbb{1}^{*q}(k)^\beta\right]. \tag{28}$$

Plugging Eq. (27) into Eq. (28) yields

$$\sup_{x\in C} \mathbb{E}_x\left[\sum_{k=1}^{\tau_C} \mathbb{1}^{*q}(k)^\beta\right] \leqslant \sup_{x\in C} \mathbb{E}_x\left[\sum_{k=0}^{\tau_C-1} V_q^\beta(\Phi_k)\right] + \left(\sum_{k=0}^{q-1} b_k\right)^\beta < \infty. \tag{29}$$

As $\mathbb{1}^{*q}(k)^\beta \sim C(k+1)^{\beta q}$, Eq. (26) follows for $\alpha=1$. For $0<\alpha<1$, we have by applying the Hölder's inequality twice

$$\begin{aligned}
&\mathbb{E}_x\left[\sum_{k=0}^{\tau_C-1} (k+1)^{q\beta\alpha} V_q^{\beta(1-\alpha)}(\Phi_k)\right] \\
&\quad \leqslant \mathbb{E}_x\left[\left(\sum_{k=0}^{\tau_C-1} (k+1)^{q\beta}\right)^\alpha \left(\sum_{k=0}^{\tau_C-1} V_q^\beta(\Phi_k)\right)^{1-\alpha}\right] \\
&\quad \leqslant \left(\mathbb{E}_x\left[\sum_{k=0}^{\tau_C-1} (k+1)^{q\beta}\right]\right)^\alpha \left(\mathbb{E}_x\left[\sum_{k=0}^{\tau_C-1} V_q^\beta(\Phi_k)\right]\right)^{1-\alpha},
\end{aligned}$$

showing that Eq. (26) for $0<\alpha<1$ follows from Eq. (24) and Eq. (29). □

Remark 5. As C is accessible (Proposition 2), then the invariant probability verifies

$$\pi(V_q^\beta) = \int_C \pi(\mathrm{d}y)\mathbb{E}_y\left[\sum_{k=0}^{\tau_C-1} V_q^\beta(\Phi_k)\right],$$

(see, Meyn and Tweedie, 1993, Theorem 10.4.9). The expectation is uniformly bounded on C which implies that $\pi(V_q^\beta) < \infty$. In some examples (e.g. when the stationary distribution π is explicitly known), it is easier to establish that P is positive recurrent and that

$$\pi(V_q^\beta) < \infty. \tag{30}$$

As shown in Meyn and Tweedie (1993, Proposition 14.2.11) Eq. (30) implies the existence of a sequence of $(f,1)$-regular sets whose union is full so that Eq. (24) is verified. Hence Eqs. (24) and (30) are actually equivalent.

Remark 6. As shown by Eq. (19), if P is ψ-irreducible aperiodic and if $D[f,q,C]$ holds for a petite set C, the maximal rate of convergence in total variation norm (that can be determined from the drift conditions) is proportional to $(n+1)^{q-1}$. Proposition 5 shows that provided Eq. (24) holds for $(q-1)/q \leqslant \beta \leqslant 1$, the maximal rate of convergence in total variation norm, is proportional to $(n+1)^{q\,\beta}$.

Remark 7. It is known (Meyn and Tweedie, 1993, Theorem 14.0.1) that when P is ψ-irreducible and aperiodic and $D[f,1,C]$ holds for some petite set C and a Borel function $f \geqslant 1$, then $\lim_n \|P^n(x,.)-\pi(.)\|_f = 0$ on the full and absorbing set $\{V_1 < \infty\}$; if in addition $\pi(V_1) < \infty$, then $\sum_{k\geqslant 1} \|P^k(x,.)-\pi(.)\|_f < \infty$ on $\{V_1 < \infty\}$. The result above shows that when $\pi(V_1^\beta) < \infty$ for $0 < \beta \leqslant 1$, then $\lim_n n^\beta \|P^n(x,.) - \pi(.)\|_{\mathrm{TV}} = 0$ for all $x \in \{V_1 < \infty\}$, providing us with a mean to sharpen the conclusions of Meyn and Tweedie (1993, Theorem 14.0.1).

Remark 8. Proposition 5 above allows to retrieve Theorem 3.6 in Jarner and Roberts (2002), when using a single drift condition of the form Eq. (22), i.e. ($PV \leqslant V - cV^{1-\delta} + b\mathbb{1}_C$, $0 < \delta < 1$, and $\sup_C V < \infty$). This condition implies that, for all $0 < \gamma \leqslant 1$,

$$PV^\gamma \leqslant V^\gamma - \gamma c V^{\gamma-\delta} + b_\gamma \mathbb{1}_C, \tag{31}$$

for some finite constant b_γ. Set $q := \lfloor 1/\delta \rfloor > 1$ and $V_k \propto V^{k\delta}$, for $k \in \{0,\dots,q\}$. Eq. (31) shows that the functions $\{V_k\}$ satisfiy the conditions $D[1,q,C]$. Set $\beta := (1-\delta)/(q\delta)$. Then, $V_q^\beta \propto V^{1-\delta}$ and Eq. (22) shows that $\pi(V_q^\beta) < \infty$ (using again e.g. Meyn and Tweedie, 1993, Theorem 14.3.7). By applying Proposition 5, noting that $V < \infty$ on $\mathscr{X}$, we have for all $\alpha \in [0,1]$, $x \in \mathscr{X}$,

$$\lim_n (n+1)^{q\beta\alpha} \|P^n(x,.) - \pi(.)\|_{V^{q\delta\beta(1-\alpha)}} = 0.$$

For all $\kappa \in [1, 1/\delta]$, by setting $\alpha := (\kappa-1)\delta/(1-\delta)$, we finally obtain

$$\lim_n (n+1)^{\kappa-1} \|P^n(x,.) - \pi(.)\|_{V^{1-\kappa\delta}} = 0. \tag{32}$$

2. Examples

In this section $\mathscr{X} = \mathbb{R}^l$, $\mathscr{B}(\mathbb{R}^l)$ is the Borel σ-field and $|.|$ is the Euclidean norm.

2.1. Random Walk Hastings Metropolis algorithm

Let P be the transition kernel of a symmetric Random Walk Hastings Metropolis algorithm (henceforth named the Metropolis algorithm) on $\mathbb{R}$ (see Robert and Casella, 1999). Multidimensional extensions can be obtained using the technique outlined in Jarner and Hansen (2000). Denote by $k(x)$ (resp. $p(x)$) the symmetric proposal density (resp. the target density) w.r.t. the Lebesgue measure. For any Borel bounded function f, Pf is given by

$$Pf(x) - f(x) = \int \{f(x+y) - f(x)\} k(y)\,\mathrm{d}y + \int_{R(x)-x} \{f(x+y) - f(x)\} \left(\frac{p(x+y)}{p(x)} - 1 \right) k(y)\,\mathrm{d}y,$$

where $R(x)$ is the so-called rejection region (where the proposed moves are rejected with a positive probability depending upon the target density)

$$R(x) := x + \{y \in \mathbb{R}, p(x+y) \leqslant p(x)\}. \tag{33}$$

Mengersen and Tweedie (1996) (resp. Roberts and Tweedie, 1996; Jarner and Hansen, 2000) have shown that the Metropolis algorithm on $\mathbb{R}^l$ is geometrically ergodic provided that the target density p is log-concave in the tails (resp. sub-exponential). Fort and Moulines (2000) have shown that the Metropolis algorithm converges at any polynomial rate, when the log density decreases hyperbolically at infinity, $\log p(x) \sim -|x|^s$, $0 < s < 1$, as $|x| \to \infty$ (as for Weibull or Benktander distributions; see Klüppelberg, 1988).

2.1.1. Regular variation in the tails

Assume that

(A1) (i) p is continuous on $\mathbb{R}$ and (ii) there exist some finite constants $s > 1$, $M > 0$, $C > 0$, a function $\rho : \mathbb{R} \to [0, \infty)$ such that for all $|x| \geqslant M$, p is strictly decreasing and for all $y \in R(x) - x$,

$$\left| \frac{p(x+y)}{p(x)} - 1 + syx^{-1} \right| \leqslant C|x|^{-1}\rho(x)y^2$$

and $\lim_{|x|\to\infty} \rho(x) = 0$.

This class contains, among other, the Pareto distributions $p(x) \propto |x|^{-s}$ (in that case, $\rho(x) := 1/|x|$) and many other "heavy tailed" distributions.

(A2) (i) there exist $\varepsilon > 0$ and $\delta < \infty$ such that $|y| \leqslant \delta \Rightarrow k(y) \geqslant \varepsilon$, (ii) the proposal density k is symmetric and there exists $\zeta \geqslant 1$ such that $\int |y|^{\zeta+3} k(y)\,dy < \infty$.

(A1i) and (A2i) ensure that P is ψ-irreducible and aperiodic and that every compact set is petite (Roberts and Tweedie, 1996, Theorem 2.2.). The key result to apply the derivations above is

Lemma 1. *Assume* (A1–2). *Set* $s_* := \zeta \wedge s$. *For all* $2 \leqslant \beta < s_* + 1$, $x \mapsto \int P(x, dy)|y|^\beta$ *is bounded on compact set and*

$$\int P(x, dy)|y|^\beta \leqslant |x|^\beta - \frac{1}{2}\sigma_k^2 \beta(s+1-\beta)|x|^{\beta-2}(1+\varepsilon(x)),$$

where $\sigma_k^2 := \int y^2 k(y)\,dy$ and $\lim_{|x|\to\infty} \varepsilon(x) = 0$.

The proof is in Appendix A.1. Set $V(x) := 1 + |x|^\beta$ for some $2 < \beta < s_* + 1$. It is easily seen that V is a solution to the JR drift condition Eq. (22) with $\delta := 2/\beta$. Since the compact sets are petite, we deduce from the discussions of Paragraph 1.3 that for any $\gamma \in [0, (s_* - 1)/2)$,

$$\lim_n (n+1)^\gamma \|P^n(x, .) - \pi(.)\|_{TV} = 0, \qquad x \in \mathbb{R},$$

and for all $r \in [0, s_* - 1)$ and $\gamma \in [0, (s_* - 1 - r)/2)$,

$$\lim_n (n+1)^\gamma \|P^n(x,.) - \pi(.)\|_{1+|x|^r} = 0, \quad x \in \mathbb{R}.$$

2.1.2. Weibull distribution on $\mathbb{R}_+$

Assume that

(A3) the target density p is a standard Weibull density on $[0, \infty)$ with shape parameter $0 < \eta < 1$, that is $p(x) := \eta x^{\eta-1} \exp(-x^\eta)$, $x > 0$,
(A4) (i) there exist $\varepsilon > 0$ and $\delta < \infty$ such that $|y| \leqslant \delta \Rightarrow k(y) \geqslant \varepsilon$, (ii) the proposal density k is symmetric and $\int_0^\infty y^{\beta_* + 2\eta + 2} \exp(\alpha y^\eta) k(y)\, \mathrm{d}y < \infty$ for some $0 < \alpha < 1$, $\beta_* \in \mathbb{R}$.

As in the preceding example, (A4i) and the continuity of the target density p ensure that the kernel is ψ-irreducible and aperiodic and that every compact set is petite.

Lemma 2. *Assume* (A3–4). *For all $\beta \leqslant \beta_*$, $x \mapsto \int P(x, \mathrm{d}y) y^\beta \exp(\alpha y^\eta)$ is bounded on compact set and*

$$\int P(x, \mathrm{d}y) y^\beta \exp(\alpha y^\eta)$$
$$= x^\beta \exp(\alpha x^\eta) - \frac{1}{2} \sigma_k^2 \alpha (1-\alpha) \eta^2 x^{2(\eta-1)}\, x^\beta \exp(\alpha x^\eta)(1 + \varepsilon(x)), \tag{34}$$

where $\sigma_k^2 := \int y^2 k(y)\, \mathrm{d}y$ and $\lim_{|x| \to \infty} \varepsilon(x) = 0$.

The proof is in Appendix A.1. Let $f(x) := (1 \vee x)^\beta \exp(\alpha x^\eta)$ for some $\beta \leqslant \beta_*$, and let q be a positive integer. Eq. (34) entails that there exists a compact set $C = C(q, \alpha, \beta)$ and some measurable functions $V_k \propto (1 \vee x)^{\beta + 2k(1-\eta)} \exp(\alpha x^\eta)$, solving $D[f, q, C]$. Thus, by Theorem 1, for all $q \geqslant 0$, $\beta \leqslant \beta_*$

$$\lim_n (n+1)^q \|P^n(x,.) - \pi(.)\|_{(1 \vee x)^\beta \exp(\alpha x^\eta)} = 0, \quad x > 0.$$

Eq. (34) can also be cast into the framework detailed in Paragraph 1.2. For example, whenever $\beta_* = 0$, Eq. (34) can be translated into a single drift condition $PV \leqslant V - \phi \circ V + b 1_C$, with $V(x) := \exp(\alpha x^\eta)$ and $\phi(x) \propto x \log^{2(\eta-1)/\eta} x$.

2.2. A non-linear state-space model

Let $F : \mathbb{R}^l \to \mathbb{R}^l$ be a measurable function. Let us consider the non-linear state-space model $\{\Phi_n\}$, defined for $n \geqslant 0$

$$\Phi_{n+1} = F(\Phi_n) + W_{n+1},$$

where

(NSS 1) $\{W_n\}$ is a sequence of i.i.d random variables with distribution $\Gamma(\mathrm{d}x) := \gamma(x)\mathrm{d}x$, and Φ_0 is independent from $\{W_n\}$.

Non-linear state-space models have received a large attention in the literature. Most of the contributions focus on conditions implying V-uniform geometric ergodicity (see Doukhan and Ghindès, 1980; Mokkadem, 1987; Tanikawa, 1996, 1999; Diaconis and Freedman, 1999). We focus here on conditions that rather imply polynomial convergence rates. The transition kernel P is given by $P(x, \mathrm{d}y) = \gamma(y - F(x))\,\mathrm{d}y$. It is assumed that

(NSS 2) P is Lebesgue-irreducible, aperiodic and every non-empty compact set is petite.

Conditions upon which this assumption holds can be found, e.g. see Meyn and Tweedie (1993, Chapter 6 and the references therein). In particular, (NSS 2) is fulfilled whenever (i) $F : \mathbb{R}^l \to \mathbb{R}^l$ is continuous, (ii) $|F(x)| \leqslant |x|$ for all $x \in \mathbb{R}^l$ and (iii) there exist $\varepsilon_\gamma > 0$ and $\delta_\gamma < \infty$ such that $|y| \leqslant \delta_\gamma \Rightarrow \gamma(y) \geqslant \varepsilon_\gamma$.

(NSS 3) There exist $0 < d \leqslant 2$, $r > 0$ and $M < \infty$ such that

$$|F(x)| \leqslant |x|(1 - r|x|^{-d}) \text{ on } |x| \geqslant M, \quad \text{and} \quad \sup_{|x| \leqslant M} |F(x)| < \infty.$$

(NSS 4) There exists $s_* \geqslant d$ such that $\Gamma(s_*) := \int |y|^{s_*} \gamma(y)\,\mathrm{d}y < \infty$.

The study of non-linear state-space models under (NSS 3–4) has been initiated among others by Tuominen and Tweedie (1994) and AngoNze (1994) with $0 < d < 1$. This model has later been worked out by Veretennikov (2000) who proved ergodicity at sub-geometrical rate (see below) for $d = 2$.

