Quantitative approximation of the invariant distribution of a Markov chain. A new approach

Loïc Hervé, James Ledoux

Quantitative approximation of the invariant distribution of a Markov chain. A new approach

Loïc HERVÉ, and James LEDOUX *

version: Tuesday 31$^{\text{st}}$ January, 2023 – 19:19

Abstract

In this paper, we deal with a Markov chain on a measurable state space $(\mathbb{X}, \mathcal{X})$ which has a transition kernel P admitting some small-set $S \in \mathcal{X}$, that is such that $P(x, A) \geq \nu(1_A)1_S(x)$ for any $x \in \mathbb{X}$, $A \in \mathcal{X}$ and for some positive measure ν. Under this condition, we propose a constructive characterisation of the existence of an P-invariant probability measure π on $(\mathbb{X}, \mathcal{X})$ such that $\pi(1_S) > 0$. When such an π exists, it is approximated in weighted or standard total variation norms by a finite linear combination of non-negative measures only depending on ν, P and S. Next, using standard drift-type conditions, we provide geometric/subgeometric convergence bounds of the approximation. Theses bounds are fully explicit and are as simple as possible. The rates of convergence are accurate, and they are optimal in the atomic case. Note that the rate of convergence for approximating the iterates of P by the finite-rank submarkovian kernels introduced in [HL20b] is also discussed. This is a new approach for approximating π in the sense that it is not based on the convergence of the iterates of P to π. Thus we need no aperiodicity condition. Moreover, the proofs are direct. They use neither the split chain in the non-atomic case, nor the renewal theory, nor the coupling method, nor the spectral theory. In some sense, this approach for Markov chains with a small-set is self-contained.

AMS subject classification : 60J05

Keywords : Drift conditions, Finite-rank approximating, Invariant probability measure, Rate of convergence, Small set

1 Introduction

Throughout the paper P is a Markov kernel on a measurable space $(\mathbb{X}, \mathcal{X})$. Let $(X_n)_{n\geq 0}$ be a Markov chain with state space $\mathbb{X}$ and transition kernel P. If $(X_n)_{n\geq 0}$ admits an invariant distribution denoted by π, the following issue is of interest for any $A \in \mathcal{X}$:

(Q) How to approximate the value of $\pi(1_A)$ and to control the error?

The standard way is to use, when n is large enough, $\mathbb{P}(X_n \in A)$ to approximate $\pi(1_A)$. This approach is supported by all the classical results related to the convergence in distribution of $(X_n)_{n\geq 0}$ to π, or in other words by all the results of convergence of the iterates P^n to

*Univ Rennes, INSA Rennes, CNRS, IRMAR-UMR 6625, F-35000, France. Loic.Herve@insa-rennes.fr, James.Ledoux@insa-rennes.fr

the rank-one operator $\pi(\cdot)1_{\mathbb{X}}$. Another classical issue is: How to approximate the value of $\mathbb{P}(X_n \in A)$ and to control the error? Of course $\pi(1_A)$ can be used to approximate $\mathbb{P}(X_n \in A)$ when n is large enough, but this approximation is effective only when π is known. In practice π is often unknown, in which case (Q) becomes a central issue. Concerning (Q), observe that π may be approximated by something other than the iterates of P, provided that the approximation procedure is effective and that the error is well controlled.

The main objective of this work is to propose a new approach to address (Q), which is not directly based on the convergence of P^n to π. Specifically, when P has a small-set S and has an invariant probability measure π such that $\pi(1_S) > 0$, we present a general and effective procedure for approximating π. The central point here is that all the convergence bounds are fully explicit and are as simple as possible. The rates of convergence are accurate, and they are optimal in the atomic case.

Let $\mathcal{M}^+$ (resp. $\mathcal{M}^+_*$) denote the set of finite non-negative (resp. positive) measures on $(\mathbb{X}, \mathcal{X})$. For any $\mu \in \mathcal{M}^+$ and any μ-integrable function $f : \mathbb{X} \to \mathbb{C}$, $\mu(f)$ denotes the integral $\int f d\mu$. Throughout the paper, the existence of a small-set S for P is assumed, that is

$$\exists S \in \mathcal{X},\ \exists \nu \in \mathcal{M}^+_*, \quad \forall x \in \mathbb{X},\ \forall A \in \mathcal{X}, \quad P(x, A) \geq \nu(1_A)\, 1_S(x). \tag{$\mathbf{S}$}$$

Under Assumption ($\mathbf{S}$), we can use the following sequence $(\beta_k)_{k\geq 1} \in (\mathcal{M}^+)^{\mathbb{N}}$ introduced in [HL20b] and defined by

$$\forall n \geq 1, \quad \beta_n = \nu \circ (P - T)^{n-1} \quad \text{with} \quad T\cdot := \nu(\cdot)1_S. \tag{1}$$

An equivalent definition of β_n is given in (11) in Section 2. Note that no spectral theory is used here in contrast to [HL20b]. Under Assumption ($\mathbf{S}$), the following results are obtained.

- In Section 2 (Theorem 2.1), we prove that there exists an $P-$invariant probability measure π on $(\mathbb{X}, \mathcal{X})$ such that $\pi(1_S) > 0$ if, and only if,

$$\sum_{k=1}^{+\infty} \beta_k(1_{\mathbb{X}}) < \infty. \tag{2}$$

  Actually, under this condition, set $\mu := \sum_{k=1}^{+\infty} \beta_k \in \mathcal{M}^+_*$. Then $\mu(1_S) = 1$ and

$$\pi := \frac{\mu}{\mu(1_{\mathbb{X}})} \tag{3}$$

  is an $P-$invariant probability measure on $(\mathbb{X}, \mathcal{X})$ such that $\pi(1_S) = 1/\mu(1_{\mathbb{X}}) > 0$.

In the next items Condition (2) is assumed to hold, and for every $n \geq 1$ we consider $\mu_n \in \mathcal{M}^+_*$ and the probability measure $\widetilde{\mu}_n$ on $(\mathbb{X}, \mathcal{X})$ defined by:

$$\mu_n := \sum_{k=1}^{n} \beta_k \quad \text{and} \quad \widetilde{\mu}_n := \frac{1}{\mu_n(1_{\mathbb{X}})}\, \mu_n. \tag{4}$$

The comments below are mainly about the error bounds obtained for the approximation of π by $\mu(1_{\mathbb{X}})^{-1}\mu_n$ or $\widetilde{\mu}_n$. For the sake of simplicity the following discussion is mainly focused on the results for the standard total variation norm $\|\cdot\|_{TV}$ (see (21)).

- In Section 3 (Theorem 3.1), we prove that the invariant distribution π given by Formula (3) can be approximated in total variation norm by either $(\mu_n/\mu(1_{\mathbb{X}}))_n$ or $(\widetilde{\mu}_n)_n$ with the following error estimates

$$\|\pi - \mu(1_{\mathbb{X}})^{-1}\mu_n\|_{TV} = \mu(1_{\mathbb{X}})^{-1}\,\varepsilon_n \leq \varepsilon_n \quad \text{and} \quad \|\pi - \widetilde{\mu}_n\|_{TV} \leq 2\,\mu(1_{\mathbb{X}})^{-1}\,\varepsilon_n \leq 2\,\varepsilon_n \tag{5}$$
$$\text{with} \quad \varepsilon_n := \sum_{k=n+1}^{+\infty} \beta_k(1_{\mathbb{X}}).$$

  Note that $\lim_n \varepsilon_n = 0$ from (2).