Lemma 3. *Assume* (NSS 1–4). *Then, for all* $d \leqslant s \leqslant s_*$, $x \mapsto \int P(x, \mathrm{d}y)|y|^s$ *is bounded on compact sets.*

(i) *Assume that* $0 < d < 1$. *Then,*

$$\int P(x, \mathrm{d}y)|y|^d \leqslant |x|^d - \lambda_1(1 + \varepsilon(x)), \tag{35}$$

with $\lambda_1 := dr - \Gamma(d)$ *and* $\lim_{|x| \to \infty} \varepsilon(x) = 0$. *In addition, for all* $d < s \leqslant s_*$,

$$\int P(x, \mathrm{d}y)|y|^s \leqslant |x|^s - \lambda_2 |x|^{s-d}(1 + \varepsilon(x)), \tag{36}$$

with $\lambda_2 := sr$.

(ii) *Assume that* $1 \leqslant d < 2$, $\int x\Gamma(\mathrm{d}x) = 0$ *and* $s_* \geqslant 2$. *Then,*
- *Eq.* (35) *holds with* $\lambda_1 := dr$.
- *If* $s_* < 4$ (*resp.* $s_* \geqslant 4$) *Eq.* (36) *holds for all* $d < s \leqslant 2$ (*resp.* $d < s \leqslant s_*$), *with* $\lambda_2 := sr$.

(iii) *Assume that* $d = 2 \leqslant s_*$ *and* $\int x\Gamma(\mathrm{d}x) = 0$. *Then,*
- *Eq.* (35) *holds with* $\lambda_1 := 2r - \Gamma(2)$.
- *If* $s_* \geqslant 4$, *then Eq.* (36) *holds for all* $2 < s \leqslant s_*$ *with* $\lambda_2 := sr - (5s - 8)\Gamma(2)/2$ *if* $s \leqslant 4$ *and* $\lambda_2 := sr - s(s-1)\Gamma(2)/2$ *if* $s \geqslant 4$.

The proof is in Appendix A.2. The case $0 < d < 1$ has been addressed by AngoNze, Theorem 2(2): the proof closely parallels the arguments used by this author. To deal with the cases (ii) and (iii), we write for $|x| \geqslant M$,

$$\int P(x, \mathrm{d}y)|y|^s = \int |F(x) + y|^s \Gamma(\mathrm{d}y)$$

$$\leqslant \int (|x|^2(1 - r|x|^{-d})^2 + |y|^2 + 2\langle F(x), y\rangle)^{s/2} \Gamma(\mathrm{d}y),$$

where $\langle .,. \rangle$ is the scalar product. An upper bound of the integrand is then obtained by Taylor expansion of the function $t \mapsto (t + a)^{s/2}$, with $a := |x|^2(1 - r|x|^{-d})^2$ and $t := |y|^2 + 2\langle F(x), y\rangle$.

Remark 9. In cases (ii) and (iii), the moment assumptions can be weakened if one assumes that the distribution Γ is symmetric. More precisely, assume that (NSS1–4) hold for some $1 \leqslant d \leqslant 2$ and that Γ is symmetric. Then, Eq. (36) holds for all $1 \leqslant d \leqslant 2$ and all $d < s \leqslant s_*$. This assertion can be proved along the same lines as the proof of Lemma 3, by setting $a := |x|^2(1 - r|x|^{-d})^2 + |y|^2$ and $t := 2\langle F(x), y\rangle$.

If $\lambda_1 > 0$, Eq. (35) shows that P possesses an unique invariant probability measure π, and, for all $x \in \mathbb{R}^l$, $\lim_n \|P^n(x,.) - \pi(.)\|_{\mathrm{TV}} = 0$. Assume that $s_* > d$ and Eq. (36) holds for all $d \leqslant s \leqslant s_*$ (that is for example either $d < 1$ or $1 \leqslant d \leqslant 2$ and $s_* \geqslant 4$) and $\lambda_2 > 0$; set $\delta := d/s_*$ and $q := \lfloor 1/\delta \rfloor > 1$. Lemma 3 shows that ($a$) $\pi(|x|^{s_*-d}) < \infty$ (by Meyn and Tweedie, 1993, Theorem 14.3.7); (b) the functions $V_k(x) \propto |x|^{kd}$, $k \in \{0, \ldots, q\}$, are solutions of $D[1, q, C]$ for some compact (and hence petite) set C. Set now $\beta := (1 - \delta)/(q\delta)$. Hence, $V_q^\beta(x) \propto |x|^{s_*-d}$ and $\pi(V_q^\beta) < \infty$. Proposition 5 establishes that for all $x \in \mathbb{R}^l$ and all $\gamma \in [1, s_*/d]$,

$$\lim_n (n + 1)^{\gamma - 1} \|P^n(x,.) - \pi(.)\|_{1+|x|^{s_* - \gamma d}} = 0. \tag{37}$$

Tuominen and Tweedie (1994, Proposition 5.2.) considered the case $0 < d < 1$. They assume that $s_* \in \mathbb{Z}_+$, $s_* \geqslant 2$ and prove that for all $x \in \mathbb{R}^l$ and all $\gamma \in \{1, \ldots, s_*\}$,

$$\lim_n (n + 1)^{\gamma - 1} \|P^n(x,.) - \pi(.)\|_{1+|x|^{s_* - \gamma}} = 0,$$

which is a weaker result than Eq. (37). AngoNze (1994, Theorem 2) establishes Eq. (37) for $0 < d < 1$ and $s_* \geqslant 1$. Veretennikov (2000, Theorem 1) studied the case $d = 2$ and showed that when $s_* > 4$,

$$\lim_n (n + 1)^{\gamma - 1} \|P^n(x,.) - \pi(.)\|_{\mathrm{TV}} = 0$$

for $\gamma \in [1, s_*/2)$ (the upper bound being not exactly obtained).

3. Computable bounds for polynomial ergodicity

The main objective of this section is to determine computable bounds, i.e. a function $B : \mathscr{X} \times \mathbb{Z}_+ \to [0, 2]$ such that, for all $x \in \mathscr{X}$ and all $n \in \mathbb{Z}_+$,

$$\|P^n(x,.) - \pi(.)\|_{TV} \leqslant B(x, n). \tag{38}$$

Roughly speaking, a bound is a computable bound, when it can be expressed as a function of quantities that can be explicitly determined from the transition kernel, such as constants appearing in drift conditions and minorizing constants on small sets. Computable bounds have been derived under conditions implying geometric ergodicity (see, e.g. Rosenthal, 1995a, b; Roberts and Tweedie, 1999, 2001, and the references therein).

As illustrated by Rosenthal (1995a) and Roberts and Tweedie (1999), computational bounds for the total-variation distance can be obtained by using the so-called Lindvall's inequality, which relates $\|P^n(x,.)-\pi(.)\|_{\mathrm{TV}}$ to the tail probability of a coupling time T. The construction of this coupling time involves defining a probability space $(\Omega,\mathscr{F},\mathbb{P})$ and a process $Z=\{(X_n,X_n',d_n)\}$ such that (a) for all positive measurable function f on $(\mathscr{X},\mathscr{B}(\mathscr{X}))$, and all $n \geqslant 0$,

$$\int f(X_n)\,\mathrm{d}\mathbb{P} = P^n f(x), \qquad \int f(X_n')\,\mathrm{d}\mathbb{P} = \pi(f),$$

(b) $X_n = X_n'$ for all n greater than the coupling time T, and $T<\infty$ $\mathbb{P}$-a.s. The auxiliary variable d_n, referred to as the bell variable, is set to 0 for $n \leqslant T$ and 1 otherwise. This construction is a standard tool for chain on countable space and consequently, for atomic general state-space chain. For non-atomic general state-space chain, such construction is possible, e.g. if one can find a set $\Delta \subset \mathscr{X}\times\mathscr{X}$, and an integer $m \geqslant 1$, such that for all $(x,x')\in\Delta$,

$$P^m(x,A)\wedge P^m(x',A) \geqslant \varepsilon \nu_m(A) \tag{39}$$

for some constant $\varepsilon>0$ and a probability measure ν_m on $\mathscr{B}(\mathscr{X})$. This is not a minimal assumption for defining a coupling procedure but it suffices for our purpose. Lindvall's inequality shows that $\|P^n(x,.)-\pi(.)\|_{\mathrm{TV}} \leqslant 2\mathbb{P}(T>n)$, and thus determining $B(x,n)$ amounts to compute an upper bound for the tail probability of T. A convenient way to determine such bound is to use a (refinement of the) Markov's inequality, which implies to compute appropriately defined moments of T. This approach, first investigated by Rosenthal (1995a), has later been improved by Roberts and Tweedie (1999), who relate moments of T to moments of the hitting-time on Δ, $\sigma := \inf\{n \geqslant 0, (X_n,X_n')\in\Delta\}$. These contributions deal with the geometrical case and compute the generating function of the coupling time $\mathbb{E}[\beta^T]$, $\beta>1$, as a function of the minorization constant ε (Eq. (39)) and of the generating function of the hitting-time σ, $\mathbb{E}[\beta^\sigma]$. The latter quantity is then classically bounded using the Foster–Lyapunov drift criterion. We extend the results of Roberts and Tweedie to polynomial rate functions $r(n)\propto n^q$, for a real $q\geqslant 0$, and substitute the Foster–Lyapunov drift criterion for the nested drift conditions $D[f,q,C]$.

In order to present the main ideas, we first complete the above program under simple conditions (Theorems 2 and 3); we will show afterwards how to improve such results (Theorems 4 and 5). Let $C,D\in\mathscr{B}(\mathscr{X})$, $C\subseteq D$. Assume that

H1a[C,D] C is accessible and D is ν_1-small: there exist $\varepsilon>0$ and a probability measure ν_1 on $\mathscr{B}(\mathscr{X})$ such that

$$\forall x\in D \quad \forall A\in\mathscr{B}(\mathscr{X}), \qquad P(x,A)\geqslant \varepsilon\nu_1(A).$$

Set $\Delta := C \times D \cup D \times C$ and define the residual kernel

$$R(x, \mathrm{d}y) := (1 - \varepsilon \mathbb{1}_D(x))^{-1}(P(x, \mathrm{d}y) - \varepsilon \mathbb{1}_D(x)\nu_1(\mathrm{d}y)). \tag{40}$$

The classical coupling construction (see, e.g. Rosenthal, 1995a) can be summarized as follows. Let λ, λ' be two probability measures on $\mathscr{B}(\mathscr{X})$. Sample X_0 (resp. X_0') from λ (resp. λ') and set $d_0 := 0$.

(1) if $d_n = 0$ and
  (i) $(X_n, X_n') \notin \Delta$, then set $d_{n+1} := 0$ and sample independently X_{n+1} and X_{n+1}' from the distribution $P(X_n, .)$ and $P(X_n', .)$.
  (ii) $(X_n, X_n') \in \Delta$, then
  - with probability ε, set $d_{n+1} := 1$, and draw $X_{n+1} = X_{n+1}' \sim \nu_1$.
  - with probability $1-\varepsilon$, set $d_{n+1} := 0$ and sample *independently* $X_{n+1} \sim R(X_n, .)$ and $X_{n+1}' \sim R(X_n', .)$.

(2) if $d_n = 1$, set $d_{n+1} := 1$, draw $X_{n+1} = X_{n+1}' \sim P(X_n, .)$.

It is easily seen that $Z = \{Z_n := (X_n, X_n', d_n)\}$ is an homogeneous Markov chain on $(\Omega, \mathscr{F}) := (\mathscr{X}^{\mathbb{Z}_+} \times \mathscr{X}^{\mathbb{Z}_+} \times \{0,1\}^{\mathbb{Z}_+}, \mathscr{B}(\mathscr{X} \times \mathscr{X} \times \{0,1\})^{\otimes \mathbb{Z}_+})$. Let P^* be its transition kernel. Denote by $P_{\lambda,\lambda',i}$ (resp. $\mathbb{E}_{\lambda,\lambda',i}$) the probability (resp. the expectation) on $(\Omega, \mathscr{F})$ for the initial distribution $\lambda \otimes \lambda' \otimes \delta_i$ $(i \in \{0,1\})$. Endow the probability space with the natural filtration $\mathscr{F}_n :=\sigma(Z_k, k \leqslant n)$. It is easily seen that for any $n \geqslant 0$, and any positive Borel function f

$$\mathbb{E}_{\lambda,\lambda',0}[f(X_n)] = \lambda P^n f \quad \text{and} \quad \mathbb{E}_{\lambda,\lambda',0}[f(X_n')] = \lambda' P^n f. \tag{41}$$

Define the $\mathscr{F}_n$-adapted coupling-time T as $T :=\inf\{n \geqslant 1, d_n = 1\}$. Setting $\lambda = \delta_x$ and $\lambda' = \pi$ the stationary distribution for P, the coupling inequality then reads

$$\|P^n(x, .) - \pi(.)\|_{\mathrm{TV}} \leqslant 2P_{x,\pi,0}(T > n). \tag{42}$$

The tails of the coupling-time are controlled by the Markov inequality, $P_{x,\pi,0}(T > n) \leqslant r(n)^{-1}\mathbb{E}_{x,\pi,0}[r(T)]$, for any positive non-decreasing sequence $\{r(n)\}$. Since by construction $T \geqslant 1$ and, for any initial probabilities λ, λ', $P_{\lambda,\lambda',0}(T > n + k | T > n) \geqslant (1 - \varepsilon)^k$ for all $k \geqslant 0$ and $n \geqslant 0$, this inequality can be refined as follows (see Theorem 4.1 in Roberts and Tweedie, 1999).

Lemma 4. *Let $T \geqslant m$ be a random-time on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ such that, for some $0 < \varepsilon < 1$, $\mathbb{P}(T > n + k | T > n) \geqslant (1 - \varepsilon)^k$ for all $k \geqslant 0$, $n \geqslant m - 1$. For a non-decreasing and positive sequence $r := \{r(n)\}$, $n \geqslant m$,*

$$\mathbb{P}(T > n) \leqslant U_{\varepsilon,m}(r; n) \quad \mathbb{E}[r(T - m)] \tag{43}$$

$$U_{\varepsilon,m}(r; n) := (r(n + 1 - m) + (1 - \varepsilon)^{-(n+1-m)} \sum_{j > n+1-m} (1 - \varepsilon)^j (r(j + 1) - r(j)))^{-1}. \tag{44}$$

The construction of an implicit bound then amounts to determine the moments $\mathbb{E}_{x,\pi,0}[r(T)]$, for the polynomial sequence $r(n) := \mathbb{1}^{*q}(n)$ (see Eq. (7)), where q is some positive integer. Define the hitting-time on Δ and the successive $\mathscr{F}_n$-adapted return-time to Δ of the process $\{(X_n, X'_n)\}$

$$T_0 := \inf\{k \geqslant 0, (X_k, X'_k) \in \Delta\},$$
$$T_n := \inf\{k \geqslant T_{n-1} + 1, (X_k, X'_k) \in \Delta\}, \quad n \geqslant 1 \tag{45}$$

with the convention that $\inf \emptyset := +\infty$ and $T_{-1} := -1$. For $l \in \{0, \ldots, q\}$, denote by

$$A(l) := (1-\varepsilon) \sup_{(x,x') \in \Delta} \int R(x, \mathrm{d}y) R(x', \mathrm{d}y') \mathbb{E}_{y,y',0}[\mathbb{1}^{*l}(1 + T_0)], \tag{46}$$

and $A(l) := 0$ otherwise. Define the (Toeplitz) matrix $A := [A_{i,j}]_{0 \leqslant i,j \leqslant q}$, where $A_{i,j} := A(i-j)$. Since A is lower triangular and that the diagonal elements are strictly less than 1 (by construction, $A(0) = 1 - \varepsilon$), $I - A$ is invertible. Define $B := (I - A)^{-1}$.