- In Section 4, geometric drift conditions are used to obtain geometric rates of convergence for the above sequence $(\varepsilon_n)_{n\geq 0}$. First, under the standard drift condition $PV \leq \delta\,V + L\,1_S$ for some constants $\delta \in (0,1)$, $L > 0$ and some measurable function $V : \mathbb{X} \to [1, +\infty)$ (e.g. see [MT93, DMPS18]), we establish that $\theta_1 := \limsup_n [\beta_n(1_{\mathbb{X}})]^{1/n} < 1$ and that: $\forall \theta \in (\theta_1, 1),\ \varepsilon_n = \mathrm{O}(\theta^n)$ (see Theorem 4.1 and Remark 4.1). Second, in order to obtain computable rates of convergence for $(\varepsilon_n)_{n\geq 1}$, the following drift condition is introduced

$$\exists \delta \in (0,1), \quad PV \leq \delta\,V + \nu(V)\,1_S. \tag{6}$$

  Under Condition (6), we prove that $\varepsilon_n \leq (1-\delta)^{-1}\nu(V)\delta^n$ (Theorem 4.2). Condition (6) holds if S is an atom (Corollary 4.1), but may fail in the non atomic case. In this last case, assuming that $\sup_S PV < \infty$ and $PV \leq \delta\,V$ on $S^c := \mathbb{X} \setminus S$, it can be shown that (6) holds with δ^{α_0} and V^{α_0} in place of δ and V for some easily computable $\alpha_0 \in (0,1)$, so that the following bound holds (Corollary 4.2)

$$\forall n \geq 1, \quad \varepsilon_n \leq \frac{\nu(V^{\alpha_0})}{1-\delta^{\alpha_0}}\,\delta^{\alpha_0 n}. \tag{7}$$

  Note that the assumptions of Corollary 4.2 are general. Indeed, if $\sup_S PV < \infty$, then the drift condition $PV \leq \delta\,V$ on $S^c := \mathbb{X}\setminus S$ is equivalent to the standard one $PV \leq \delta\,V + L\,1_S$. Finally some properties involved in the proof of Theorem 4.2 (resp. of Corollary 4.2) are used in Theorem 4.3 to derive a rate of convergence to 0 for $P^n - T_n$ in V-weighted operator norm, where T_n is the submarkovian finite-rank kernel defined in (13a). This rate of convergence enables us to specify the error bound obtained in [HL20b, HL20a] for the $V-$geometrical ergodicity of P. Using the triangle inequality, any such error bounds can be combined with (5)-(7) to approximate $P^n(x, A)$ for any $A \in \mathcal{X}$ (Theorem 4.4).

- In Section 5 the following subgeometric drift-type conditions using $P - T = P - \nu(\cdot)1_S$ are introduced to study the rate of convergence of $(\varepsilon_n)_{n\geq 1}$: for $m \geq 1$ there exist $m+1$ measurable functions $V_i : \mathbb{X} \to [1, +\infty),\ i = 0, \ldots, m$, such that

$$\forall i \in \{0, \ldots, m-1\}, \quad (P-T)V_i \leq V_i - V_{i+1}. \tag{8}$$

  Under Condition (8), $(\varepsilon_n)_{n\geq 1}$ is proved to satisfy $\lim_n n^{m-1}\varepsilon_n = 0$ (Theorem 5.1). The sequence $(\beta_k(V_m))_k$ is investigated in Theorem 5.2 to obtain computable rates of convergence for $(\varepsilon_n)_{n\geq 1}$. In particular the following property is stated in Corollary 5.1: if $m \geq 2$, then

$$\forall n \geq 1, \quad \varepsilon_n \leq \frac{C_m\,\nu(V_0)}{(m-1)\,n^{m-1}} \quad \text{with} \quad C_m := 2^{\frac{m(m+1)}{2}-1}. \tag{9}$$

Next it is shown in Corollary 5.2 that the subgeometric drift conditions (8) are fulfilled under the more explicit following ones:

$$\forall i \in \{0, \ldots, m-1\}, \quad \begin{cases} V_{i+1} \leq V_i \\ PV_i \leq \big(V_i - V_{i+1}\big) + \nu(V_i)\, 1_S. \end{cases} \tag{10}$$

Assumption (10) corresponds to standard subgeometric drift conditions (e.g. see [JR02, DMPS18]), but using the particular constant $\nu(V_i)$ in the second condition. It turns out that (10) is our target subgeometric drift condition, as is Condition (6) in the geometric case. Indeed appropriate procedures to return to this condition when starting with more realistic subgeometric drift conditions can also be provided. Again the atomic case is simpler since (10) can be applied directly. Indeed, if S is an atom, the first condition in (10) implies the second one. Then, using an iterative procedure based on [JR02, Lem. 3.5], we prove that, if $\sup_S PV < \infty$ and if P satisfies the condition $PV \leq V - c_1 V^{\alpha}$ on S^c for some constants $\alpha \in [0,1)$, $c_1 > 0$, and some measurable function $V : \mathbb{X} \to [1, +\infty)$, then the bound (9) holds with $m := \lfloor (1-\alpha)^{-1} \rfloor$, where $\lfloor \cdot \rfloor$ denotes the integer part function on $\mathbb{R}$ (see Corollary 5.4). In the non atomic case the second condition in (10) may hold with a constant $b_i > \nu(V_i)$. However the iterative procedure of the atomic case can be adapted, provided that $PV \leq V - c_1 V^{\alpha}$ is replaced by $P\widehat{V} \leq \widehat{V} - \widehat{c}_1 \widehat{V}^{\widehat{\alpha}}$ with $\widehat{V} = V^{\eta_0}$ for some explicit $\eta_0 \in (0,1]$. Then, if $\eta_0 \geq 1 - \alpha$ and if V, PV are bounded on S, the bound (9) holds with $m := \lfloor \eta_0 (1-\alpha)^{-1} \rfloor$ (see Corollary 5.5). Hence the conditions of Corollary 5.5 are general, but the initial subgeometric drift condition needs to be adjusted with suitable powers. Finally in Theorem 5.3 the rate of convergence for $P^n - T_n$ with T_n given in (13a) is specified under the subgeometric drift conditions (8).

- In Section 6, we illustrate our results on standard examples of Markov chains. The numerical findings support the idea that the quality of the bounds in (5) is good, and in particular that the use of $\alpha_0 \in (0,1)$ in the bound (7) is relevant even when α_0 is small.
- The above error bounds actually hold in $W-$weighted total variation norm (see (21)) for suitable functions $W \geq 1$, namely for any $W \geq 1$ such that $\mu(W) < \infty$ in Section 3, for $W = V^{\alpha_0}$ in Section 4, and finally for $W = V_j$ in Section 5.

We recall that this work is not directly based on the convergence in distribution of the Markov chain $(X_n)_{n \geq 0}$. In particular no aperiodicity condition is introduced. We use neither renewal theory, nor coupling method, nor spectral theory. Actually our main statements are concerned with the rate of convergence in (5), in which the positive measure μ_n and the probability measure $\widetilde{\mu}_n$ can be written as a linear combination of the non-negative measures $\nu, \nu \circ P, \ldots, \nu \circ P^{n-1}$ with explicit coefficients only depending on ν, P and S (see (4) and (11)). Therefore, precise qualitative or quantitative comparisons with the classical works recalled below are difficult to address.

The basic fact is that our assumptions are quite close to usual ones. Indeed, the central assumption (**S**) is the existence of a small-set S. But we do not introduce the strong aperiodicity condition $\nu(1_S) > 0$ in order to get a minorization condition as in [MT09, p 98] or in [DMPS18, Chap. 11]. Here no use of the split chain is needed for proving our results in the non-atomic case. Next, Condition (2) is proved to be equivalent to $\sum_{k=1}^{+\infty} \beta_k(1_S) = 1$ and $\sum_{k=1}^{+\infty} k\, \beta_k(1_S) < \infty$ in Theorem 2.1. When S is an atom, this last condition is nothing else but the usual condition of finite expectation of the first return time in S, see (20).

Formula (3), which has been obtained in the V-geometric ergodicity context [HL20b], extends a well-known formula when P satisfies the Doeblin condition ($\mathbb{X}$ is a small-set), see [LC14], or when P is irreducible and recurrent positive according to [Num84, p 74]. Next, the use of geometric or subgeometric drift conditions is standard for investigating the rate of convergence of the iterates P^n of the transition kernel P to π. The error term is usually computed in some weighted operator norm. Under irreducibility and aperiodicity conditions, if P satisfies Assumption (**S**) and the geometric drift condition $PV \le \delta V + L 1_S$ for some constants $\delta \in (0,1)$, $L > 0$, and some measurable function $V : \mathbb{X} \to [1, +\infty)$, then P is V-geometrically ergodic, e.g. see [MT93, RR04, Bax05, MT09, DMPS18] (see also [Hen06, Hen07, HM11, Del17, HL20b] for alternative approaches). Moreover the previous drift condition has been proved to be useful to derive computable rates of convergence in the V-geometric ergodicity property, e.g. see [MT94, LT96, RT99, RT00, Ros02, Bax05]. However recall that deriving effective and accurate bounds in the V-geometric ergodicity property is a difficult issue. Similarly non-geometric (for instance polynomial) rates of convergence can be derived under subgeometric drift conditions, see [DMPS18, and the references therein] and [Del17] for an operator-type approach. Various subgeometric drift conditions can be found in [DFMS04, DMPS18] and quantitative bounds of polynomial rates for the convergence of P^n to π are obtained in [AF05, AFV15].

The estimates in (5) do not give direct information on the convergence of the iterates of P to π, but they do provide an approximation of $\pi(1_A)$ for all $A \in \mathcal{X}$. The error bounds obtained in both geometrical case (Section 4) and subgeometrical case (Section 5) are simple and explicit. Note that μ_n, thus $\widetilde{\mu}_n$, are computable whenever the iterates of P are available. The proofs in this paper can be thought of as self-contained. It appears that the initial idea of approximating π by μ_n or $\widetilde{\mu}_n$ rather than with the iterates of P simplifies the error computations. Moreover, introducing from the minorization measure ν the ideal drift conditions (6) or (10), and then adjusting with a power to get back to them when starting from general drift conditions, seems to be a new and efficient idea to find explicit error bounds.

Note that this new approach might also be used as an alternative theoretical tool in problems usually involving the iterates of Markov kernels. For example, if P_θ is a perturbed Markov kernel of P_{θ_0}, then the quantities $\pi_\theta - \widetilde{\mu}_{n,\theta}$ defined from P_θ can be used as intermediate error terms to control $\pi_\theta - \pi_{\theta_0}$, where π_θ (resp. π_{θ_0}) is the invariant probability measure for P_θ (resp. P_{θ_0}). Note that only the error bounds for $\pi_\theta - \widetilde{\mu}_{n,\theta}$ are useful in this perturbation issue: neither $\widetilde{\mu}_{n,\theta}$ nor $\widetilde{\mu}_{n,\theta_0}$ need to be computed. The resulting error bounds for $\pi_\theta - \pi_{\theta_0}$ will be more accurate than those obtained with the intermediate term $\pi_\theta - P_\theta^n$, simply because the error bounds for $\pi_\theta - \widetilde{\mu}_{n,\theta}$ are better.

2 Existence of π under Assumption (S)

We denote by $\mathcal{B}$ the space of real-valued bounded measurable functions on $(\mathbb{X}, \mathcal{X})$, equipped with its usual supremum norm: $\forall f \in \mathcal{B},\ \|f\| := \sup_{x\in\mathbb{X}} |f(x)|$. If Q_1 and Q_2 are bounded linear operators on $\mathcal{B}$, we write $Q_1 \le Q_2$ when the following property holds: $\forall f \in \mathcal{B},\ f \ge 0,\ Q_1 f \le Q_2 f$.

Let P be a Markov kernel satisfying Condition (**S**). Note that P is a bounded linear operator on $\mathcal{B}$ since P is a Markov kernel, and that $f \mapsto \nu(f)$ is a continuous linear form on

$\mathcal{B}$, with $\nu \in \mathcal{M}_*^+$ given in (**S**). Let us introduce the following continuous linear forms on $\mathcal{B}$

$$\forall f \in \mathcal{B}: \ \beta_1(f) := \nu(f) \ \text{ and } \ \forall n \geq 2, \ \beta_n(f) := \nu\big(P^{n-1}f\big) - \sum_{k=1}^{n-1} \nu\big(P^{n-k-1}1_S\big)\, \beta_k(f). \tag{11}$$

Moreover let T be the rank-one operator on $\mathcal{B}$ defined by :

$$\forall f \in \mathcal{B}, \quad Tf := \nu(f)\, 1_S = \beta_1(f)\, 1_S. \tag{12}$$

It follows from the positivity of ν and from (**S**) that $0 \leq T \leq P$. Some basic facts proved in [HL20b] are collected in the following proposition. A proof is postponed in Annex A for the convenience of the reader.

Proposition 2.1 *Assume that P satisfies Condition (**S**). Set $T_0 := 0$ and $T_n := P^n - (P-T)^n$ for any $n \geq 1$. Then*

$$\forall n \geq 1, \quad 0 \leq T_n \leq P^n \quad \text{and} \quad T_n = \sum_{k=1}^{n} \beta_k(\cdot) P^{n-k} 1_S \tag{13a}$$

$$T_n - T_{n-1}P = (P^{n-1} - T_{n-1})T. \tag{13b}$$

Moreover, for every $n \geq 1$, $\beta_n \in \mathcal{M}^+$, that is: there exists a non-negative measure on $(\mathbb{X}, \mathcal{X})$ (still denoted by β_n) such that $\int_{\mathbb{X}} d\beta_n < \infty$ and, such that, for every $f \in \mathcal{B}$, we have $\beta_n(f) = \int_{\mathbb{X}} f\, d\beta_n$. Finally we have

$$\forall n \geq 1, \quad \beta_n = \nu \circ (P^{n-1} - T_{n-1}) = \nu \circ (P-T)^{n-1} \quad \text{and} \quad \beta_{n+1} = \beta_n \circ (P-T) \tag{14}$$

with the convention that P^0 and $(P-T)^0$ stand for the identity map on $\mathcal{B}$.

The definition (1) of β_n given in Introduction is equivalent to the above definition (11) from (14). Now we can prove the main theorem of this section.

Theorem 2.1 *Assume that P satisfies Condition (**S**). Then the four following assertions are equivalent.*

(i) *There exists a $P-$invariant probability measure π on $(\mathbb{X}, \mathcal{X})$ such that $\pi(1_S) > 0$.*

(ii) $\displaystyle\sum_{k=1}^{+\infty} \beta_k(1_{\mathbb{X}}) < \infty.$

(iii) $\displaystyle\sum_{k=1}^{+\infty} \beta_k(1_S) = 1$ *and* $\displaystyle\sum_{k=1}^{+\infty} k\, \beta_k(1_S) < \infty.$

(iv) $\displaystyle\lim_{k \to +\infty} \beta_k(1_{\mathbb{X}}) = 0$ *and* $\displaystyle\sum_{k=1}^{+\infty} k\, \beta_k(1_S) < \infty.$

Moreover, under any of these four conditions, we have

$$\sum_{k=1}^{+\infty} k\, \beta_k(1_S) = \frac{1}{\nu(1_{\mathbb{X}})} \sum_{k=1}^{+\infty} \beta_k(1_{\mathbb{X}}) \tag{15}$$

and

$$\pi := \frac{1}{\sum_{k=1}^{+\infty} \beta_k(1_{\mathbb{X}})} \sum_{k=1}^{+\infty} \beta_k \tag{16}$$

is an $P-$invariant probability measure on $(\mathbb{X}, \mathcal{X})$ such that

$$\pi(1_S) = \frac{1}{\sum_{k=1}^{+\infty} \beta_k(1_{\mathbb{X}})} > 0. \tag{17}$$

Proof. Assume that Assertion (i) holds. We deduce from (13a) that

$$0 \le \pi\big((P^n - T_n)1_{\mathbb{X}}\big) = 1 - \pi(T_n 1_{\mathbb{X}}) = 1 - \pi(1_S) \sum_{k=1}^{n} \beta_k(1_{\mathbb{X}}),$$

from which it follows that $\sum_{k=1}^{+\infty} \beta_k(1_{\mathbb{X}}) \le 1/\pi(1_S) < \infty$ since $\pi(1_S) > 0$ by hypothesis. This gives Property (ii). Conversely assume that Assertion (ii) holds. Then

$$\mu := \sum_{k=1}^{+\infty} \beta_k \ \in \mathcal{M}_*^+$$

since $\mu(1_{\mathbb{X}}) \ge \beta_1(1_{\mathbb{X}}) = \nu(1_{\mathbb{X}}) > 0$. Also note that, for every $f \in \mathcal{B}$, the series $\sum_{k=1}^{+\infty} \beta_k(f)$ absolutely converges in $\mathbb{C}$ since $|\beta_k(f)| \le \|f\|\, \beta_k(1_{\mathbb{X}})$. We obtain that, for every $f \in \mathcal{B}$,

$$\begin{aligned}
\mu(Pf) &= \sum_{k=1}^{+\infty} \nu\big(P^k f - T_{k-1}Pf\big) \qquad \text{(from (14))} \\
&= \sum_{k=1}^{+\infty} \nu\big(P^k f - T_k f\big) + \sum_{k=1}^{+\infty} \nu\big(P^{k-1}Tf - T_{k-1}Tf\big) \qquad \text{(from (13b))} \\
&= \mu(f) - \nu(f) + \mu(Tf) \qquad \text{(from (14) and } \beta_1(f) = \nu(f)) \\
&= \mu(f) - \nu(f) + \mu(1_S)\, \nu(f) \qquad \text{(from the definition of } T) \\
&= \mu(f) - \nu(f)\big(1 - \mu(1_S)\big).
\end{aligned}$$

Note that the second equality holds since both series in the right-hand side are equal to $\sum_{k=1}^{+\infty} \beta_{k+1}(f)$ and $\sum_{k=1}^{+\infty} \beta_k(Tf)f$ respectively, which are absolutely convergent. With $f := 1_{\mathbb{X}}$ the previous equality gives $\mu(1_S) = 1$ since $P1_{\mathbb{X}} = 1_{\mathbb{X}}$ and $\nu(1_{\mathbb{X}}) > 0$. Thus μ is an $P-$invariant non-negative measure such that $\mu(1_{\mathbb{X}}) > 0$ and $\mu(1_S) = 1$, so that $\pi := \mu/\mu(1_{\mathbb{X}})$ is an $P-$invariant probability measure on $(\mathbb{X}, \mathcal{X})$ such that $\pi(1_S) = 1/\mu(1_{\mathbb{X}}) > 0$. We have proved that Assertions (i) and (ii) are equivalent, and that Equality (17) is valid under any of these two assertions.