Remark 10. Observe that for all $l \in \{0, \ldots, q\}$, $A(l) \leqslant (1-\varepsilon)\alpha^{l/q}$ where

$$\alpha := \sup_{n \geqslant 1, l \in \{1,\ldots,q\}} \frac{[\mathbb{1}^{*l}(n)]^{q/l}}{\mathbb{1}^{*q}(n)} \sup_{(x,x') \in \Delta} \int R(x, \mathrm{d}y) R(x', \mathrm{d}y') \mathbb{E}_{y,y',0}[\mathbb{1}^{*q}(1 + T_0)].$$

This implies that for $0 \leqslant l \leqslant k \leqslant q$, $j \geqslant 0$, $[A^{j+1}]_{k,l} \leqslant (1-\varepsilon)^{j+1} \alpha^{(k-l)/q} \mathbb{1}^{*j}(k - l + 1)$. For $k = l$, $\varepsilon B_{k,k} = 1$. For $0 \leqslant l < k \leqslant q$, $B_{k,l} \leqslant (1-\varepsilon)\ \alpha^{(k-l)/q} \sum_{j \geqslant 0} (1-\varepsilon)^j \mathbb{1}^{*j}(k-l+1)$. By definition of the sequence $\mathbb{1}^{*j}$, for $n \geqslant 1$,

$$\sum_{j \geqslant 0} (1-\varepsilon)^j \mathbb{1}^{*j}(n) = 1 + (1-\varepsilon) \sum_{r=1}^{n} \sum_{j \geqslant 0} (1-\varepsilon)^j \mathbb{1}^{*j}(r),$$

and by a trivial induction (on n), it is established that $\sum_{j \geqslant 0} (1-\varepsilon)^j \mathbb{1}^{*j}(n) = \varepsilon^{-n}$. Hence, for $l < k$, $\varepsilon B_{k,l} \leqslant (1-\varepsilon)\ \alpha^{(k-l)/q} \varepsilon^{-(k-l)}$. Finally, $B_{k,l} = 0$ for all $l > k$.

Proposition 6. *Let $C, D \in \mathscr{B}(\mathscr{X})$, $C \subseteq D$, such that* H1a$[C, D]$ *holds. Let q be a positive integer. For any integer $k \in \{0, \ldots, q\}$, and any initial probabilities λ, λ' on $\mathscr{B}(\mathscr{X})$ such that $\mathbb{E}_{\lambda,\lambda',0}[\mathbb{1}^{*q}(T_0)] < \infty$,*

$$\mathbb{E}_{\lambda,\lambda',0}[\mathbb{1}^{*k}(T - 1)] \leqslant \varepsilon \sum_{l=0}^{q} B_{k,l} \mathbb{E}_{\lambda,\lambda',0}[\mathbb{1}^{*l}(T_0)]. \tag{47}$$

Remark 11. From Remark 10 and (47), it is easily seen that

$$\mathbb{E}_{\lambda,\lambda',0}[\mathbb{1}^{*k}(T - 1)] \leqslant \mathbb{E}_{\lambda,\lambda',0}[\mathbb{1}^{*k}(T_0)] + (1-\varepsilon) \sum_{l=0}^{k-1} \alpha^{(k-l)/q} \varepsilon^{-(k-l)} \mathbb{E}_{\lambda,\lambda',0}[\mathbb{1}^{*l}(T_0)].$$

Proof. Eq. (47) is trivial for $k = 0$. We show, by induction on j, that for all $k \in \{1, \ldots, q\}$,

$$\mathbb{E}_{\lambda,\lambda',0}[\mathbb{1}^{*k}(T_j) \mathbb{1}_{\{0\}}(d_{T_j})] \leqslant \sum_{l=0}^{q} [A^j]_{k,l} \quad \mathbb{E}_{\lambda,\lambda',0}[\mathbb{1}^{*l}(T_0)]. \tag{48}$$

For $j=0$, Eq. (48) is obvious. Assume now that Eq. (48) holds for $j \geqslant 0$. The induction assumption implies that $T_j \mathbb{1}_{\{0\}}(d_{T_j}) < \infty$ $P_{\lambda,\lambda',0}$-a.e. Note that $\{d_{T_{j+1}}=0\} = \{d_{T_j+1}=0\} \subset \{d_{T_j}=0\}$. Eq. (8) implies that

$$\begin{aligned}
&\mathbb{E}_{\lambda,\lambda',0}[\mathbb{1}^{*k}(T_{j+1})\mathbb{1}_{\{0\}}(d_{T_{j+1}})] \\
&\quad = \sum_{l=0}^{k} \mathbb{E}_{\lambda,\lambda',0}[\mathbb{1}^{*l}(T_j)\ \mathbb{1}^{*(k-l)}(T_{j+1}-T_j)\ \mathbb{1}_{\{0\}}(d_{T_{j+1}})] \\
&\quad = \sum_{l=0}^{k} \mathbb{E}_{\lambda,\lambda',0}[\mathbb{E}_{\lambda,\lambda',0}\left[\mathbb{1}^{*(k-l)}(T_{j+1}-T_j)\mathbb{1}_{\{0\}}(d_{T_{j+1}})|\mathscr{F}_{T_j}\right]\mathbb{1}^{*l}(T_j)\ \mathbb{1}_{\{0\}}(d_{T_j})].
\end{aligned}$$

By construction, $T_{j+1} = 1 + T_j + T_0 \circ \theta^{1+T_j}$ where θ is the usual shift operator, and, for $(x,x') \in \Delta$, and any positive Borel function $\phi : \mathscr{X} \times \mathscr{X} \mapsto \mathbb{R}^+$,

$$\mathbb{E}_{x,x',0}[\phi(X_1,X_1')\mathbb{1}_{\{0\}}(d_1)] = (1-\varepsilon)\int R(x,\mathrm{d}z)R(x,\mathrm{d}z')\phi(z,z').$$

The strong Markov property shows that, for any $l \in \{0,\dots,q\}$,

$$\begin{aligned}
&\mathbb{E}_{\lambda,\lambda',0}[\mathbb{1}^{*l}(T_{j+1}-T_j)\mathbb{1}_{\{0\}}(d_{T_{j+1}})|\mathscr{F}_{T_j}] \\
&\quad = \mathbb{E}_{\lambda,\lambda',0}[\mathbb{E}_{\lambda,\lambda',0}[\mathbb{1}^{*l}(1+T_0)|\mathscr{F}_{T_j+1}]\mathbb{1}_{\{0\}}(d_{T_j+1})|\mathscr{F}_{T_j}] \\
&\quad = (1-\varepsilon)\int R(X_{T_j},\mathrm{d}z)R(X'_{T_j},\mathrm{d}z')\mathbb{E}_{z,z',0}[\mathbb{1}^{*l}(1+T_0)] \leqslant A(l),
\end{aligned}$$

which implies, by using the induction assumption

$$\begin{aligned}
\mathbb{E}_{\lambda,\lambda',0}[\mathbb{1}^{*k}(T_{j+1})\ \mathbb{1}_{\{0\}}(d_{T_{j+1}})] &\leqslant \sum_{l=0}^{k} A(k-l)\mathbb{E}_{\lambda,\lambda',0}[\mathbb{1}^{*l}(T_j)\ \mathbb{1}_{\{0\}}(d_{T_j})] \\
&\leqslant \sum_{l=0}^{q} [A^{j+1}]_{k,l}\mathbb{E}_{\lambda,\lambda',0}[\mathbb{1}^{*l}(T_0)].
\end{aligned}$$

We now conclude the proof. Since $P_{\lambda,\lambda',0}(d_{T_j}=0) = (1-\varepsilon)^j$, $P_{\lambda,\lambda',0}(d_{T_j}=0$ infinitely often$)=0$. Thus, the sets $\{\{d_{T_j}=0\} \cap \{d_{T_j+1}=1\}\}$, $j \geqslant 0$, define a partition of Ω; since $P_{\lambda,\lambda',0}(d_{T_j+1}=1|\mathscr{F}_{T_j})\mathbb{1}_{\{0\}}(d_{T_j}) = \varepsilon$, then

$$\begin{aligned}
\mathbb{E}_{\lambda,\lambda',0}[\mathbb{1}^{*k}(T-1)] &= \sum_{j\geqslant 0} \mathbb{E}_{\lambda,\lambda',0}[\mathbb{1}^{*k}(T_j)\mathbb{1}_{\{0\}}(d_{T_j})\mathbb{1}_{\{1\}}(d_{T_j+1})] \\
&= \varepsilon \sum_{j\geqslant 0} \mathbb{E}_{\lambda,\lambda',0}[\mathbb{1}^{*k}(T_j)\ \mathbb{1}_{\{0\}}(d_{T_j})] \\
&\leqslant \varepsilon \sum_{l=0}^{q} B_{k,l}\ \mathbb{E}_{\lambda,\lambda',0}[\mathbb{1}^{*l}(T_0)]. \qquad \square
\end{aligned}$$

Since B is lower triangular, $\mathbb{E}_{\lambda,\lambda',0}[\mathbb{1}^{*k}(T-1)]$ is a linear combination of the moments of the hitting time on the set Δ, $\mathbb{E}_{\lambda,\lambda',0}[\mathbb{1}^{*l}(T_0)]$ for $l \in \{0,\dots,k\}$, and the estimation of the polynomial moments of the coupling time boils down to estimate the corresponding moments of T_0. Following the discussions in Section 1, these quantities can be estimated from Proposition 3, provided that one can construct a set of drift functions satisfying $D[1,k,\Delta]$ for the extended transition kernel P^*. Unfortunately, deriving drift conditions for P^* is hopeless: it is more appropriate to formulate drift conditions for the original kernel P. The fact that drift conditions for P can be used to estimate moments of the hitting time of Z to Δ is of course ultimately linked with the fact that, prior to entering Δ (or between two visits to Δ before coupling), the extended chain Z behaves as two independent copies of the original chain.

Let q be a positive integer and $C, D \in \mathscr{B}(\mathscr{X})$, $C \subseteq D$.

H2$[q,C,D]$ There exist some measurable functions on $\mathscr{X}$, $1 \leqslant V_0 \leqslant \cdots \leqslant V_q < \infty$ and some constants $0 < a_k < 1$ and $b_k < \infty$, $k \in \{0,\dots,(q-1)\}$, such that $\sup_D V_q < \infty$ and for all $k \in \{0,\dots,(q-1)\}$,

$$PV_{k+1} \leqslant V_{k+1} - V_k + b_k \mathbb{1}_C, \tag{49}$$

$$V_k \geqslant b_k + a_k V_k \quad \text{on } D^c. \tag{50}$$

We know from Proposition 3 that $V_k(x)$ is an upper bound of the moment $\mathbb{E}_x[\mathbb{1}^{*k}(\sigma_C)]$ of the hitting-time on C of the original chain Φ. The additional minorization condition (50) is required to derive an upper bound of the hitting-time on Δ of the extended chain Z starting from $\mathscr{X} \times \mathscr{X} \times \{0\}$.

Proposition 7. *Let $C, D \in \mathscr{B}(\mathscr{X})$, $C \subseteq D$ and q be a positive integer. Assume* H2 $[q,C,D]$. *Then, for any $l \in \{0,\dots,(q-1)\}$ and any $(x,x') \notin \Delta$*

$$\left(\prod_{k=0}^{l} a_k\right) \mathbb{E}_{x,x',0}\left[\sum_{k=0}^{T_0} \mathbb{1}^{*l}(k+1)\{V_0(X_k)+V_0(X'_k)\}\right] \leqslant V_{l+1}(x)+V_{l+1}(x') \tag{51}$$

so that

$$\mathbb{E}_{x,x',0}[\mathbb{1}^{*(l+1)}(T_0+1)] \leqslant \mathbb{1}_\Delta(x,x') + \mathbb{1}_{\Delta^c}(x,x')\left(\prod_{k=0}^{l} a_k\right)^{-1} m(V_0)^{-1}(V_{l+1}(x)+V_{l+1}(x')),$$

where $m(V_0) := \inf_{(x,x')\in\Delta^c}\{V_0(x)+V_0(x')\}$.

Proof. The first step is to prove that for any $l \in \{0,\dots,(q-1)\}$, $(x,x') \in \Delta^c$,

$$a_l\ \mathbb{E}_{x,x',0}\left[\sum_{k=0}^{T_0}\{V_l(X_k)+V_l(X'_k)\}\right] \leqslant V_{l+1}(x)+V_{l+1}(x'). \tag{52}$$

Define $W_l(x,x',d) = W_l(x,x') := V_l(x)+V_l(x')$, $d \in \{0,1\}$. Then for $(x,x') \notin \Delta$,

$$\begin{aligned} P^* W_{l+1}(x,x',0) &\leqslant W_{l+1}(x,x') - W_l(x,x') + b_l\{\mathbb{1}_C(x)+\mathbb{1}_C(x')\} \\ &\leqslant W_{l+1}(x,x') - W_l(x,x') + b_l\{\mathbb{1}_{C\times D^c}(x,x')+\mathbb{1}_{D^c\times C}(x,x')\}. \end{aligned}$$

Note that, for $x \in D^c$, $b_l \leqslant (1-a_l)V_l(x)$. Thus,

$$P^* W_{l+1}(x,x',0) \leqslant W_{l+1}(x,x') - W_l(x,x'), \quad (x,x') \in C^c \times C^c,$$

$$P^* W_{l+1}(x,x',0) \leqslant W_{l+1}(x,x') - V_l(x) - V_l(x') + (1-a_l)V_l(x'), \quad (x,x') \in C \times D^c,$$

$$P^* W_{l+1}(x,x',0) \leqslant W_{l+1}(x,x') - V_l(x) - V_l(x') + (1-a_l)V_l(x), \quad (x,x') \in D^c \times C.$$

Combining these inequalities, we thus have for $(x,x') \in \Delta^c$,

$$P^* W_{l+1}(x,x',0) \leqslant W_{l+1}(x,x') - a_l W_l(x,x'),$$

which implies Eq. (52) by use of the Comparison Theorem. Note that since Eq. (52) T_0 is finite $P_{x,x',0}$-a.e. The proof is concluded by induction: Eq. (52) establishes the proposition for $l=0$. Assume that Eq. (51) holds for some $l \in \{0,\dots,(q-2)\}$, i.e. for all $(x,x') \in \Delta^c$,

$$\left(\prod_{k=0}^{l} a_k\right) \mathbb{E}_{x,x',0}\left[\sum_{k=0}^{T_0} \mathbb{1}^{*l}(k+1)\{V_0(X_k)+V_0(X'_k)\}\right] \leqslant V_{l+1}(x) + V_{l+1}(x'). \quad (53)$$