Next we obtain from (14) and the definition of T

$$\forall k \ge 1, \quad \beta_{k+1}(1_{\mathbb{X}}) = \beta_k \circ (P - T)(1_{\mathbb{X}}) = \beta_k(1_{\mathbb{X}}) - \nu(1_{\mathbb{X}})\, \beta_k(1_S). \tag{18}$$

Set $b_k := \beta_k(1_{\mathbb{X}})$ and $c_k := \nu(1_{\mathbb{X}})\, \beta_k(1_S)$ for any $k \ge 1$. Note that $b_k, c_k \ge 0$ and that $(b_k)_{k\ge1}$ is decreasing. We have $c_k = b_k - b_{k+1}$ from (18) so that

$$\sum_{k=1}^{n} k\, c_k = \sum_{k=1}^{n} \big(b_k \ - b_{n+1}\big) = \sum_{k=1}^{+\infty} \phi^{(n)}(k) \qquad \text{with} \quad \phi^{(n)}(k) := \big(b_k \ - b_{n+1}\big)\, 1_{[1,n]}(k).$$

Note that $0 \leq \phi^{(n)} \leq \phi^{(n+1)}$. Moreover, if $\lim_n b_n = 0$, then we have $\forall k \geq 1,\ \lim_n \phi^{(n)}(k) = b_k$ and the following equalities hold in $[0, +\infty]$

$$\sum_{k=1}^{+\infty} k\, c_k = \lim_{n\to+\infty} \sum_{k=1}^{n} k\, c_k = \lim_{n\to+\infty} \sum_{k=1}^{+\infty} \phi^{(n)}(k) = \sum_{k=1}^{+\infty} b_k$$

from the monotone convergence theorem with respect to the counting measure. This proves the equivalence of Assertions (ii) and (iv), and Equality (15). Equivalence of Assertions (iii) and (iv) follows from

$$\forall n \geq 1, \quad \sum_{k=1}^{n-1} c_k = \sum_{k=1}^{n-1} \big(b_k - b_{k+1}\big) = \nu(1_{\mathbb{X}}) - b_n \tag{19}$$

due to $\beta_1(1_{\mathbb{X}}) = \nu(1_{\mathbb{X}})$.

□

Recall that a set $S \in \mathcal{X}$ is said to be an atom for P if: $\forall (a, a') \in S^2,\ P(a, \cdot) = P(a', \cdot)$. Then Condition $(\mathbf{S})$ holds for $\nu(\cdot) := P(a_0, \cdot)$ with some (any) $a_0 \in S$. In the atomic case, Assertions (ii) or (iii) of Theorem 2.1 correspond to the well-known conditions involving the first return time in an atom. More precisely, let us assume that S is an atom for P and define R_S as the first return time in S:

$$R_S := \inf\{n \geq 1 : X_n \in S\}.$$

Then, we have

$$\forall n \geq 1, \quad \beta_n(1_S) = \mathbb{P}_{a_0}\big(R_S = n\big) \qquad \text{and} \qquad \beta_n(1_{\mathbb{X}}) = \mathbb{P}_{a_0}\big(R_S \geq n\big) \tag{20}$$

with $\beta_n(\cdot)$ defined from S and $\nu(\cdot) := P(a_0, \cdot)$ with some $a_0 \in S$. Hence Assertion (ii) of Theorem 2.1 rewrites as $\sum_{k=1} \mathbb{P}(R_S \geq k) < \infty$ and Assertion (iii) as $\mathbb{P}_{a_0}(R_S < \infty) = 1$, and $\sum_{k=1}^{+\infty} k\, \mathbb{P}(R_S = k) < \infty$. Both assertions read as the usual moment condition of the return time in $S : \mathbb{E}_{a_0}[R_S] < \infty$.

3 Approximation of π in total variation norms

If $W : \mathbb{X} \to [1, +\infty)$ is a measurable function, then the W-weighted total variation norm $\|\lambda_1 - \lambda_2\|_W$ for any $(\lambda_1, \lambda_2) \in (\mathcal{M}^+)^2$ is defined by

$$\|\lambda_1 - \lambda_2\|_W := \sup_{|f| \leq W} \big|\lambda_1(f) - \lambda_2(f)\big|. \tag{21}$$

If $W := 1_{\mathbb{X}}$, then $\|\lambda_1 - \lambda_2\|_{1_X} = \|\lambda_1 - \lambda_2\|_{TV}$ is the standard total variation norm. If λ_1 and λ_2 are probability measures on $(\mathbb{X}, \mathcal{X})$, then $\|\lambda_1 - \lambda_2\|_{TV}$ corresponds to their standard total variation distance.

Under Assumption $(\mathbf{S})$ recall that, if $\sum_{k=1}^{+\infty} \beta_k(1_{\mathbb{X}}) < \infty$ (equivalently if one of the three Assertions (i), (iii) or (iv) of Theorem 2.1 holds), then we denote by $\mu := \sum_{k=1}^{+\infty} \beta_k$ the $P-$invariant positive measure given in Theorem 2.1. For every $n \geq 1$ let us define on $(\mathbb{X}, \mathcal{X})$ the following finite positive measure μ_n and probability measure $\widetilde{\mu}_n$:

$$\forall n \geq 1, \quad \mu_n := \sum_{k=1}^{n} \beta_k \quad \text{and} \quad \widetilde{\mu}_n := \frac{1}{\mu_n(1_{\mathbb{X}})}\, \mu_n.$$

Theorem 3.1 *Assume that P satisfies Condition* $(\mathbf{S})$ *and that for some measurable function $W \geq 1_{\mathbb{X}}$ we have $\mu(W) < \infty$ (thus $\mu(1_{\mathbb{X}}) < \infty$). Let $\pi := \mu/\mu(1_{\mathbb{X}})$. Then*

$$\forall n \geq 1, \quad \|\pi - \mu(1_{\mathbb{X}})^{-1}\mu_n\|_W = \mu(1_{\mathbb{X}})^{-1}\,\varepsilon_{n,W} \leq \varepsilon_{n,W} \tag{22a}$$

$$\forall n \geq 1, \quad \|\pi - \widetilde{\mu}_n\|_W \leq \mu(1_{\mathbb{X}})^{-1}\big(\varepsilon_{n,W} + \mu_n(W)\,\mu_n(1_{\mathbb{X}})^{-1}\,\varepsilon_n\big) \tag{22b}$$

$$\textit{with} \quad \forall n \geq 1, \quad \varepsilon_{n,W} := \sum_{k=n+1}^{+\infty} \beta_k(W) \quad \textit{and} \quad \varepsilon_n := \varepsilon_{n,1_{\mathbb{X}}} = \sum_{k=n+1}^{+\infty} \beta_k(1_{\mathbb{X}}). \tag{23}$$

Such an estimate on $\|\pi - \widetilde{\mu}_n\|_W$ allows us to quantify the quality of the approximation by $(\tilde{\mu}_n(f))_{n\geq 1}$ of $\pi(f)$ when f is some unbounded function such that $|f| \leq W$. Note that under the assumptions of Theorem 3.1, since $\mu(1_{\mathbb{X}}) \leq \mu(W) < \infty$, we can always use Estimates (22a)-(22b) for $W := 1_{\mathbb{X}}$ to get the following bound for the standard total variation norm

$$\forall n \geq 1, \quad \|\pi - \widetilde{\mu}_n\|_{TV} \leq 2\mu(1_{\mathbb{X}})^{-1}\,\varepsilon_n \leq 2\,\varepsilon_n. \tag{24}$$

Proof. We have

$$\|\pi - \mu_n/\mu(1_{\mathbb{X}})\|_W = (\mu - \mu_n)(W)/\mu(1_{\mathbb{X}}) = \varepsilon_{n,W}/\mu(1_{\mathbb{X}})$$

since $\pi = \mu/\mu(1_{\mathbb{X}})$ and $\mu - \mu_n \in \mathcal{M}^+$, so that $\|\mu - \mu_n\|_W = (\mu - \mu_n)(W) = \varepsilon_{n,W}$ from (23). The last inequality in (22a) follows from $\mu(1_{\mathbb{X}}) \geq \mu(1_S) = 1$ (see Theorem 2.2-(iii)). To prove (22b) consider any measurable function $f : \mathbb{X} \to \mathbb{R}$ such that $|f| \leq W$. Then

$$\begin{aligned}
\big|\pi(f) - \widetilde{\mu}_n(f)\big| &= \left|\pi(f) - \frac{\mu_n(f)}{\mu_n(1_{\mathbb{X}})}\right| \\
&\leq \left|\pi(f) - \frac{\mu_n(f)}{\mu(1_{\mathbb{X}})}\right| + |\mu_n(f)| \times \left|\frac{1}{\mu(1_{\mathbb{X}})} - \frac{1}{\mu_n(1_{\mathbb{X}})}\right| \\
&\leq \frac{\varepsilon_{n,W}}{\mu(1_{\mathbb{X}})} + \mu_n(W)\left|\frac{\mu_n(1_{\mathbb{X}}) - \mu(1_{\mathbb{X}})}{\mu(1_{\mathbb{X}})\mu_n(1_{\mathbb{X}})}\right|
\end{aligned}$$

by using the triangle inequality, (22a) and $|\mu_n(f)| \leq \mu_n(W)$, and finally $|\mu_n(1_{\mathbb{X}}) - \mu(1_{\mathbb{X}})| = (\mu - \mu_n)(1_{\mathbb{X}}) = \varepsilon_n$ from (23). □

Remark 3.1 *Under the assumptions of Theorem 3.1, alternative bounds to (22a)-(22b) can be proposed according to the needs.*

- *From (17) the constant $\mu(1_{\mathbb{X}})^{-1}$ in (22a)-(22b) is equal to $\pi(1_S)$ which is less than 1. But this cannot be used here since π is supposed to be unknown.*
- *We have*

$$\forall n \geq 1, \quad \|\pi - \widetilde{\mu}_n\|_W \leq \nu(1_{\mathbb{X}})^{-1}\big(\varepsilon_{n,W} + \mu(W)\,\nu(1_{\mathbb{X}})^{-1}\,\varepsilon_n\big) \tag{25}$$

which follows from $\mu_n(W) \leq \mu(W)$ and from $\mu(1_{\mathbb{X}}) \geq \mu_n(1_{\mathbb{X}}) \geq \beta_1(1_{\mathbb{X}}) = \nu(1_{\mathbb{X}})$. In atomic case we have $\nu(1_{\mathbb{X}}) = 1$, so that

$$\forall n \geq 1, \quad \|\pi - \widetilde{\mu}_n\|_W \leq \varepsilon_{n,W} + \mu(W)\,\varepsilon_n \qquad \textit{(atomic case)}. \tag{26}$$

- *In the non-atomic case $\nu(1_{\mathbb{X}})$ is small in general, in which case Estimate (25) can be replaced with the following one. Set $n^* := \min\big\{\ell \geq 1 : \varepsilon_\ell \leq 1/2\big\}$. Then*

$$\forall n \geq n^*, \quad \|\pi - \widetilde{\mu}_n\|_W \leq \mu(1_{\mathbb{X}})^{-1}\big(\varepsilon_{n,W} + 2\mu(W)\,\mu(1_{\mathbb{X}})^{-1}\,\varepsilon_n\big) \leq \varepsilon_{n,W} + 2\mu(W)\,\varepsilon_n. \quad (27)$$

  Note that the last bound in (27) is that of the atomic case in (26), up to the factor 2 in the second term, and that a priori (27) only holds for $n \geq n^*$ in the non-atomic case. If n^* is replaced with $n^*_\eta := \min\big\{\ell \geq 1 : \varepsilon_\ell \leq \eta\big\}$ for $\eta \in (0, 1/2)$, then the factor 2 in (27) is replaced with a factor close to one when $\eta \to 0$, but it is worth noticing that $n^*(\eta)$ is larger and larger when $\eta \to 0$. To prove (27), first note that n^* is well-defined since $\lim_n \varepsilon_n = 0$, and that

$$\forall n \geq n^*, \quad \mu_n(1_{\mathbb{X}}) \geq \mu(1_{\mathbb{X}}) - 1/2 \geq \mu(1_{\mathbb{X}})/2$$

  from the definition of ε_n and n^*, and from $\mu(1_{\mathbb{X}}) \geq \mu(1_S) = 1$ (see Theorem 2.2-(iii)). Thus the first inequality in (27) follows from (22b). The second one follows from $\mu(1_{\mathbb{X}}) \geq 1$.

Note that the constants associated with $\varepsilon_{n,W}$ and ε_n in (25) and (27) do not depend on n.

Remark 3.2 *If P satisfies Condition* (**S**) *for $S := \mathbb{X}$, then P is uniformly ergodic, and we have $\sup_{x\in\mathbb{X}} \|P^n(x,\cdot) - \pi\|_{TV} \leq (1 - \nu(1_{\mathbb{X}}))^n$, e.g. see [RR04]. In this case, note that we have $\varepsilon_n = (1 - \nu(1_{\mathbb{X}}))^n$ in* (24) *since an easy induction provides: $\forall k \geq 1,\ \beta_k(1_{\mathbb{X}}) = \nu(1_{\mathbb{X}})\,(1 - \nu(1_{\mathbb{X}}))^{k-1}$.*

It is clear from Estimates (22a)-(22b) or (24) and from Definition (23) of $\varepsilon_{n,W}$ and ε_n that the rate of convergence to 0 of $\|\pi - \tilde{\mu}_n\|_W$ can be derived from good estimates of the convergence of the sequences $(\beta_n(W))_{n\geq 1}$ and $(\beta_n(1_{\mathbb{X}}))_{n\geq 1}$. This is the main objective of Sections 4-5 where appropriate drift conditions are introduced in order to obtain geometric or subgeometric convergence to 0 of $(\beta_n(W))_{n\geq 1}$ and $(\beta_n(1_{\mathbb{X}}))_{n\geq 1}$ and so of $(\varepsilon_{n,W})_{n\geq 1}$ and $(\varepsilon_n)_{n\geq 1}$.

4 Explicit bound under geometric drift conditions

Throughout the Sections 4 and 5, any measurable function $V : \mathbb{X} \to [1, +\infty)$ will be called a Lyapunov function. For the sake of simplicity, any Lyapunov function V in this section is assumed to satisfy: $\forall x \in \mathbb{X},\ (PV)(x) < \infty$. Hence, under Assumption (**S**), we have

$$\nu(V) < \infty.$$

Recall that we have set $\mu := \sum_{k=1}^{+\infty} \beta_k$ under Assumption (**S**). The following theoretical statement is derived from Theorem 3.1 and [HL20b, Th. 3.1].

Theorem 4.1 ([HL20b]) *Assume that P satisfies Condition* (**S**). *Moreover assume that*

$$\exists \delta \in (0,1),\ \exists L > 0, \quad PV \leq \delta\, V + L\, 1_S \qquad (\mathbf{D})$$

with respect to some Lyapunov function V. Then we have

$$\theta_V := \limsup_n [\beta_n(V)]^{\frac{1}{n}} < 1,$$

that is, for every $\theta \in (\theta_V, 1)$ there exists a positive constant C_θ such that

$$\forall n \geq 1, \quad \beta_n(1_{\mathbb{X}}) \leq \beta_n(V) \leq C_\theta\, \theta^n. \tag{28}$$

Moreover we have

$$\mu(1_{\mathbb{X}}) \leq \mu(V) \leq \max\big(C_\theta\, \theta/(1-\theta)\,,\, L/(1-\delta)\big) < \infty.$$

Finally Estimates (22a)-(22b) *hold with $W := V$ and*

$$\varepsilon_n \leq \varepsilon_{n,V} \leq \frac{C_\theta}{1-\theta}\, \theta^{n+1}. \tag{29}$$

Proof. Under Assumptions (**S**) and (**D**), we know from [HL20b, Th. 3.1] that $\theta_V < 1$. This gives (28). Then it follows from (28) that $\mu(V) \leq C_\theta\, \theta/(1-\theta) < \infty$. Moreover we deduce from (**D**) and from the $P-$invariance of μ that $\mu(V) \leq \delta\mu(V) + L$ since $\mu(1_S) = 1$, hence $\mu(V) \leq L/(1-\delta)$. Finally (29) follows from (23) and (28).