Then, for $(x,x') \in \Delta^c$, Eq. (52), the induction assumption Eq. (53) and the Markov property imply

$$\begin{aligned}
&\{V_{l+2}(x) + V_{l+2}(x')\} \\
&\quad \geqslant a_{l+1}\mathbb{E}_{x,x',0}\left[\sum_{k=0}^{T_0}\{V_{l+1}(X_k)+V_{l+1}(X'_k)\}\right] \\
&\quad \geqslant a_{l+1}\left(\prod_{k=0}^{l} a_k\right) \mathbb{E}_{x,x',0}\left[\sum_{k=0}^{T_0}\mathbb{E}_{X_k,X'_k,0}\left[\sum_{j=0}^{T_0}\mathbb{1}^{*l}(j+1)\{V_0(X_j)+V_0(X'_j)\}\right]\right] \\
&\quad \geqslant \left(\prod_{k=0}^{l+1} a_k\right) \mathbb{E}_{x,x',0}\left[\sum_{k=0}^{T_0}\mathbb{1}^{*(l+1)}(k+1)\{V_0(X_k)+V_0(X'_k)\}\right],
\end{aligned}$$

showing the induction hypothesis and thus concluding the proof. □

We have now at hands all the necessary ingredients to evaluate the bound $B(x,n)$. Define $W_0(x,x') := 1$ and, for $l \in \{1,\dots,q\}$,

$$W_l(x,x') := \mathbb{1}_{\Delta}(x,x') + \mathbb{1}_{\Delta^c}(x,x')\left(\prod_{k=0}^{l-1} a_k\right)^{-1} m(V_0)^{-1}\{V_l(x)+V_l(x')\}, \quad (54)$$

$$W(x,x') := [W_0(x,x'),\dots,W_q(x,x')]^{\mathrm{T}}. \quad (55)$$

Set $\delta_x \otimes \pi(W) := \int \delta_x(\mathrm{d}y)\pi(\mathrm{d}y')W(y,y')$. For $l \in \{0,\dots,q\}$, denote by

$$\hat{A}(l) := (1-\varepsilon) \sup_{(x,x') \in \Delta} \int R(x,\mathrm{d}y)R(x',\mathrm{d}y')W_l(y,y'),$$

and $\hat{A}(l) := 0$ otherwise. Define $\hat{A} := [\hat{A}_{i,j}]_{0 \leqslant i,j \leqslant q}$, where $\hat{A}_{i,j} := \hat{A}(i-j)$. Set $\hat{B} := (I - \hat{A})^{-1}$. As in Remark 10, it may be proved that for all $0 \leqslant k \leqslant q$, $\varepsilon \hat{B}_{k,k} = 1$, and for $0 \leqslant l < k \leqslant q$, $\varepsilon \hat{B}_{k,l} \leqslant (1-\varepsilon)\ \hat{\alpha}^{(k-l)}\ \varepsilon^{-(k-l)}$ where

$$\hat{\alpha} := \sup_{l \in \{1,\dots,q\},(x,x') \in \Delta} \int R(x,\mathrm{d}y) R(x',\mathrm{d}y') W_l(y,y')$$
$$\leqslant \sup_{(x,x') \in \Delta} \int R(x,\mathrm{d}y) R(x',\mathrm{d}y') W_q(y,y'). \tag{56}$$

Finally, if $l > k$, $\hat{B}_{k,l} = 0$.

Theorem 2. *Assume that P is ψ-irreducible and aperiodic. Let q be a positive integer and $C, D \in \mathscr{B}(\mathscr{X})$, $C \subseteq D$ such that* H1a$[C,D]$ *and* H2$[q,C,D]$ *hold. There exists an unique invariant probability measure π and for all $x \in \mathscr{X}$, $n \geqslant 1$,*

$$\|P^n(x,.) - \pi(.)\|_{\mathrm{TV}} \leqslant 2\varepsilon \min_{k \in \{0,\dots,(q-1)\}} \left\{ U_{\varepsilon,1}(\mathbb{1}^{*k};n) \sum_{l=0}^{q} \hat{B}_{k,l} \delta_x \otimes \pi(W_l) \right\}, \tag{57}$$

where $U_{\varepsilon,1}(\mathbb{1}^{*k};.)$ is given by Eq. (44).

The proof follows from Eq. (42), Lemma 4 and Proposition 6. $U_{\varepsilon,1}(\mathbb{1}^{*k};n)$ is equivalent to n^{-k}; the maximum rate of convergence is thus of order $n^{(q-1)}$. In comparison, application of Proposition 1 under the assumptions of Theorem 2 shows that $\lim_{n\to\infty} n^{(q-1)} \|P^n(x,.) - \pi(.)\|_{\mathrm{TV}} = 0$, which is indeed, stronger than Eq. (57) from an asymptotic standpoint.

Computable bound for the NSS model. For the purpose of numerical illustration, it is assumed that

$$F(x) := x(1 - r/|x|), \quad |x| \geqslant M,$$
$$F(x) := 0.1x^3 + (1 - r/M - 0.1M^2)x, \quad |x| \leqslant M,$$

with $r := 3$ and $M := 2$ (see Fig. 2). We assume in addition that Γ is Pareto in the tails, i.e.

$$\gamma(y) \propto M_\gamma^{-(g+1)} \mathbb{1}_{[-M_\gamma, M_\gamma]}(y) + |y|^{-(g+1)} \mathbb{1}_{[-M_\gamma, M_\gamma]^c}(y),$$

with $M_\gamma := 1$ and $g := 4.1$ (see in Figs. 1 and 2 a typical trace of 4000 samples). For this model, the conditions (NSS1–3) are verified since $\gamma(y)$ is continuous and positive on $\mathbb{R}$, and F is continuous on $\mathbb{R}$; and the condition (NSS 4) holds for all $1 \leqslant s_* < g$. Lemma 3 shows that for all $1 \leqslant s < g$, for all $0 < \lambda < 1$, there exist some constant $b < \infty$ and a petite set C such that $\int P(x,\mathrm{d}y)|y|^s \leqslant |x|^s - \lambda s r |x|^{s-1} + b\mathbb{1}_C(x)$.

Let $\lambda := 1/20$. We set

$$V_0(x) := 1 + |x|, \quad V_1(x) := (2r\lambda)^{-1}(0.33 + |x|^2),$$
$$V_2(x) := (2r\lambda 3r\lambda)^{-1}(0.16 + |x|^3), \quad V_3(x) := (2r\lambda 3r\lambda 4r\lambda)^{-1}(0.11 + |x|^4)$$

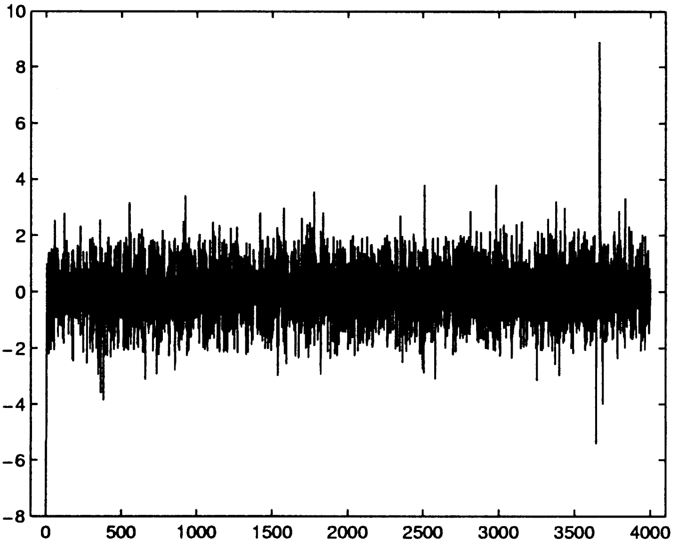

Fig. 1. NSS model, Example 2.2: a path of the Markov chain $\{\phi_n\}$, $\phi_{n+1} = F(\phi_n) + W_{n+1}$ starting from -8. $\{W_n\}$ are i.i.d. Pareto samples (and independent of ϕ_0) with density $p(y) \sim |y|^{-5.1}$; F is plotted in Fig. 2.

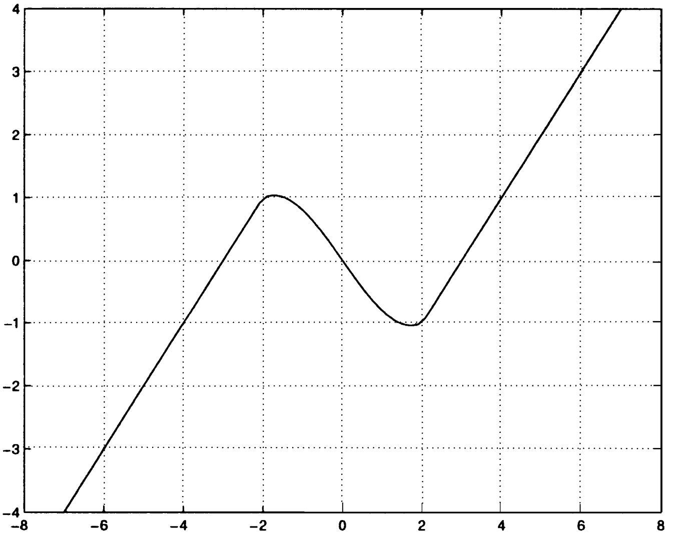

Fig. 2. F function: $F(x) := x(1 - 3/|x|)$ if $|x| \geqslant 2$ and $F(x) = 0.1x^3 - 0.9x$ otherwise.

so that $1 \leqslant V_0 \leqslant V_1 \leqslant V_2 \leqslant V_3$. The drift conditions $PV_{k+1} \leqslant V_{k+1} - V_k + b_k \mathbb{1}_C$, $k \in \{0,1,2\}$, are verified with

$$b_0 := 3.47, \quad b_1 := 13.9, \quad b_2 := 138, \quad C := [-2.06, 2.06].$$

For $a_k := 3/7$, the inequality $V_k - b_k \geqslant a_k V_k$ on D^c is verified, for $k \in \{0,1,2\}$, by setting

$$D := [-5.1, 5.1] \supset C.$$

For all $x \in [-a, a]$, it is easily seen that

$$P(x, \mathrm{d}y) \geqslant \mathbb{1}_{(-\infty,0]}(y)\ \gamma(y - F_*)\,\mathrm{d}y + \mathbb{1}_{[0,\infty)}(y)\ \gamma(y + F_*)\,\mathrm{d}y,$$

with $F_* := \sup_{[-a,a]} F = -\min_{[-a,a]} F$. Hence, H1a$[C,D]$ is verified with $\varepsilon := 9.36 \times 10^{-3}$ and

$$\varepsilon \nu(\mathrm{d}y) := \mathbb{1}_{(-\infty,0]}(y)\ \gamma(y - 2.1)\,\mathrm{d}y + \mathbb{1}_{[0,\infty)}(y)\ \gamma(y + 2.1)\,\mathrm{d}y.$$

We need finally to determine an upper bound for $\pi(V_k)$, $k \in \{1,2\}$. Since π is not known, it is required to use for such purpose drift conditions. Recall indeed that, for a ψ-irreducible and aperiodic kernel P, if there exist a petite set K and two functions $0 < V, f < \infty$ such that $PV \leqslant V - f + b\mathbb{1}_K$ for some $b < \infty$, then $\pi(f) \leqslant b$ (Meyn and Tweedie, 1993, Theorem 14.3.7). We can of course use $V = V_2$ (resp. $V = V_3$), $f = V_1$ (resp. $f = V_2$) and $K = C$, but this choice does not necessarily provide the best bound. Note indeed that the bound for $\pi(V_k)$ does not depend upon the minorization constant on the petite set K, whereas the choice of C results from a compromise (choosing C too large yields vanishingly small minorization constant). By crude optimization on K, we obtain, $\pi(V_1) \leqslant 2.14$ and $\pi(V_2) \leqslant 6.99$.

In Figs. 3 and 4, we plot the bound given by Theorem 2 for different values of the number of iterations n and of the starting point x. For a given n, the bound is an increasing function of x and for given x, the bound is a decreasing function of n. For large x and small n, the computed bound does not improve the trivial one that is $B(x,n) = 2$ (reached with $l = 0$ in Eq. (57)).

Interpolated rates. Assumption H2$[q,C,D]$ does not imply $\pi(V_q) < \infty$; hence, $\sum_{l=0}^{q} \hat{B}_{q,l}\ \delta_x \otimes \pi(W_l)$ may well be infinite. If for some $0 < \beta \leqslant 1$, $\pi(V_q^\beta) < \infty$ it is possible (similar to Proposition 5) to sharpen the bounds. Define for $l \in \{0,\dots,q\}$,

$$\hat{A}^{(\beta)}(l) := (1-\varepsilon) \sup_{(x,x') \in \Delta} \int R(x,\mathrm{d}y) R(x',\mathrm{d}y') W_l^\beta(y,y'),$$

and $\hat{A}^{(\beta)}(l) := 0$ otherwise. Denote by $\hat{A}^{(\beta)} := [\hat{A}^{(\beta)}_{i,j}]_{0 \leqslant i,j \leqslant q}$, $\hat{A}^{(\beta)}_{i,j} := \hat{A}^{(\beta)}(i-j)$. Finally, set $\hat{B}^{(\beta)} := (I - \hat{A}^{(\beta)})^{-1}$. As in Remark 10, it may be proved that for all $0 \leqslant k \leqslant q$, $\varepsilon \hat{B}^{(\beta)}_{k,k} = 1$, and for $0 \leqslant l < k \leqslant q$, $\varepsilon \hat{B}^{(\beta)}_{k,l} \leqslant (1-\varepsilon)\ \bar{\alpha}^{k-l}\ \varepsilon^{-(k-l)}$ where

$$\begin{aligned} \bar{\alpha} &:= \sup_{l \in \{1,\dots,q\}} \sup_{(x,x') \in \Delta} \int R(x,\mathrm{d}y) R(x',\mathrm{d}y') W_l^\beta(y,y') \\ &\leqslant \sup_{(x,x') \in \Delta} \int R(x,\mathrm{d}y) R(x',\mathrm{d}y') W_q^\beta(y,y'). \end{aligned} \tag{58}$$

Finally, if $l > k$, $\hat{B}^{(\beta)}_{k,l} = 0$.

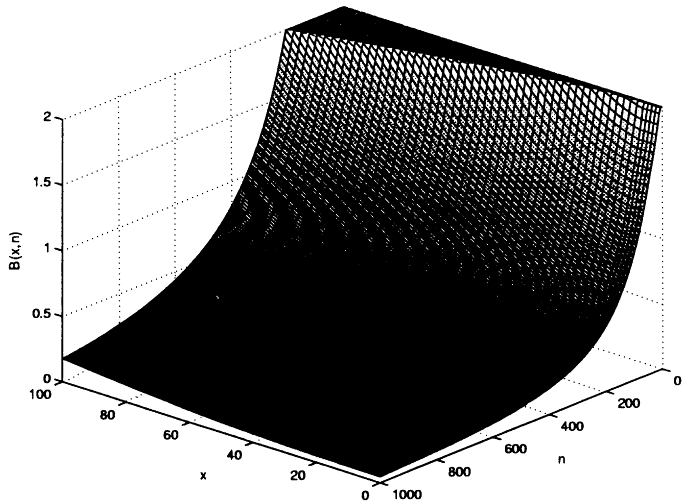

Fig. 3. NSS model, Example 2.2: control of ergodicity in total variation norm given by Theorem 2 for the NSS Markov chain. The upper-bound $B(x,n)$ is plotted for different values of $x \in [0,\ 100]$ and different values of $n \in [1,\ 1000]$. For large x and small n, the bound is not lower than 2, which is a trivial upper bound.