□

Remark 4.1 *From $\theta_V < 1$ we deduce that*

$$\theta_1 := \limsup_n [\beta_n(1_{\mathbb{X}})]^{\frac{1}{n}} \leq \theta_V < 1$$

since $1_X \leq V$ and $\beta_n \in \mathcal{M}^+$. If $\theta_1 < \theta_V$, then we obtain a more accurate bound for ε_n, that is: for every $\theta \in (\theta_1, 1)$, $\varepsilon_n \leq D_\theta \theta^{n+1}/(1-\theta)$ for some positive constant D_θ. Actually we do not know if the inequality $\theta_1 < \theta_V$ is true (nor if it holds for some instances).

As mentioned in [HL20b, Rem. 5.4], the real number $\limsup_n [\beta_n(V)]^{1/n}$ may be strictly less than the so-called spectral gap related to the V-geometric ergodicity of P. In this case the rate of convergence in (29) is better than that given by the V-geometric ergodicity, see Subsection 6.1.1. Recall that finding explicit rate and bound in the V-geometric ergodicity property are difficult issues. Similarly, finding explicit bounds for θ_V and for the constant C_θ in (29) are difficult a priori, because the inequality $\limsup_n [\beta_n(V)]^{1/n} < 1$ is obtained in [HL20b, Th. 3.1] thanks to spectral arguments.

Below various statements specify the explicit control of the error term ε_n in (22a)-(22b) under Assumption (**S**) and the following drift condition:

$$\exists \delta \in (0,1),\ \forall x \in S^c, \quad (PV)(x) \leq \delta\, V(x). \tag{$\boldsymbol{D_{S^c}}$}$$

Note that Condition ($\boldsymbol{D_{S^c}}$) is equivalent to (**D**) when PV is bounded on S.

Theorem 4.2 *Assume that P satisfies Condition* (**S**) *for some $S \in \mathcal{X}$ and $\nu \in \mathcal{M}_*^+$. Moreover assume that there exists a Lyapunov function V such that P satisfies ($\boldsymbol{D_{S^c}}$) and the following condition on S*

$$\forall x \in S, \quad (PV)(x) \leq \delta\, V(x) + \nu(V). \tag{$\boldsymbol{D_S}$}$$

Then

$$\forall n \geq 1, \quad \beta_n(1_{\mathbb{X}}) \leq \beta_n(V) \leq \nu(V)\, \delta^{n-1}. \tag{30}$$

Moreover we have

$$\mu(1_{\mathbb{X}}) \le \mu(V) \le \frac{\nu(V)}{1-\delta} < \infty. \tag{31}$$

Finally Estimates (22a)-(22b) *hold with* $W := V$ *and*

$$\forall n \ge 1, \quad \varepsilon_n \le \varepsilon_{n,V} \le \frac{\nu(V)}{1-\delta}\,\delta^n. \tag{32}$$

Note that Conditions ($\boldsymbol{D_{S^c}}$)-($\boldsymbol{D_S}$) rewrite in a single inequality as

$$PV \le \delta\, V + \nu(V)1_S.$$

However, in view of the proof of the next Corollary 4.2, it may be convenient to separate the condition on S^c and that on S.

Proof. Recall that $T = \nu(\cdot)1_S$. Then it follows from the last inequality that

$$(P-T)V = PV - \nu(V)1_S \le \delta V.$$

Using $P - T \ge 0$ and iterating the previous inequality gives

$$0 \le (P-T)^n V \le \delta^n\, V. \tag{33}$$

Next, it follows from (14) that

$$\forall n \ge 1, \quad \beta_n(V) = \nu\big((P-T)^{n-1}V\big) \le \nu(V)\,\delta^{n-1}.$$

This gives (30) and (31) due to $1_{\mathbb{X}} \le V$ and to the positivity of β_n and μ. Finally (32) follows from the definition of ε_n and $\varepsilon_{n,V}$ in (23) and from (30). $\square$

Remark 4.2 *As in the Theorem 4.1 (see Remark 4.1), it is worth noticing that the computable geometric bound (32) is the same for* $(\varepsilon_n)_{n\ge1}$ *and* $(\varepsilon_{n,V})_{n\ge1}$. *However the constants differ according that we deal with the error bound* (24) *in total variation distance or with the error bound* (22b) *in* $V-$*weighted total variation distance. Anyway note that (31) gives a computable bound of* $\mu(V)$ *which is useful in* (22b) *(applied here to* $W := V$*).*

If Condition (**S**) holds for an atom S and for $\nu(\cdot) := P(a_0, \cdot)$ with some (any) $a_0 \in S$, then Condition ($\boldsymbol{D_S}$) is fulfilled since

$$\forall x \in S, \quad PV(x) - \delta\, V(x) - \nu(V) = -\delta\, V(x) \le 0.$$

Consequently we obtain the following corollary of Theorem 4.2.

Corollary 4.1 (Atomic case) *Assume that* P *satisfies Condition* (**S**) *with an atom* S *and with* $\nu(\cdot)$ *defined by* $\nu(\cdot) := P(a_0, \cdot)$ *with some (any)* $a_0 \in S$. *Moreover assume that there exists a Lyapunov function* V *such that* P *satisfies the drift condition* ($\boldsymbol{D_{S^c}}$). *Then Estimates* (22a)-(22b) *hold with* $W := V$ *and with* $(\varepsilon_n)_{n\ge1}$ *and* $(\varepsilon_{n,V})_{n\ge1}$ *satisfying (32) (recall that (26) is an alternative bound to (22b) in the atomic case).*

Remark 4.3 *In the atomic case, the bound $\beta_n(1_{\mathbb{X}}) \le \nu(V)\,\delta^{n-1}$ (see (30)) may be derived from well-known results under Assumption $(\boldsymbol{D_{S^c}})$. Indeed we know from (20) that $\beta_n(1_{\mathbb{X}}) = \mathbb{P}_{a_0}\big(R_S \ge n\big)$, where R_S is the first return time in S. Moreover $(\boldsymbol{D_{S^c}})$ gives*

$$PV \le \delta\, V + (c - \delta\vartheta)\, 1_S \qquad \text{with } c := \nu(V) \text{ and } \vartheta := \inf_{x\in S} V(x).$$

Then we deduce from [DMPS18, Prop.4.3.3(ii)] that

$$\forall x \in \mathbb{X}, \quad \mathbb{E}_x\big[\delta^{-R_S}\big] \le V(x) + (c - \delta\vartheta)\,\delta^{-1}. \tag{34}$$

hence

$$\mathbb{E}_{a_0}\big[\delta^{-R_S}\big] \le \vartheta + (c-\delta\vartheta)\delta^{-1} = c\,\delta^{-1}.$$

The same estimate as in (30) is obtained using Markov's inequality

$$\beta_n(1_{\mathbb{X}}) = \mathbb{P}_{a_0}\big(R_S \ge n\big) = \mathbb{P}_{a_0}\big(\delta^{-R_S} \ge \delta^{-n}\big) \le c\,\delta^{n-1}.$$

According to the previous discussion on the atomic case, the bound $\beta_n(1_{\mathbb{X}}) \le \nu(V)\,\delta^{n-1}$ obtained in (30), and consequently the resulting bound (32) for ε_n, are not only simple and explicit but also optimal. Although Condition $(\boldsymbol{D_S})$ is automatically satisfied in the atomic case and may hold in the non atomic case too, this condition is nevertheless restrictive. In the next corollary, the function V is replaced with V^{α_0} for some suitable $\alpha_0 \in (0,1]$ in Condition $(\boldsymbol{D_S})$ and Condition $(\boldsymbol{D_{S^c}})$ is preserved. The price to be paid is that the geometrical bound (32) for ε_n will hold with $\delta^{\alpha_0 n}$ in place of the expected rate δ^n. But the benefit will be that the bound for ε_n is still simple and explicit.

Let V be a Lyapunov function such that PV is bounded on S. Then

$$\exists \alpha_0 \in (0,1],\ \forall x \in S, \quad (PV^{\alpha_0})(x) \le \delta^{\alpha_0}\, V(x)^{\alpha_0} + \nu(V^{\alpha_0}). \tag{35}$$

Indeed, set $M_S := \sup_S PV$. Then, for every $\alpha \in (0,1]$, we have $1 \le \sup_S PV^\alpha \le {M_S}^\alpha$ from $1_{\mathbb{X}} \le V^\alpha$ and $PV^\alpha \le (PV)^\alpha$ using Jensen's inequality. Moreover

$$\forall x \in S, \quad (PV^\alpha)(x) - \delta^\alpha\, V(x)^\alpha - \nu(V^\alpha) \le {M_S}^\alpha - \delta^\alpha - \nu(1_{\mathbb{X}})$$

from $1_{\mathbb{X}} \le V$. Passing to the limit when $\alpha \to 0$ gives (35) since $\nu(1_{\mathbb{X}}) > 0$.