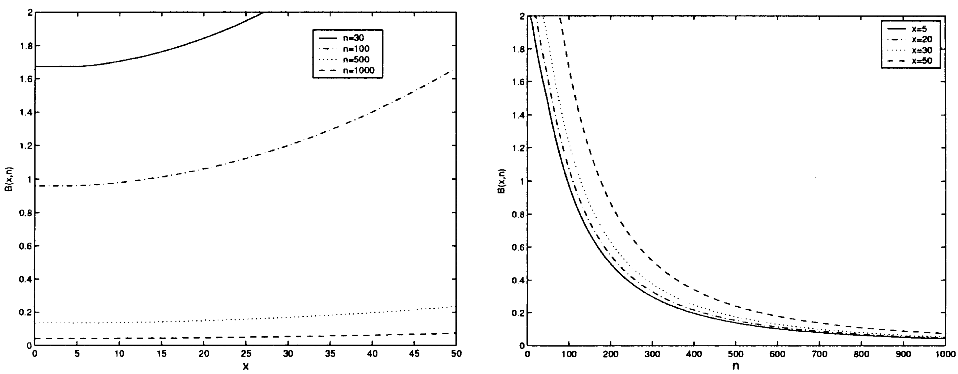

Fig. 4. NSS model, Example 2.2: Evolution of the upper-bound $B(x,n)$. (left) $B(.,n)$ for different values of n; $B(x,.)$ for different values of x.

Theorem 3. *Assume that P is ψ-irreducible and aperiodic. Let q be a positive integer and $C, D \in \mathscr{B}(\mathscr{X})$, $C \subseteq D$ such that* H1a$[C,D]$ *and* H2$[q,C,D]$ *hold, and assume that $\pi(V_q^\beta) < \infty$ for some $0 < \beta \leqslant 1$. For all $x \in \mathscr{X}$, $n \geqslant 1$,*

$$\|P^n(x,.) - \pi(.)\|_{\mathrm{TV}} \leqslant 2\varepsilon \min_{k \in \{0,\dots,q\}} \left\{ U_{\varepsilon,1}((\mathbb{1}^{*k})^\beta; n) \sum_{l=0}^{q} \hat{B}_{k,l}^{(\beta)}\, \delta_x \otimes \pi(W_l^\beta) \right\}. \tag{59}$$

Proof. The proof is based on Eq. (42) and Lemma 4. As above, the moment of the coupling time may be expressed as a series of the moments of the successive hitting times on Δ

$$\mathbb{E}_{x,x',0}[\mathbb{1}^{*k}(T-1)^{\beta}] = \sum_{j\geqslant 0} \mathbb{E}_{x,x',0}[\mathbb{1}^{*k}(T_j)^{\beta}\mathbb{1}_{\{0\}}(d_{T_j})\mathbb{1}_{\{1\}}(d_{T_j+1})]$$

$$= \varepsilon \sum_{j\geqslant 0} \mathbb{E}_{x,x',0}[\mathbb{1}^{*k}(T_j)^{\beta}\mathbb{1}_{\{0\}}(d_{T_j})].$$

The proof follows provided that we may show that, for any non-negative integer j,

$$\mathbb{E}_{x,x',0}[\mathbb{1}^{*k}(T_j)^{\beta}\mathbb{1}_{\{0\}}(d_{T_j})] \leqslant \sum_{l=0}^{q} [(\hat{A}^{(\beta)})^j]_{k,l}\ \delta_x \otimes \pi(W_l^{\beta}).$$

This last equation can be established by induction along the same lines as in Proposition 6, using, for $k \geqslant 1$, $n+m \geqslant 1$,

$$(\mathbb{1}^{*k}(m+n))^{\beta} \leqslant \sum_{l=0}^{k} (\mathbb{1}^{*l}(n))^{\beta}(\mathbb{1}^{*(k-l)}(m))^{\beta}. \qquad \square$$

Remark 12. The order of $U_{\varepsilon,1}((\mathbb{1}^{*q})^{\beta};n)$ is $n^{-q\beta}$, and Eq. (59) gives an explicit bound for $\sup_{n\geqslant 1} n^{q\beta}\|P^n(x,.)-\pi(.)\|_{\mathrm{TV}}$ (which implies that $\lim_n n^{\gamma}\|P^n(x,.)-\pi(.)\|_{\mathrm{TV}} = 0$, for all $0 \leqslant \gamma < q\beta$). Once again, application of Proposition 5 shows, under the stated assumption, that $\lim_n n^{q\beta}\|P^n(x,.)-\pi(.)\|_{\mathrm{TV}} = 0$, which is stronger from an asymptotic standpoint.

Remark 13. Successful coupling can be achieved in sets which are not necessarily small. For example, instead of H1a$[C,D]$, we may assume that there exists a kernel $\rho_{x,x'}(.)$ from Δ to $\mathscr{X}$ such that, for all $(x,x') \in \Delta$, and all $A \in \mathscr{B}(\mathscr{X})$,

$$P(x,A) \wedge P(x',A) \geqslant \rho_{x,x'}(A) \quad \text{and} \quad \varepsilon^- := \inf_{(x,x')\in\Delta} \rho_{x,x'}(\mathscr{X}) > 0. \tag{60}$$

This condition is verified by setting, for any kernel $\mu_{x,x'}(\mathrm{d}y)$ from Δ to $\mathscr{X}$,

$$\rho_{x,x'}(A) := \int_A \frac{\mathrm{d}P(x,.)}{\mathrm{d}\mu_{x,x'}}(y) \wedge \frac{\mathrm{d}P(x',.)}{\mathrm{d}\mu_{x,x'}}(y)\, \mu_{x,x'}(\mathrm{d}y),$$

where $\mathrm{d}P(x,.)/\mathrm{d}\mu_{x,x'}$ is the derivative of the absolutely continuous part of $P(x,\mathrm{d}y)$ w.r.t. $\mu_{x,x'}(\mathrm{d}y)$. By choosing $\mu_{x,x'}(\mathrm{d}y) := P(x',\mathrm{d}y)$, (60) is equivalently written as

$$\varepsilon^- := \inf_{(x,x')\in\Delta} \int \left\{ 1 \wedge \frac{\mathrm{d}P(x,\mathrm{d}y)}{\mathrm{d}P(x',\mathrm{d}y)} \right\} P(x',\mathrm{d}y) > 0,$$

a condition referred to as a *local Doeblin condition* in Veretennikov (2000). Extensions to local Doeblin conditions can be obtained along the same line as above, and yield alternate bounds. This is addressed in Fort (2001).

Remark 14. In previous contributions on computational bounds for geometrical ergodicity, the set D is chosen to be equal to C, where C is ν_1-small. Assumption H2$[q,C,D]$

is replaced by (a) a Foster–Lyapunov drift criterion, i.e. there exists some Borel function $V \geqslant 1$ such that $PV \leqslant \lambda V + b\mathbb{1}_C$ with $0 < \lambda < 1$ and $b < \infty$ and (b) a minorization condition on C^c of the drift function V, which involves a rather tricky relation between the seemingly unrelated constants λ, b and $\sup_C V$,

$$\sup_C V \geqslant \frac{b}{2(1-\lambda)} - 1, \tag{61}$$

(see for example, Theorem 12 of Rosenthal, 1995a; and Eq. (49) of Roberts and Tweedie, 1996). When the condition Eq. (61) is not checked, it is suggested to choose a ν_1-small set larger than C, but there is no guarantee that Eq. (61) can be verified: it is generally difficult to control the relative rate of growth of $\sup_C V$ and b as a function of C (note that the drift function V itself may also depend on C). In addition, increasing C has an adverse effect on the minorizing constant ε in the smallness condition, which may become unacceptably small. Note finally that while for ψ-irreducible chain, the state space can be covered by a countable family of small sets, there is no guarantee that the state space can be covered by a ν_1-small set, which means that it is not possible to choose arbitrarily large ν_1-small set.

We would like to argue that choosing $C \neq D$ answers some of the problems outlined above. In most cases the minorization condition Eq. (50) holds by choosing $D \supset C$ large enough as V is often unbounded off petite set; since C is fixed here, enlarging D does not modify the constants involved in the "basic" drift condition (contrary to the solution outlined above). Nevertheless, it may happen that increasing D in such a way that Eq. (50) above holds, yields to sets that are no longer ν_1-small but rather ν_m-small for some $m > 1$. This problem is answered below.

The adaptations to the case where D is ν_m-small instead of ν_1-small are not straightforward. The construction and the main results are derived below. We substitute the assumption H1a$[C,D]$ for the condition H1b$[C,D]$

H1b$[C,D]$ C is accessible and D is ν_m-small: there exist $m > 1$, a constant $\varepsilon > 0$ and a probability measure ν_m on $\mathscr{B}(\mathscr{X})$ such that

$$\forall x \in D, \qquad \forall A \in \mathscr{B}(\mathscr{X}), \qquad P^m(x,A) \geqslant \varepsilon \nu_m(A). \tag{62}$$

When D is ν_m-small, the coupling construction should be adapted (see e.g. Rosenthal, 2001 or Kalashnikov, 1994). We briefly present the necessary adaptations. Define the residual kernel

$$R_m(x, \mathrm{d}y) := (1 - \varepsilon\mathbb{1}_D(x))^{-1}(P^m(x, \mathrm{d}y) - \varepsilon\mathbb{1}_D(x)\nu_m(\mathrm{d}y)). \tag{63}$$

The coupling construction is slightly more complicated, and involves the definition of two processes on the extended probability space, Z and $\tilde{Z}$. The process Z is a Markov chain defined as above substituting the residual kernel R for R_m, but this time, Z does not verify the condition Eq. (41). From this process Z, a companion process $\tilde{Z}$ is defined by an appropriately chosen random change of time in such a way that

Eq. (41) holds. This construction is outlined below. Set $Q_0 := 0$ and $\tilde{Z}_0 := Z_0$. Then, $\tilde{Z}$ is defined from Z recursively as follows:

(1) if $d_n = 0$,
- if $(X_n, X'_n) \notin \Delta$, set $Q_{n+1} := Q_n + 1$ and $\tilde{Z}_{Q_{n+1}} := Z_{n+1}$.
- if $(X_n, X'_n) \in \Delta$, set $Q_{n+1} := Q_n + m$ and $\tilde{Z}_{Q_{n+1}} := Z_{n+1}$. Sample $\{\tilde{X}_{Q_n+k}\}$, for $k \in \{1, \dots, (m-1)\}$ from the distribution of $P(\Phi_1 \in \cdot, \dots, \Phi_{m-1} \in \cdot | \Phi_m = X_{n+1}, \Phi_0 = X_n)$ where $\{\Phi_n\}$ is a Markov chain with transition kernel P; sample similarly and independently $\{\tilde{X}'_{Q_n+k}\}$, $k \in \{1, \dots, (m-1)\}$. Finally, set $\tilde{d}_{Q_n+k} := \tilde{d}_{Q_n}$, $k \in \{1, \dots, (m-1)\}$.

(2) if $d_n = 1$, set $Q_{n+1} := Q_n + 1$ and $\tilde{Z}_{Q_{n+1}} := Z_{n+1}$.

In words, every time $Z \in \Delta \times \{0\}$, a gap of size $(m-1)$ is inserted and filled in conditionally independently from the conditional distribution of the Markov chain given the *initial* and *final* values. The resulting process $\tilde{Z}$ is no longer a Markov chain. We denote respectively by $P_{\lambda,\lambda',0}$ and $\tilde{P}_{\lambda,\lambda',0}$ the probability distributions generated by the processes Z and $\tilde{Z}$ on the canonical space. When $Z_0 \sim \lambda \otimes \lambda' \otimes \delta_0$, it is easily verified that, for any $n \geqslant 0$, for any positive Borel function f,

$$\tilde{\mathbb{E}}_{\lambda,\lambda',0}[f(\tilde{X}_n)] = \lambda P^n f \quad \text{and} \quad \tilde{\mathbb{E}}_{\lambda,\lambda',0}[f(\tilde{X}'_n)] = \lambda' P^n f.$$

The coupling-time is now defined as $T := \inf\{n \geqslant 1, \tilde{d}_n = 1\}$: thus, $T \geqslant m$ $P_{\lambda,\lambda',0}$ a.e., and with this definition, the Lindvall's inequality Eq. (42) still applies. Define

$$A_m(l) := (1-\varepsilon) \sup_{(x,x') \in \Delta} \int R_m(x, \mathrm{d}y) R_m(x', \mathrm{d}y') \mathbb{E}_{y,y',0}[\mathbb{1}^{*l}(T_0 + m)].$$

Using the property Eq. (8) of the sequence $\mathbb{1}^{*l}$, Proposition 7 and the definition Eq. (54) of W_i, $A_m(l)$ is bounded for $l \geqslant 1$ by

$$\hat{A}_m(l) := (1-\varepsilon) \sup_{(x,x') \in \Delta} \sum_{i=0}^{l} \mathbb{1}^{*(l-i)}(m-1) \int R_m(x, \mathrm{d}y) R_m(x', \mathrm{d}y') W_i(y, y'). \quad (64)$$

Set $\hat{A}_m(0) := 1 - \varepsilon$ and $\hat{A}_m(l) := 0$ otherwise. Denote by $\hat{A}_m := [\hat{A}_{m;i,j}]_{0 \leqslant i,j \leqslant q}$ where $\hat{A}_{m;i,j} := \hat{A}_m(i-j)$. Set $\hat{B}_m := (I - \hat{A}_m)^{-1}$. An upper bound for $\hat{B}_{m;k,l}$ can be computed similarly to what is done for $\hat{B}_{k,l}$, by substituting $\hat{\alpha}$ given by (56) for

$$\hat{\alpha}_m := \sup_{l \in \{1,\dots,q\}} \sup_{(x,x') \in \Delta} \sum_{i=0}^{l} \mathbb{1}^{*(l-i)}(m-1) \int R_m(x, \mathrm{d}y) R_m(x', \mathrm{d}y') W_i(y, y').$$

Theorem 4. *Let q be a positive integer and let $C, D \in \mathscr{B}(\mathscr{X})$, $C \subseteq D$. Assume that P is ψ-irreducible aperiodic and that* H1b$[C,D]$ *and* H2$[q,C,D]$ *hold. There exists an unique invariant probability measure π and for all $x \in \mathscr{X}$, $n \geqslant m$,*

$$\|P^n(x,.) - \pi(.)\|_{\mathrm{TV}} \leqslant 2\varepsilon \min_{k \in \{0,\dots,(q-1)\}} \left\{ U_{\varepsilon,m}(\mathbb{1}^{*k}; n) \sum_{l=0}^{q} \hat{B}_{m;k,l} \delta_x \otimes \pi(W_l) \right\}.$$

Proof. The proof follows from the coupling inequality Eq. (42) and Lemma 4 upon showing

$$\mathbb{E}_{x,x',0}[\mathbb{1}^{*k}(T-m)] \leqslant \varepsilon \sum_{l=0}^{q} \hat{B}_{m;k,l}\ W_l(x,x'). \tag{65}$$

Set $T_j' := T_j + j(m-1)$, $j \geqslant 0$. We first prove by induction that for all $j \geqslant 0$, and all $l \in \{0,\dots,q\}$,

$$\mathbb{E}_{x,x',0}[\mathbb{1}^{*k}(T_j')\ \mathbb{1}_{\{0\}}(d_{T_j})] \leqslant \sum_{l=0}^{q} (\hat{A}_m^j)_{k,l}\ W_l(x,x'). \tag{66}$$

For $j=0$, the property holds by Proposition 6. Assume that Eq. (66) holds for some $j \geqslant 0$. Note that $d_{T_{j+1}} = d_{T_j+1}$ and $T_{j+1}' - T_j' = T_{j+1} - T_j + m - 1 = m + T_0 \circ \theta^{1+T_j}$. Proceeding as above, we have

$$\begin{aligned}
&\mathbb{E}_{x,x',0}[\mathbb{1}^{*k}(T_{j+1}')\mathbb{1}_{\{0\}}(d_{T_{j+1}})]\\
&= \sum_{l=0}^{k} \mathbb{E}_{x,x',0}[\mathbb{1}^{*l}(T_j')\ \mathbb{1}^{*(k-l)}(T_{j+1} - T_j + m - 1)\ \mathbb{1}_{\{0\}}(d_{T_{j+1}})]\\
&= \sum_{l=0}^{k} \mathbb{E}_{x,x',0}[\mathbb{E}_{x,x',0}[\mathbb{E}_{x,x',0}[\mathbb{1}^{*(k-l)}(m + T_0 \circ \theta^{T_j+1})|\mathscr{F}_{T_j+1}]\mathbb{1}_{\{0\}}(d_{T_j+1})|\mathscr{F}_{T_j}]\\
&\quad \times \mathbb{1}^{*k}(T_j')\ \mathbb{1}_{\{0\}}(d_{T_j})].
\end{aligned}$$

The strong Markov property implies that

$$\begin{aligned}
&\mathbb{E}_{x,x',0}[\mathbb{1}^{*k}(T_{j+1}')\ \mathbb{1}_{\{0\}}(d_{T_{j+1}})]\\
&= \sum_{l=0}^{k} \mathbb{E}_{x,x',0}\Bigg[(1-\varepsilon)\int R_m(X_{T_j},\mathrm{d}z)R_m(X_{T_j}',\mathrm{d}z')\\
&\quad \times \mathbb{E}_{z,z',0}[\mathbb{1}^{*(k-l)}(m+T_0)]\mathbb{1}^{*l}(T_j')\ \mathbb{1}_{\{0\}}(d_{T_j})\Bigg]\\
&\leqslant \sum_{l=0}^{k} \hat{A}_m(k-l)\mathbb{E}_{x,x',0}[\mathbb{1}^{*l}(T_j')\ \mathbb{1}_{\{0\}}(d_{T_j})]\\
&\leqslant \sum_{l=0}^{q} (\hat{A}_m^j)_{k,l}\mathbb{E}_{x,x',0}[\mathbb{1}^{*l}(T_j')\ \mathbb{1}_{\{0\}}(d_{T_j})].
\end{aligned}$$

By use of the induction assumption, it holds

$$\mathbb{E}_{x,x',0}[\mathbb{1}^{*k}(T_{j+1}')\ \mathbb{1}_{\{0\}}(d_{T_{j+1}})] \leqslant \sum_{l=0}^{q} [\hat{A}_m^{j+1}]_{k,l}\, W_l(x,x').$$

We now conclude the proof as in Proposition 6. $P_{x,x',0}$-a.e., the sequence $\{d_{T_j}\}$ is equal to zero for finitely many values of j and $\{\{d_{T_j}=0\} \cap \{d_{T_{j+1}}=1\}\}$ define a partition of Ω. Thus,

$$\mathbb{E}_{x,x',0}[\mathbb{1}^{*k}(T-m)] = \sum_{j\geqslant 0} \mathbb{E}_{x,x',0}[\mathbb{1}^{*k}(T'_j)\mathbb{1}_{\{0\}}(d_{T_j})\mathbb{1}_{\{1\}}(d_{T_{j+1}})]$$

$$= \varepsilon \sum_{j\geqslant 0} \mathbb{E}_{x,x',0}[\mathbb{1}^{*k}(T'_j)\ \mathbb{1}_{\{0\}}(d_{T_j})] \leqslant \varepsilon \sum_{k=0}^{q} \hat{B}_{m;k,l} W_l(x,x'). \quad \square$$

Define $\hat{A}_m^{(\beta)}(0) := (1-\varepsilon)$, for $l \in \{1,\dots,q\}$

$$\hat{A}_m^{(\beta)}(l) := (1-\varepsilon) \sup_{(x,x')\in\Delta} \sum_{i=0}^{l} \mathbb{1}^{*(l-i)}(m-1) \int R_m(x,\mathrm{d}y) R_m(x',\mathrm{d}y') W_i^{\beta}(y,y'), \quad (67)$$

and $\hat{A}_m^{(\beta)}(l) := 0$ otherwise. Denote by $\hat{A}_m^{(\beta)} := [\hat{A}_{m;i,j}^{(\beta)}]_{0\leqslant i,j\leqslant q}$ where $\hat{A}_{m;i,j}^{(\beta)} = \hat{A}_m^{(\beta)}(i-j)$. Set $\hat{B}_m^{(\beta)} := (I - \hat{A}_m^{(\beta)})^{-1}$. An upper bound for $\hat{B}_{m;k,l}^{(\beta)}$ can be computed similarly to what is done for $\hat{B}_{k,l}^{(\beta)}$, by substituting $\bar{\alpha}$ given by (58) for

$$\bar{\alpha}_m := \sup_{l\in\{1,\dots,q\}} \sup_{(x,x')\in\Delta} \sum_{i=0}^{l} \mathbb{1}^{*(l-i)}(m-1) \int R_m(x,\mathrm{d}y) R_m(x',\mathrm{d}y') W_i^{(\beta)}(y,y').$$

Theorem 5. *Assume that P is ψ-irreducible aperiodic. Let q be a positive integer and $C, D \in \mathscr{B}(\mathscr{X})$, $C \subseteq D$ such that* H1b$[C,D]$ *and* H2$[q,C,D]$ *hold, and assume that there exists $0 < \beta \leqslant 1$ such that $\pi(V_q^{\beta}) < \infty$. For all $x \in \mathscr{X}$, $n \geqslant m$,*

$$\|P^n(x,.) - \pi(.)\|_{\mathrm{TV}} \leqslant 2\varepsilon \min_{k\in\{0,\dots,q\}} \left\{ U_{\varepsilon,m}((\mathbb{1}^{*k})^{\beta}; n) \sum_{l=0}^{q} \hat{B}_{m;k,l}^{(\beta)} \delta_x \otimes \pi(W_l^{\beta}) \right\}.$$

The proof is omitted for brevity.

Remark 15. It is interesting to relate the conditions stated above and the single drift conditions derived in Section 1. Let ϕ be a Borel function, q be a positive integer and C be a petite set. Assume $S[\phi,q,C]$, and $\phi_0 \circ V$ is unbounded off petite set (where ϕ_0 and V are defined in $S[\phi,q,C]$). The nested drift conditions Eq. (49) are verified with $V_q := (\prod_{l=0}^{q-1} c_l)\phi_q \circ V$, $V_k := (\prod_{l=0}^{k-1} c_l)\phi_k \circ V$, for $k \in \{1,\dots,(q-1)\}$, $V_0 := \phi_0 \circ V$, $b_{q-1} := c_q b$ and $b_k := \sup_{x\in C}[PV_{k+1} - V_{k+1} + V_k]$, for $k \in \{0,\dots,(q-2)\}$. In addition, since the constants c_k are greater than 1, then $V_0 \leqslant V_1 \leqslant \cdots \leqslant V_q$. For any $a_k \in (0,1)$, $k \in \{0,\dots,(q-1)\}$, define

$$D := C \cup \bigcup_{k=0}^{q-1} \{V_k \leqslant b_k/(1-a_k)\}.$$

It is easily verified that, with these definitions, H2$[q,C,D]$ holds. Since $V_k \geqslant V_0$, then $\{V_k \leqslant l\} \subset \{V_0 \leqslant l\}$, $k \in \{0,\dots,(q-1)\}$. By assumption, the level set $\{V_0 \leqslant l\}$ is petite (for any $l \geqslant 1$), and thus all the level sets $\{V_k \leqslant l\}$, $k \in \{0, \dots, (q-1)\}$ are petite. Hence, D is also petite (as a finite union of petite sets, Meyn and Tweedie, 1993, Proposition 5.5.5.), and hence ν_m-small for some $m \geqslant 1$, which establishes H1b$[C, D]$.

4. Computable bounds for polynomial f-ergodicity

Let $W : \mathscr{X} \to [1, \infty)$ be a Borel function. In this section, we focus on computable bounds for the W-norm, $\|P^n(x, .) - \pi(.)\|_W$ (we use here W instead of f to avoid confusion with the function f used in the nested drift conditions). Let λ, λ' be two probability measures on $\mathscr{B}(\mathscr{X})$. Under H1a$[C, D]$ or H1b$[C, D]$, we can define a coupling-time T on a probability space $(\Omega, \mathscr{F}, P_{\lambda,\lambda',0})$ and write that for all measurable function $|g| \leqslant W^{1-\alpha}$, $0 \leqslant \alpha \leqslant 1$,

$$
\begin{aligned}
|\lambda P^n g - \lambda' P^n g| &\leqslant \int \lambda(\mathrm{d}x)\lambda'(\mathrm{d}x')|P^n g(x) - P^n g(x')| \\
&\leqslant \mathbb{E}_{\lambda,\lambda',0}[\{W^{1-\alpha}(X_n) + W^{1-\alpha}(X'_n)\}\mathbb{1}_{T>n}],
\end{aligned}
$$

that is

$$
\|\lambda P^n - \lambda' P^n\|_{W^{1-\alpha}} \leqslant \mathbb{E}_{\lambda,\lambda',0}[\{W^{1-\alpha}(X_n) + W^{1-\alpha}(X'_n)\}\mathbb{1}_{T>n}].
$$

Then using the Hölder's inequality, for all $0 < \alpha < 1$,

$$
\|\lambda P^n - \lambda' P^n\|_{W^{1-\alpha}} \leqslant 2^{\alpha}\mathbb{E}_{\lambda,\lambda',0}[\{W(X_n) + W(X'_n)\}\mathbb{1}_{T>n}]^{1-\alpha} P_{\lambda,\lambda',0}(T > n)^{\alpha}, \quad (68)
$$

and this inequality remains true for $\alpha = 0$ (the case $\alpha = 1$ was addressed in Section 3). Remind that the bounds for polynomial ergodicity in total variation norm given in Section 3, are computed from an estimation of the tail probability $P_{\lambda,\lambda',0}(T > n)$. Therefore, bounds for polynomial ergodicity in W-norm are determined under an additional assumption, allowing to compute a bound for $\mathbb{E}_{x,x',0}[\{W(X_n) + W(X'_n)\}\mathbb{1}_{T>n}]$, for all $(x, x') \in \mathscr{X} \times \mathscr{X}$ (which has to be $\delta_x \otimes \pi$-integrable).

Let $W : \mathscr{X} \to [1, \infty)$ be a Borel function, q be a positive integer and $C, D \in \mathscr{B}(\mathscr{X})$, $C \subseteq D$.

H3$[W, C, D]$ $\sup_D W < \infty$ and there exist a Borel function $1 \leqslant w \leqslant W$ and a constant $b < \infty$ such that

$$
PW \leqslant W - w + b\mathbb{1}_C,
$$

$$
w \geqslant b, \quad \text{on } D^c.
$$

Proposition 8. *Let $C, D \in \mathscr{B}(\mathscr{X})$, $C \subseteq D$ and $W \geqslant 1$ be a Borel function. Assume that* H1b$[C, D]$ *and* H3$[W, C, D]$ *hold. Then, for all $(x, x') \in \mathscr{X} \times \mathscr{X}$, and all $n \geqslant 0$,*

$$
\mathbb{E}_{x,x',0}[\{W(X_n) + W(X'_n)\}\mathbb{1}_{T>n}] \leqslant \mathbb{M}(W; x, x'),
$$

$$
\mathbb{M}(W; x, x') := W(x) + W(x')
$$

$$+\varepsilon^{-1}\left((1-\varepsilon)\sup_{(x,x')\in\Delta}\{R_mW(x)+R_mW(x')\}\right.$$

$$\left.+\sup_{(x,x')\in\Delta,k\in\{1,\dots,m-1\}}\{P^kW(x)+P^kW(x')\}\right).$$

Proof. To simplify the notations, it is assumed here that $m=1$. The first step of the proof consists in showing that, for all $(x,x')\in\mathscr{X}\times\mathscr{X}$ and all $n\geqslant 0$,

$$\mathbb{E}_{x,x',0}[(W(X_n)+W(X_n'))\mathbb{1}_{T_0\geqslant n}]\leqslant W(x)+W(x'), \tag{69}$$

where T_0 is the hitting-time on $\Delta\times\{0\}$ of Z. The proof is by induction on n. The induction assumption Eq. (69) is obvious for $n=0$. Assume now that Eq. (69) holds for some $n\geqslant 0$. We have

$$\begin{aligned}&\mathbb{E}_{x,x',0}[(W(X_{n+1})+W(X_{n+1}'))\ \mathbb{1}_{T_0\geqslant n+1}]\\&\quad=\mathbb{E}_{x,x',0}[(W(X_{n+1})+W(X_{n+1}'))\mathbb{1}_{T_0\geqslant n}\mathbb{1}_{\Delta^c\times 0}(Z_n)],\\&\quad=\mathbb{E}_{x,x',0}[\mathbb{E}_{Z_n}[(W(X_1)+W(X_1'))]\mathbb{1}_{T_0\geqslant n}\mathbb{1}_{\Delta^c\times 0}(Z_n)].\end{aligned} \tag{70}$$

For $(x,x')\notin\Delta$, we may write

$$\begin{aligned}&\mathbb{E}_{x,x',0}[(W(X_1)+W(X_1'))\mathbb{1}_{\{0\}}(d_1)]\\&\quad=PW(x)+PW(x')\\&\quad\leqslant W(x)+W(x')-(w(x)-b\mathbb{1}_C(x'))-(w(x')-b\mathbb{1}_C(x)).\end{aligned}$$

Under the stated assumption $w(x)-b\mathbb{1}_C(x')\geqslant 0$ and $w(x')-b\mathbb{1}_C(x)\geqslant 0$ for $(x,x')\notin\Delta$ showing that

$$\mathbb{E}_{x,x',0}[(W(X_1)+W(X_1'))\mathbb{1}_{\{0\}}(d_1)]\mathbb{1}_{\Delta^c}(x,x')\leqslant W(x)+W(x'). \tag{71}$$

Plugging this relation into Eq. (70) yields the desired result. Write

$$\begin{aligned}&\mathbb{E}_{x,x',0}[\{W(X_n)+W(X_n')\}\mathbb{1}_{T>n}]\\&\quad=\mathbb{E}_{x,x',0}[\{W(X_n)+W(X_n')\}\mathbb{1}_{0\leqslant n\leqslant T_0}]\\&\qquad+\sum_{j\geqslant 1}\mathbb{E}_{x,x',0}[\{W(X_n)+W(X_n')\}\mathbb{1}_{T_{j-1}<n\leqslant T_j}\mathbb{1}_{\{0\}}(d_{T_j})].\end{aligned} \tag{72}$$

Note that $T_j=T_{j-1}+1+T_0\circ\theta^{T_{j-1}+1}$ and $d_{T_j}=d_{T_{j-1}+1}$. For $j\geqslant 1$,

$$\mathbb{E}_{x,x',0}[\{W(X_n)+W(X_n')\}\mathbb{1}_{T_{j-1}<n\leqslant T_j}\mathbb{1}_{\{0\}}(d_{T_j})]=\int U(\omega,\theta^{1+T_{j-1}}(\omega))\,\mathrm{d}P_{x,x',0}(\omega),$$

where

$$\begin{aligned}U(\omega_1,\omega_2):=&\{W(X_{n-(T_{j-1}(\omega_1)+1)}(\omega_2))+W(X'_{n-(T_{j-1}(\omega_1)+1)}(\omega_2))\}\\&\times\mathbb{1}_{0\leqslant n-(T_{j-1}(\omega_1)+1)\leqslant T_0(\omega_2)}\mathbb{1}_{\{0\}}(d_{T_{j-1}+1}(\omega_1)).\end{aligned}$$