Corollary 4.2 *Assume that P satisfies Condition* $(\mathbf{S})$ *and that there exists a Lyapunov function V such that P satisfies Condition $(\boldsymbol{D_{S^c}})$ and PV is bounded on S (so that the usual drift condition* $(\mathbf{D})$ *holds). Let $\alpha_0 \in (0,1]$ provided by Property (35). Then*

$$\forall n \ge 1, \quad \beta_n(1_{\mathbb{X}}) \le \beta_n(V^{\alpha_0}) \le \nu(V^{\alpha_0})\,\delta^{\alpha_0(n-1)}. \tag{36}$$

Moreover we have

$$\mu(1_{\mathbb{X}}) \le \mu(V^{\alpha_0}) \le \frac{\nu(V^{\alpha_0})}{1-\delta^{\alpha_0}} < \infty. \tag{37}$$

Finally Estimates (22a)-(22b) *hold with $W := V^{\alpha_0}$ and*

$$\forall n \ge 1, \quad \varepsilon_n \le \varepsilon_{n,V^{\alpha_0}} \le \frac{\nu(V^{\alpha_0})}{1-\delta^{\alpha_0}}\,\delta^{\alpha_0 n}. \tag{38}$$

Proof. We have

$$\forall x \in S^c, \quad (PV^{\alpha_0})(x) \leq \delta^{\alpha_0}\, V(x)^{\alpha_0} \tag{39}$$

from $PV^{\alpha_0} \leq (PV)^{\alpha_0}$ (Jensen's inequality) and from ($\boldsymbol{D}_{\boldsymbol{S^c}}$). Moreover (35) holds. Then Corollary 4.2 follows from Theorem 4.2 applied to V^{α_0} and δ^{α_0} in place of V and δ. □

Recall that Inequality (35) holds with $\alpha_0 = 1$ in the atomic case (see Corollary 4.1). Inequality (35) may be also fulfilled with $\alpha_0 = 1$ in the non atomic case (e.g. see Subsection 6.1.2 and Table 1 in Subsection 6.3). If $\alpha_0 = 1$ does not work, the following statement is useful to find $\alpha_0 \in (0, 1)$ in (35).

Proposition 4.1 *Assume that P satisfies Condition* (**S**) *and that S is not an atom. Let $\sigma := 1 - \nu(1_{\mathbb{X}})$. Then we have for any Lyapunov function V:*

$$\forall \alpha \in (0,1],\ \forall x \in S, \quad (PV^{\alpha})(x) - \nu(V^{\alpha}) \leq \frac{\sigma}{\sigma^{\alpha}}\big[(PV)(x) - \nu(V)\big]^{\alpha}. \tag{40}$$

Proof. Let $x \in S$. Note that $\sigma_x(\cdot) := P(x,\cdot) - \nu(\cdot)$ is a non-negative measure on $(\mathbb{X}, \mathcal{X})$ from Assumption (**S**), and that $\sigma_x(1_{\mathbb{X}}) = 1 - \nu(1_{\mathbb{X}})$ does not depend on x. The case $\sigma = 0$ corresponds to the atomic case. Here we assume that $\sigma > 0$. Define the following probability measure on $(\mathbb{X}, \mathcal{X})$: $\widetilde{\sigma}_x(\cdot) = \sigma_x(\cdot)/\sigma$. Let $\alpha \in (0, 1]$. It follows from Jensen's inequality that

$$\frac{(PV^{\alpha})(x) - \nu(V^{\alpha})}{\sigma} = \widetilde{\sigma}_x(V^{\alpha}) \leq \big[\widetilde{\sigma}_x(V)\big]^{\alpha} = \frac{\big[(PV)(x) - \nu(V)\big]^{\alpha}}{\sigma^{\alpha}},$$

from which we deduce (40). □

The previous statements only concern the approximation of the stationary distribution π. To conclude this section recall that, in [HL20b, Cor. 2.3], the V-weighted operator norm of $P^n - T_n$ with T_n given in (13a) is proved to converge to zero with a geometric rate of convergence under Assumptions (**S**) and (**D**). Using Inequality (33), we specify this rate of convergence under the assumptions of Theorem 4.2 or Corollary 4.2.

Theorem 4.3 *Assume that P satisfies the assumptions of Theorem 4.2. Then*

$$\sup_{|f|\leq V}\ \sup_{x\in\mathbb{X}} \frac{\big|(P^n f)(x) - (T_n f)(x)\big|}{V(x)} \leq \delta^n \quad \textit{with} \quad T_n f = \sum_{k=1}^{n} \beta_k(f) P^{n-k} 1_S \tag{41}$$

where the functions f are assumed to be real-valued and measurable on $(\mathbb{X}, \mathcal{X})$. Similarly, if P satisfies the assumptions of Corollary 4.2, then Inequality (41) holds with V^{α_0} and δ^{α_0} in place of V and δ.

Proof. If P satisfies the assumptions of Theorem 4.2, then it follows from (33) that for every measurable function $f : \mathbb{X} \to \mathbb{R}$ such that $|f| \leq V$

$$|P^n f - T_n f| = |(P - T)^n f| \leq (P - T)^n |f| \leq (P - T)^n V \leq \delta^n\, V, \tag{42}$$

from which we deduce (41). Under the assumptions of Corollary 4.2, we know that Theorem 4.2 applies with V^{α_0} and δ^{α_0} in place of V and δ, so that (42) holds with V^{α_0} and δ^{α_0} in place of V and δ too.

□

Given any $A \in \mathcal{X}$, the error bound (24) combined with any bound on $|P^n(x,A) - \pi(1_A)|$ can be used to approximate $P^n(x,A)$. This is illustrated in the next theorem. If P satisfies the assumptions of Theorem 4.2 and the strong aperiodicity condition $\nu(1_S) > 0$, then the bound (41) can be used to obtain a rate of convergence in the V-geometrical ergodicity property, which simply depends on $\delta \in (0,1)$ in $(\boldsymbol{D_S})$ and on the real number

$$\varrho_S := \limsup_{n \to +\infty} \left(\sup_{x \in \mathbb{X}} \frac{\big|(P^n 1_S)(x) - \pi(1_S)\big|}{V(x)} \right)^{\frac{1}{n}}$$

introduced in [HL20b]. More precisely we know from [HL20b] that $\varrho_S < 1$. In Theorem 4.4 below we consider any $\varrho \in (\varrho_S, 1)$ and we define

$$\alpha := \max(\delta, \varrho) \qquad \text{and} \quad D_\varrho := \sup_{n \geq 0} \varrho^{-n} \sup_{x \in \mathbb{X}} \frac{\big|(P^n 1_S)(x) - \pi(1_S)\big|}{V(x)} < \infty.$$

Theorem 4.4 *Assume that P satisfies the assumptions of Theorem 4.2 and that $\nu(1_S) > 0$. Then we have*

$$\sup_{|f| \leq V} \sup_{x \in \mathbb{X}} \frac{\big|(P^n f)(x) - \pi(f)\big|}{V(x)} \leq \frac{\nu(V) + 1 - \delta}{1 - \delta}\ \delta^n + \frac{\nu(V)\, D_\varrho}{\delta}\, n\, \alpha^n. \tag{43}$$

Moreover, setting $c := 2\,\mu(1_{\mathbb{X}})^{-1} \leq 2$, the following inequality holds for every $n \geq 1$ and for every $A \in \mathcal{X}$:

$$\left| P^n(x,A) - \frac{\mu_n(1_A)}{\mu_n(1_{\mathbb{X}})} \right| \leq \left(\frac{(1+c)\nu(V) + 1 - \delta}{1 - \delta}\ \delta^n + \frac{\nu(V)\, D_\varrho}{\delta}\, n\, \alpha^n \right) V(x). \tag{44}$$

Similarly, if P satisfies the assumptions of Corollary 4.2 and if $\nu(1_S) > 0$, then the bounds (43) and (44) hold with V^{α_0} and δ^{α_0} in place of V and δ (the function V must be replaced by V^{α_0} in the definitions of ϱ_S and D_ϱ).

Proof. Property (43) can be easily obtained by using the bound (41) in the proof of [HL20b, Th. 5.3]. Then Inequality (44) follows from (43), (24) and (32) using $V \geq 1_{\mathbb{X}}$ and the triangle inequality. □

Similar inequality to (44) can be obtained with $\mu_n(1_A)/\mu(1_{\mathbb{X}})$ from (22a) with $W := 1_{\mathbb{X}}$. Of course any bound known for $|P^n(x,A) - \pi(1_A)|$ combined with (22a) (with $W := 1_{\mathbb{X}}$) or (24) can be used to obtain an approximate value of $P^n(x,A)$.