Applying a variant of the strong Markov property (see Lemma 6 in Appendix B) with $\tau = T_{j-1}+1$ yields

$$\mathbb{E}_{x,x',0}[\{W(X_n)+W(X'_n)\}\mathbb{1}_{T_{j-1}<n\leqslant T_j}\mathbb{1}_{\{0\}}(d_{T_j})]$$
$$\leqslant \int \mathbb{1}_{\{0\}}(d_{T_{j-1}+1}(\omega_1))\int\{W(X_{n-(T_{j-1}(\omega_1)+1)}(\omega_2))+W(X'_{n-(T_{j-1}(\omega_1)+1)}(\omega_2))\}$$
$$\times \mathbb{1}_{T_0(\omega_2)\geqslant n-(T_{j-1}(\omega_1)+1)\geqslant 0}\ \mathrm{d}P_{X_{1+T_j}(\omega_1),X'_{1+T_j}(\omega_1),0}(\omega_2)\,\mathrm{d}P_{x,x',0}(\omega_1).$$

Then, using Eq. (69) and the strong Markov property

$$\mathbb{E}_{x,x',0}[\{W(X_n)+W(X'_n)\}\mathbb{1}_{T_{j-1}<n\leqslant T_j}\mathbb{1}_{\{0\}}(d_{T_j})]$$
$$\leqslant \int \mathbb{1}_{\{0\}}(d_{T_{j-1}+1}(\omega_1))\{W(X_{1+T_j}(\omega_1))+W(X'_{1+T_j}(\omega_1))\}\,\mathrm{d}P_{x,x',0}(\omega_1)$$
$$\leqslant \mathbb{E}_{x,x',0}[\mathbb{1}_{\{0\}}(d_{T_{j-1}+1})\{W(X_{1+T_{j-1}})+W(X'_{1+T_{j-1}})\}]$$
$$\leqslant \mathbb{E}_{x,x',0}[\mathbb{1}_{\{0\}}(d_{T_{j-1}})\sup_{(x,x')\in\Delta}(1-\varepsilon)\int R(x,\mathrm{d}y)R(x',\mathrm{d}y')\{W(y)+W(y')\}]$$
$$\leqslant (1-\varepsilon)^{j-1}(1-\varepsilon)\sup_{(x,x')\in\Delta}\int R(x,\mathrm{d}y)R(x',\mathrm{d}y')\{W(y)+W(y')\}. \tag{73}$$

The proof is concluded by combining Eqs. (72) and (73). □

If P is ψ-irreducible and aperiodic, then under H1b$[C,D]$ and H3$[W,C,D]$, P has an unique invariant probability measure π and $\mathbb{M}(W;x,x')$ is $\delta_x\otimes\pi$-integrable if and only if $\pi(W)<\infty$.

Theorem 6. *Assume that P is ψ-irreducible and aperiodic. Let q be a positive integer and $C,D\in\mathscr{B}(\mathscr{X})$, $C\subseteq D$, and $W\geqslant 1$ be a Borel function. Assume* H1b$[C,D]$, H2$[q,C,D]$ *and* H3$[W,C,D]$. *Then, P has an unique invariant probability measure π, and for all $x\in\mathscr{X}$, $n\geqslant m$ and $0\leqslant\alpha\leqslant 1$,*

$$\|P^n(x,.)-\pi(.)\|_{W^{1-\alpha}}$$
$$\leqslant 2^\alpha\varepsilon^\alpha \min_{k\in\{0,\dots,(q-1)\}}\left\{U_{\varepsilon,m}(\mathbb{1}^{*k};n)^\alpha\int\delta_x(\mathrm{d}y)\pi(\mathrm{d}y')\ \{\mathbb{M}(W;y,y')\}^{1-\alpha}\right.$$
$$\left.\times\left\{\sum_{l=0}^{q}\hat{B}_{m;k,l}W_l(y,y')\right\}^\alpha\right\},$$

where (ε,m), the sequence $U_{\varepsilon,m}(\mathbb{1}^{*k};.)$, the vector W_k, the matrix $\hat{B}_m := (I-\hat{A}_m)^{-1}$, are given by Eqs. (62), (44), (54), (64), *respectively. Assume in addition that there* *exists* $0 < \beta \leqslant 1$ *such that* $\pi(V_q^\beta) < \infty$. *Then for all* $x \in \mathscr{X}$, $n \geqslant m$, $0 \leqslant \alpha \leqslant 1$,

$$\|P^n(x,.) - \pi(.)\|_{W^{1-\alpha}}$$
$$\leqslant 2^\alpha \varepsilon^\alpha \min_{k \in \{0,\dots,q\}} \left\{ U_{\varepsilon,m}((\mathbb{1}^{*k})^\beta; n)^\alpha \int \delta_x(\mathrm{d}y)\pi(\mathrm{d}y')\{\mathbb{M}(W; y, y')\}^{1-\alpha} \right.$$
$$\left. \times \left\{ \sum_{l=0}^{q} \hat{B}_{m;k,l}^{(\beta)} W_l^\beta(y, y') \right\}^\alpha \right\},$$

where the matrix $\hat{B}_m^{(\beta)} := (I - \hat{A}_m^{(\beta)})^{-1}$ *is defined in Eq.* (67).

Theorem 6 is a consequence of Eq. (68), Proposition 8 and of Theorems 4 and 5. For $\alpha = 1$, the bounds above coincides with the one given in Theorems 4 and 5 for total variation norm. Note that the bounds above are of interest only if $\pi(W) < \infty$.

Rates of convergence. Assume that there exist a drift function V unbounded off petite set, some $0 < \delta < 1/2$, a petite set C and some constants $b < \infty$, $c > 0$, such that $\sup_C V < \infty$ and $PV \leqslant V - cV^{1-\delta} + b\mathbb{1}_C$.

Set $q := \lfloor 1/\delta \rfloor$ and $\phi(x) := |x|^{1-\delta}$. Then, for any $0 < \eta \leqslant 1 - q\delta$ Eq. (31) shows that the function $V^{\eta+q\delta}$ satisfies the condition $S[\phi, q, C]$. As $\eta > 0$, the function $\phi_0 \circ V^{\eta+q\delta} \propto V^\eta$ is unbounded off petite set and (see Remark 15), there exist (a) a ν_m-small set $D_\eta \supseteq C$ such that H1b$[C, D_\eta]$ holds and (b) some functions $V_k \propto V^{\eta+k\delta}$ satisfying H2$[q, C, D_\eta]$. In addition, the JR drift condition implies that $\pi(V^{1-\delta}) < \infty$ that is $\pi(V_q^\beta) < \infty$ with $\beta := (1-\delta)/(\eta + q\delta)$. Finally, as $\delta < 1/2$, the condition H3$[V^{1-\delta}, C, D_\eta]$ also holds.

Hence, when P is ψ-irreducible and aperiodic, Theorem 6 applies and by setting $\alpha := (\kappa - 1)\delta/(1-\delta)$, we have, for all $x \in \mathbb{R}^l$, $\kappa \in [1, 1/\delta]$,

$$\sup_n (n+1)^{q(\kappa-1)\delta/(\eta+q\delta)} \|P^n(x,.) - \pi(.)\|_{V^{1-\kappa\delta}} < \infty.$$

By taking $\eta \downarrow 0$, we finally obtain for all $x \in \mathbb{R}^l$, $\kappa \in [1, 1/\delta]$ and $0 < \gamma < \kappa$,

$$\sup_n (n+1)^{\gamma-1} \|P^n(x,.) - \pi(.)\|_{V^{1-\kappa\delta}} < \infty. \tag{74}$$

In comparison, application of Proposition 5 under the same assumptions shows (see Eq. (32)) that $\lim_{n\to\infty} n^{\kappa-1}\|P^n(x,.) - \pi(.)\|_{V^{1-\kappa\delta}} = 0$, which is indeed, stronger than Eq. (74) from an asymptotic standpoint. Eq. (74) allows however to evaluate an explicit upper bound for the f-norm.

Acknowledgements

We are grateful to Soren Jarner, Gareth Roberts and Alexandre Veretennikov for useful discussions and for providing us preprint versions of on-going work on this topic. A special thanks to Pierre Priouret for his many comments and insights on earlier version of the manuscript and for the many hours spent to proofread the paper. We acknowledge partial support from the EU TMR network ERB-FMRX-CT96-0095 on "Computational and Statistics Methods for the Analysis of Spatial data."

Appendix A. Proofs of Section 2

A.1. Proof of Lemmas 1 and 2

A.1.1. A general result

Let P be the transition kernel of a Hastings Metropolis algorithm on $\mathbb{R}$ with proposal density (w.r.t. the Lebesgue measure) $k(y)$ and target density $p(y)$. Assume that p is such that

(i) there exist $0 \leqslant \eta < 1$, some measurable functions $\dot{l} : \mathbb{R} \to \mathbb{R}$, $\rho : \mathbb{R} \to [0, \infty)$, $\lim_{|x|\to\infty} |\rho(x)| = 0$, and some constants $C < \infty$, $M < \infty$ such that for all $|x| \geqslant M$, and all $y \in R(x) - x$,

$$\left| \frac{p(x+y)}{p(x)} - 1 - \dot{l}(x)y \right| \leqslant C|x|^{\eta-1}\rho(x)y^2,$$

$$|x|^{1-\eta}\ |\dot{l}(x)| \leqslant C.$$

Assume that the proposal density k is such that

(ii) there exist $0 < \alpha < 1$ and $\beta \in \mathbb{R}$ if $\eta > 0$, and $\beta \geqslant 2$ if $\eta = 0$ such that

$$\int \{|y|^{\beta+2\eta+2} \exp(\alpha|y|^{\eta}) + |y|^3\} k(y)\, \mathrm{d}y < \infty.$$

Lemma 5. *Assume* (i–ii). *Set* $V(x) := |x|^{\beta} \exp(\alpha|x|^{\eta})$ *if* $\eta \neq 0$ *and* $V(x) := |x|^{\beta}$ *otherwise. Then*

$$PV(x) - V(x) \leqslant \int \tilde{I}_1(x,y)k(y)\,\mathrm{d}y + \int \tilde{I}_2(x,y)k(y)\,\mathrm{d}y + |x|^{2\eta-2}V(x)\varepsilon(x)$$

where $\lim_{|x|\to\infty} \varepsilon(x) = 0$ *and*

$$\tilde{I}_1(x,y) := (\alpha\eta|x|^{\eta-1} + \beta|x|^{-1})V(x)\mathrm{sign}(x)y + \left(\frac{\beta(\beta-1)}{2}|x|^{-2} + \frac{\alpha^2\eta^2}{2}|x|^{2\eta-2} \right) V(x)y^2,$$

$$\tilde{I}_2(x,y) := \mathbb{1}_{R(x)-x}(y)(\alpha\eta|x|^{\eta-1} + \beta|x|^{-1})V(x)\,\mathrm{sign}(x)\ \dot{l}(x)y^2.$$

Proof. We will use the following two results, the proofs of which are omitted. (a) For all $y \in \mathbb{R}$,

$$V(x+y) - V(x) = (\alpha\eta|x|^{\eta-1} + \beta|x|^{-1})\mathrm{sign}(x)V(x)y + (\alpha^2\eta^2/2|x|^{2\eta-2} + \beta(\beta-1)/2|x|^{-2})V(x)y^2 + |x|^{2\eta-2}V(x)\varepsilon(x)$$

and $\lim_{|x|\to\infty}\varepsilon(x) = 0$. (b) There exist some finite constants K, M' such that for all $|x| \geqslant M'$, $y \in \mathbb{R}$,

$$V(x+y) - V(x) = (\beta|x|^{-1} + \alpha\eta|x|^{\eta-1})\,\mathrm{sign}(x)V(x)y + R(x,y), \tag{A.1}$$

where $|R(x,y)| \leqslant K|x|^{2\eta-2}V(x)|y|^{\beta+2\eta}\exp(\alpha|y|^{\eta})$. In order to prove the lemma, we write

$$\begin{aligned} PV(x) - V(x) &= \int (V(x+y) - V(x))k(y)\,\mathrm{d}y \\ &\quad + \int_{R(x)-x} (V(x+y) - V(x))\left(\frac{p(x+y)}{p(x)} - 1\right)k(y)\,\mathrm{d}y \\ &=: \int I_1(x,y)k(y)\,\mathrm{d}y + \int I_2(x,y)k(y)\,\mathrm{d}y \end{aligned}$$

and show that $\int \{I_i(x,y) - \tilde{I}_i(x,y)\}k(y)\,\mathrm{d}y = |x|^{2\eta-2}V(x)\varepsilon(x)$, $i = 1,2$, and $\lim_{|x|\to\infty}\varepsilon(x) = 0$; more precisely, we show that $\int H_i(x,y)k(y)\,\mathrm{d}y = \varepsilon_i(x)$, $\lim_{|x|\to\infty}\varepsilon_i(x) = 0$, by setting

$$H_i(x,y) := |x|^{-2\eta+2}V^{-1}(x)\{I_i(x,y) - \tilde{I}_i(x,y)\}.$$

We deduce easily that for all $y \in \mathbb{R}^l$, $H_1(x,y) = \varepsilon(x)$. In addition, from (A.1), it holds that for all $|x| \geqslant M'$,

$$H_1(x,y) = |x|^{-2\eta+2}V^{-1}(x)R(x,y) - (\beta(\beta-1)/2|x|^{-2\eta} + \alpha^2\eta^2/2)y^2.$$

Thus, there exists a finite constant K' such that for all $|x| \geqslant M'$, $|H_1(x,y)| \leqslant K'\{|y|^2 + |y|^{\beta+2\eta}\exp(\alpha|y|^{\eta})\}$ and this upper bound is $k(y)\,\mathrm{d}y$-integrable. It follows from the dominated convergence Theorem that $\int H_1(x,y)k(y)\,\mathrm{d}y = \varepsilon_1(x)$ and $\lim_{|x|\to\infty}\varepsilon_1(x) = 0$.

Similarly, there exist some constants $K'', M'' < \infty$ such that for all $|x| \geqslant M''$

$$|H_2(x,y)| \leqslant K''\{(1 + |x|^{-\eta})|y|^3\rho(x) + |x|^{\eta-1}|y|^{\beta+2\eta+2}e^{\alpha|y|^{\eta}}\}.$$

Thus $\lim_{|x|\to\infty} H_2(x,y) = 0$ for any given y and $\sup_{|x|\geqslant M''}|H_2(x,y)|$ is $k(y)\,\mathrm{d}y$-integrable, and the dominated convergence theorem again yields $\int H_2(x,y)k(y)\,\mathrm{d}y = \varepsilon_2(x)$ and $\lim_{|x|\to\infty}\varepsilon_2(x) = 0$. □

A.1.2. Proof of Lemma 1

Under (A1), the conditions (i–ii) of Lemma 5 are verified by setting $\eta = 0$ and $\dot{l}(x) = -sx^{-1}$. (A2ii) implies (ii) for all $2 \leqslant \beta \leqslant \zeta + 1$. Note finally that, since k is symmetric,

$$\int yk(y)\,\mathrm{d}y = 0 \quad \text{and} \quad \lim_{|x|\to\infty}\int_{R(x)-x} y^2k(y)\,\mathrm{d}y = \sigma_k^2/2,$$

and the result follows.

A.1.3. Proof of Lemma 2

Under (A3), the condition (i) of Lemma 5 holds with η equal to the shape parameter of the Weibull distribution and $\dot{l}(x) = -\eta x^{\eta-1}$. Finally, (A4ii) implies (ii) for all $\beta \in \mathbb{R}$ and the proof follows from the symmetry of k.

A.2. Proof of Lemma 3

For $|x| \leqslant M$, we have

$$\int P(x, \mathrm{d}y)|y|^s = \int |y + F(x)|^s \gamma(y)\,\mathrm{d}y \leqslant 2^{(s-1)\vee 0} \left\{ \sup_{|x| \leqslant M} |F(x)|^s + \Gamma(s) \right\} < \infty.$$

Assume now that $|x| \geqslant M$. We write

$$\begin{aligned} |F(x) + y|^s &\leqslant (|F(x)|^2 + |y|^2 + 2\langle F(x), y\rangle)^{s/2} \\ &\leqslant (|x|^2(1 - r|x|^{-d})^2 + |y|^2 + 2\langle F(x), y\rangle)^{s/2} =: N(x, y) \end{aligned}$$

and we determine an upper bound for $N(x, y)$. The following inequalities hold $\gamma(y)\,\mathrm{d}y$-a.e.

(i) If $s \leqslant 1$, the sub-additivity of the function $t \mapsto |t|^s$ implies $|F(x)+y|^s \leqslant |F(x)|^s + |y|^s$. If $s > 1$, we adapt here the proof of AngoNze (2000, Theorem 2). We write

$$\begin{aligned} \left(\int |F(x) + y|^s \Gamma(\mathrm{d}y)\right)^{1/s} &\leqslant |F(x)| + \Gamma(s)^{1/s} \leqslant \|x|(1 - r|x|^{-d})| + \Gamma(s)^{1/s} \\ &\leqslant |x|(1 - \{r - \Gamma(s)^{1/s}|x|^{d-1}\}|x|^{-d}). \end{aligned}$$

As $\Gamma(s) < \infty$, there exists a finite constant C such that

$$\begin{aligned} \int P(x, \mathrm{d}y)|y|^s &\leqslant |x|^s(1 - \{r - \Gamma(s)^{1/s}|x|^{d-1}\}|x|^{-d})^s \\ &\leqslant |x|^s(1 - sr|x|^{-d} + C|x|^{-d}\varepsilon(x)) \end{aligned}$$

and $\lim_{|x|\to\infty} \varepsilon(x) = 0$. Hence Lemma 3(i) follows.

(ii) If $d \leqslant s \leqslant 2$,

$$\begin{aligned} &|F(x) + y|^s \\ &\quad\leqslant \|x|(1 - r|x|^{-d})|^s \\ &\qquad + s/2(|y|^2 + 2\langle F(x), y\rangle) \int_0^1 (|x|^2(1-r|x|^{-d})^2 + u|y|^2 + 2u\langle F(x), y\rangle)^{s/2-1}\,\mathrm{d}u \\ &\quad\leqslant \|x|(1 - r|x|^{-d})|^s + s/2(|y|^2 + 2\langle F(x), y\rangle)\|x|(1 - r|x|^{-d})|^{s-2} \end{aligned}$$

and the result follows since $\int \langle F(x), y\rangle \Gamma(\mathrm{d}y) = 0$.

If $s > 2$, we write

$$\begin{aligned} |F(x) + y|^s \leqslant{}& \|x|(1 - r|x|^{-d})|^s + s/2(|y|^2 + 2\langle F(x), y\rangle)\|x|(1 - r|x|^{-d})|^{s-2} \\ &+ (s/2 - 1)(|y|^2 + 2\langle F(x), y\rangle)^2 \int_0^1 (1 - u)(|x|^2(1 - r|x|^{-d})^2 \\ &+ u|y|^2 + 2u\langle F(x), y\rangle)^{s/2-2}\,\mathrm{d}u. \end{aligned}$$

Observe that if $s > 4$, then there exists a finite constant C such that

$$\int_0^1 (|x|^2(1-r|x|^{-d})^2 + u|y|^2 + 2u\langle F(x), y\rangle)^{s/2-2}\, \mathrm{d}u \leqslant C|x|^{s-4}|y|^{s-4} \qquad \text{(A.2)}$$

and if $2 < s \leqslant 4$, $(|x|^2(1-r|x|^{-d})^2 + u|y|^2 + 2u\langle F(x), y\rangle) \geqslant (1-u)|x|^2(1-r|x|^{-d})^2$ so that Eq. (A.2) still holds. Hence,

$$|F(x)+y|^s \leqslant \|x|(1-r|x|^{-d})|^s + s/2(|y|^2 + 2\langle F(x), y\rangle)\|x|(1-r|x|^{-d})|^{s-2}$$
$$+ C|x|^{s-2}|y|^s.$$

And Lemma 3(ii) follows.

(iii) If $s=2$, $|F(x)+y|^2 = |F(x)|^2 + |y|^2 + 2\langle F(x), y\rangle$ and Eq. (35) follows since $\int \langle F(x), y\rangle \Gamma(\mathrm{d}y) = 0$.

If $2 < s \leqslant 6$, we write

$$|F(x)+y|^s \leqslant \|x|(1-r|x|^{-2})|^s + s/2(|y|^2 + 2\langle F(x), y\rangle)\|x|(1-r|x|^{-2})|^{s-2}$$
$$+ s/2(s/2-1)(|y|^2+2\langle F(x), y\rangle)^2 \int_0^1 (1-u)(|x|^2(1-r|x|^{-2})^2$$
$$+ u|y|^2 + 2u\langle F(x), y\rangle)^{s/2-2}\, \mathrm{d}u.$$

If $2 < s \leqslant 4$, since $|x|^2(1-r|x|^{-2})^2 + u|y|^2 + 2u\langle F(x), y\rangle \geqslant (1-u)|x|^2(1-r|x|^{-2})^2$, there exists a finite constant C such that

$$(|y|^2 + 2\langle F(x), y\rangle)^2 \int_0^1 (1-u)(|x|^2(1-r|x|^{-2})^2 + u|y|^2 + 2u\langle F(x), y\rangle)^{s/2-2}\, \mathrm{d}u$$
$$\leqslant (|y|^2 + 2\langle F(x), y\rangle)^2 \|x|(1-r|x|^{-2})|^{s-4} \int_0^1 (1-u)^{s/2-1}\, \mathrm{d}u$$
$$\leqslant 8/s|x|^{s-2}|y|^2 + C|x|^{s-3}|y|^4.$$

Hence,

$$|F(x)+y|^s \leqslant \|x|(1-r|x|^{-2})|^s + s/2|x|^{s-2}|y|^2 + s\langle F(x), y\rangle \|x|^{s-2}(1-r|x|^{-2})|^{s-2}$$
$$+ 4(s/2-1)|x|^{s-2}|y|^2 + C|x|^{s-3}|y|^4.$$

If $4 < s \leqslant 6$, since

$$(|x|^2(1-r|x|^{-2})^2 + u|y|^2 + 2u\langle F(x), y\rangle)^{s/2-2}$$
$$\leqslant \|x|(1-r|x|^{-2})|^{s-4} + \||y|^2 + 2\langle F(x), y\rangle|^{s/2-2}$$

then there exists a finite constant C such that

$$s/2(s/2-1)(|y|^2 + 2\langle F(x), y\rangle)^2 \int_0^1 (1-u)(|x|^2(1-r|x|^{-2})^2$$
$$+ u|y|^2 + 2u\langle F(x), y\rangle)^{s/2-2}\, \mathrm{d}u$$
$$\leqslant s/2(s/2-1)|x|^{s-2}|y|^2 + C|x|^{s/2}|y|^s.$$

Hence,

$$|F(x)+y|^s \leqslant \|x|(1-r|x|^{-2})|^s + s/2|x|^{s-2}|y|^2 + s\langle F(x),y\rangle \|x|(1-r|x|^{-2})|^{s-2}$$
$$+ s/2(s/2-1)|x|^{s-2}|y|^2 + C|x|^{s/2}|y|^s. \quad \text{(A.3)}$$

Finally, if $s>6$, we write

$$|F(x)+y|^s$$
$$\leqslant \|x|(1-r|x|^{-2})|^s + s/2(|y|^2+2\langle F(x),y\rangle)\|x|(1-r|x|^{-2})|^{s-2}$$
$$+ s/4(s/2-1)(|y|^2+2\langle F(x),y\rangle)^2\|x|(1-r|x|^{-2})|^{s-4}$$
$$+ s/4(s/2-1)(s/2-2)(|y|^2+2\langle F(x),y\rangle)^3\int_0^1 (1-u)^2(|x|^2(1-r|x|^{-2})^2$$
$$+ u|y|^2 + 2u\langle F(x),y\rangle)^{s/2-3}\,\mathrm{d}u.$$

Since $s/2-3>0$, there exists a finite constant C such that

$$s/4(s/2-1)(|y|^2+2\langle F(x),y\rangle)^2\|x|(1-r|x|^{-2})|^{s-4}$$
$$\leqslant s(s/2-1)|x|^{s-2}|y|^2 + C|x|^{s-3}|y|^4$$

and

$$(|y|^2+2\langle F(x),y\rangle)^3\int_0^1 (1-u)^2(|x|^2(1-r|x|^{-2})^2 + u|y|^2$$
$$+ 2u\langle F(x),y\rangle)^{s/2-3}\,\mathrm{d}u \leqslant C|x|^{s-3}|y|^s.$$

We then conclude as in Eq. (A.3).

Appendix B. Proof of Lemma 6

The following lemma is a strengthening of the strong Markov property (see Dynkin, 1963, Lemma 5.5.); we have included it for the sake of completeness.

Lemma 6. *Let* $\Phi = (\Omega, \mathscr{F}, \{\mathscr{F}_n\}, P_x, \{\Phi_n\})$ *be a Markov process. Let* τ *be a* $\mathscr{F}_n$*-stopping-time and* U *be a* $\mathscr{F}_\tau \otimes \mathscr{F}$*-measurable positive function. Then for any initial probability measure* λ,

$$\int_{\{\tau(\omega)<\infty\}} U(\omega, \theta^\tau\omega)\,\mathrm{d}P_\lambda(\omega) = \int_{\{\tau(\omega_1)<\infty\}}\int U(\omega_1,\omega_2)\,\mathrm{d}P_{\Phi_{\tau(\omega_1)}}(\omega_2)\,\mathrm{d}P_\lambda(\omega_1).$$

Proof. It is sufficient to prove the lemma for $U(\omega,\omega') := U_1(\omega)U_2(\omega')$ for U_1, U_2 two positive random variables and U_1 $\mathscr{F}_\tau$-measurable. In that case, we have

$$\int_{\tau<\infty} U_1(\omega)U_2(\theta^\tau\omega)P_\lambda(\mathrm{d}\omega) = \int_{\tau<\infty} U_1(\omega)\mathbb{E}_\lambda[U_2(\theta^\tau)|\mathscr{F}_\tau].$$

Using the strong Markov property,

$$\int_{\tau<\infty} U_1(\omega)U_2(\theta^{\tau}\omega)P_\lambda(\mathrm{d}\omega) = \int_{\tau<\infty} U_1(\omega)\mathbb{E}_{\Phi_\tau(\omega)}[U_2]P_\lambda(\mathrm{d}\omega)$$

$$= \int_{\tau<\infty} U_1(\omega)\int U_2(\omega')P_{\Phi_\tau}(\mathrm{d}\omega')P_\lambda(\mathrm{d}\omega)$$

which concludes the proof. □

References

AngoNze, P., 1994. Critères d'ergodicité de modèles markoviens. Estimation non-paramétrique sous des hypothèses de dépendance. Ph.D. Thesis, Univ. Paris IX- Dauphine, Paris.

AngoNze, P., 2000. Geometric and subgeometric rates for Markovian processes: a robust approach. Preprint-Université de Lille III.

Diaconis, P., Freedman, D., 1999. Iterated random functions. SIAM Rev. 41 (1), 45–76.

Doukhan, P., Ghindès, M., 1980. Etude du processus $x_{n+1} = f(x_n) + \varepsilon_n$. C.R. Acad. Sci. Paris Sér. A-B 291 (1), 61–63.

Dynkin, E.B., 1963. Théorie des processus markoviens. Dunod, Paris.

Fort, G., 2001. Contrôle explicite d'ergodicité de chaîne de Markov: applications à l'analyse de convergence de l'algorithme Monte-Carlo EM. Ph.D. Thesis, Université Pierre et Marie Curie, Paris.

Fort, G., Moulines, E., 2000. V-subgeometric ergodicity for a Hastings-Metropolis algorithm. Statist. Probab. Lett. 49 (4), 401–410.

Jarner, S.F., Hansen, E., 2000. Geometric ergodicity of Metropolis algorithms. Stochastic Processes Appl. 85 (2), 341–361.

Jarner, S., Roberts, G.O., 2002. Polynomial convergence rates of Markov Chains. Ann. Appl. Probab. 12 (1), 224–247.

Kalashnikov, V.V., 1994. Regeneration and general Markov Chains. J. Appl. Math. Stochastic Anal. 7 (3), 357–371.

Klüppelberg, C., 1988. Subexponential distributions and integrated tails. J. Appl. Probab. 25 (1), 132–141.

Mengersen, K.L., Tweedie, R.L., 1996. Rates of convergence of the Hastings and Metropolis algorithms. Ann. Statist. 24 (1), 101–121.

Meyn, S.P., Tweedie, R.L., 1993. Markov Chains and Stochastic Stability. Springer, London Ltd., London.

Mokkadem, A., 1987. Sur un modèle autorégressif non linéaire: ergodicité et ergodicité géométrique. J. Time Ser. Anal. 8 (2), 195–204.

Robert, C.P., Casella, G., 1999. Monte Carlo statistical methods. Springer, New York.

Roberts, G.O., Tweedie, R.L., 1996. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. Biometrika 83 (1), 95–110.

Roberts, G.O., Tweedie, R.L., 1999. Bounds on regeneration times and convergence rates for Markov Chains. Stochastic Process. Appl. 80 (2), 211–229.

Roberts, G.O., Tweedie, R.L., 2001. Corrigendum to: "Bounds on regeneration times and convergence rates for Markov chains". Stochastic Process. Appl. 91 (2), 337–338.

Rosenthal, J.S., 1995a. Minorization conditions and convergence rates for Markov chain Monte Carlo. J. Amer. Statist. Assoc. 90 (430), 558–566.

Rosenthal, J.S., 1995b. Correction: "Minorization conditions and convergence rates for Markov chain Monte Carlo". J. Amer. Statist. Assoc. 90 (431), 1136.

Rosenthal, J.S., 2001. A review of asymptotic convergence for general state space Markov Chains. Far East J. Math. Sci. 5, 37–50.

Stramer, O., Tweedie, R.L., 1999. Langevin-Type Models I: diffusions with given stationary distributions, and their discretizations. Methodol. Comput. Appl. Probab. 1 (3), 283–306.

Tanikawa, A., 1996. The rate of convergence of a homogeneous Markov chain arising from two-queue networks. Comput. Math. Appl. 31 (2), 103–109.

Tanikawa, A., 1999. Geometric ergodicity of nonlinear first order autoregressive models. Comm. Statist. Stochastic Models 15 (2), 227–245.

Tuominen, P., Tweedie, R.L., 1994. Subgeometric rates of convergence of f-ergodic Markov Chains. Adv. Appl. Probab. 26 (3), 775–798.

Tweedie, R.L., 1983. Criteria for rates of convergence of Markov chains, with application to queueing and storage theory. In: Kingman, J.F.C., Reuter, G.E.H. (Eds.), Probability, Statistics and Analysis. Cambridge University Press, Cambridge, pp. 260–276.

Veretennikov, A.Y., 2000. On polynomial mixing and the rate of convergence for stochastic difference and differential equations. Theory Probab. Appl. 44 (2), 361–374.