Remark 4.4 *Let r be the spectral radius of the operator $P - T$ on the $V-$weighted supremum space $(\mathcal{B}_V, \|\cdot\|_V)$ composed of the complex-valued measurable functions $f : \mathbb{X} \to \mathbb{C}$ such that $\|f\|_V := \sup_{\mathbb{X}} |f|/V < \infty$. Then (41) gives $r \leq \delta$. Consequently the proofs of [HL20b, Th. 5.3] and [HL20a, Th. A.1] can be easily adapted to obtain the following alternative:*

- *either $\varrho_S \leq \delta$*
- *or $\varrho_S = \theta^{-1}$ with $\theta := \min\big\{|z| : z \in \mathbb{C},\ 1 < |z| < 1/\delta,\ B_{1_{\mathbb{X}}}(z) = 0\big\}$, where $B_{1_{\mathbb{X}}}$ is the power series defined by $B_{1_{\mathbb{X}}}(z) := \sum_{k=1}^{+\infty} \beta_k(1_{\mathbb{X}})\, z^k$.*

This alternative is due to the following fact: if $\lambda \in \mathbb{C}$ is such that $\delta < |\lambda| \leq 1$, then λ is an eigenvalue of P on $\mathcal{B}_V$ if, and only if, $B_{1_{\mathbb{X}}}(\lambda^{-1}) = 0$ (see [HL20a, Prop. A.2]). We can observe that the bound (32) in Theorem 4.2 or Corollary 4.1 does not take into account the possible eigenvalues λ of P such that $\delta < |\lambda| < 1$. However note that the bound (43), thus (44), depends on the real number ϱ_S. If P admits eigenvalues in the annulus $\{z \in \mathbb{C} : \delta < |z| < 1\}$, then we have $\varrho_S = \theta^{-1}$ which is strictly greater than δ. Such atomic instances occur, see Subsection 6.1.1.

5 Explicit bounds under subgeometric drift conditions

For the sake of simplicity, any Lyapunov function V in this section is assumed to satisfy: $\forall x \in \mathbb{X},\ (PV)(x) < \infty$.

5.1 Theoretical results

Let P be a Markov kernel on $(\mathbb{X}, \mathcal{X})$ satisfying Condition (**S**). Let $T(\cdot) := \nu(\cdot)1_S$. For any integer $m \geq 1$, let us introduce the following condition: there exists a collection $\{V_i\}_{i=0}^m$ of Lyapunov functions such that

$$\forall i \in \{0, \dots, m-1\}, \quad (P - T)V_i \leq V_i - V_{i+1}. \tag{45}$$

Note that the properties (45) and $P - T \geq 0$ from (**S**) give

$$V_m \leq V_{m-1} \leq \cdots \leq V_1 \leq V_0.$$

Since $(PV_0)(\cdot) < \infty$ by hypothesis, we have under Assumption (**S**)

$$\nu(V_0) < \infty. \tag{46}$$

In this section, first we present a theoretical result which shows that Estimates (22a)-(22b) hold with a polynomial rate of convergence under Assumptions (45). Second we propose further statements in which an explicit polynomial rate of convergence is obtained. Denote by $(\vartheta_j)_{j\geq 0}$ the recurrent sequence of positive real numbers defined by

$$\vartheta_0 := 1 \quad \text{and} \quad \forall \ell \geq 1,\ \vartheta_\ell := \sum_{j=0}^{\ell-1} C_\ell^j \vartheta_j \quad \text{with} \quad C_\ell^j := \frac{\ell\,!}{j\,!(\ell - j)\,!}. \tag{47}$$

Theorem 5.1 *Let P be a Markov kernel on $(\mathbb{X}, \mathcal{X})$ satisfying Condition (**S**). Moreover, assume that there exists a collection $\{V_i\}_{i=0}^m$ of Lyapunov functions satisfying Conditions (45). Then we have*

$$\forall j \in \{1, \dots, m\}, \quad \sum_{k=1}^{+\infty} k^{j-1}\, \beta_k(V_j) \leq \vartheta_{j-1}\, \nu(V_0) < \infty. \tag{48}$$

Moreover for every $j = 1, \cdots, m$ we have $\pi(V_j) \leq \mu(V_j) = \sum_{k=1}^{+\infty} \beta_k(V_j) < \infty$, and Estimates (22a)-(22b) hold with $W := V_j$ and with $(\varepsilon_{n,V_j})_{n\geq 1}$ and $(\varepsilon_n)_{n\geq 1}$ satisfying

$$\forall j = 1, \cdots, m,\ \lim_{n\to+\infty} n^{j-1}\varepsilon_{n,V_j} = 0 \qquad \text{and} \qquad \lim_{n\to+\infty} n^{m-1}\varepsilon_n = 0. \tag{49}$$

Proof. Let us prove Inequality (48) by an induction on m. Assume that (45) holds with $m=1$, that is $(P-T)V_0 \le V_0 - V_1$, or equivalently: $V_1 \le V_0 - (P-T)V_0$. Then

$$\forall k \ge 0, \quad (P-T)^k V_1 \le (P-T)^k V_0 - (P-T)^{k+1} V_0$$

from which we deduce that

$$\forall n \ge 1, \quad \sum_{k=0}^{n} (P-T)^k V_1 \le \sum_{k=0}^{n} \left[(P-T)^k V_0 - (P-T)^{k+1} V_0\right] \le V_0.$$

It follows from (14) that

$$\forall n \ge 1, \quad \sum_{k=1}^{n+1} \beta_k(V_1) \le \nu(V_0).$$

This proves (48) when $m=1$. Now suppose that Inequalities (48) hold for some $m \ge 1$. Assume that (45) holds at order $m+1$. Then using $V_{m+1} \le V_m - (P-T)V_m$, we get

$$\forall k \ge 0, \quad (P-T)^k V_{m+1} \le (P-T)^k V_m - (P-T)^{k+1} V_m$$

so that we have for every $n \ge 1$

$$\begin{aligned}
\sum_{k=0}^{n} (k+1)^m \, (P-T)^k V_{m+1} &\le \sum_{k=0}^{n} (k+1)^m \, (P-T)^k V_m - \sum_{k=0}^{n+1} k^m \, (P-T)^k V_m \\
&\le \sum_{k=0}^{n} \left[(k+1)^m - k^m\right] (P-T)^k V_m \\
&\le \sum_{j=0}^{m-1} C_m^j \sum_{k=0}^{n} k^j \, (P-T)^k V_m \\
&\le \sum_{j=1}^{m} C_m^{j-1} \sum_{k=0}^{n} k^{j-1} \, (P-T)^k V_j
\end{aligned}$$

using $\forall j \in \{1,\ldots,m\},\ V_m \le V_j$ for the last inequality. It follows from (14) that

$$\begin{aligned}
\sum_{k=1}^{+\infty} k^m \, \beta_k(V_{m+1}) \le \sum_{j=1}^{m} C_m^{j-1} \sum_{k=0}^{+\infty} k^{j-1} \, \beta_{k+1}(V_j) &\le \sum_{j=1}^{m} C_m^{j-1} \sum_{k=1}^{+\infty} k^{j-1} \, \beta_k(V_j) \\
&\le \left(\sum_{j=1}^{m} C_m^{j-1} \vartheta_{j-1}\right) \nu(V_0) \quad = \vartheta_m \, \nu(V_0)
\end{aligned}$$

from the induction hypothesis. This gives Inequalities (48) at order $m+1$.

Now let us prove the last assertion of Theorem 5.1. First note that for every $j = 1, \cdots, m$ we have $\pi(V_j) \le \mu(V_j) = \sum_{k=1}^{+\infty} \beta_k(V_j) < \infty$ from (16) and (48). Next we have

$$\forall j = 1, \cdots, m, \quad \varepsilon_{n,V_j} = \sum_{k=n+1}^{+\infty} \beta_k(V_j) \le \frac{1}{(n+1)^{j-1}} \sum_{k=n+1}^{+\infty} k^{j-1} \beta_k(V_j).$$

Then the first assertion in (49) follows from (48). In particular we have $\lim_n n^{m-1} \varepsilon_{n,V_m} = 0$, so that $\lim_{n \to +\infty} n^{m-1} \varepsilon_n = 0$ since $\varepsilon_n \le \varepsilon_{n,V_m}$ from $1_{\mathbb{X}} \le V_m$.

□

In the following statement, under the assumptions of Theorem 5.1, we specify the asymptotic behaviour of the sequence $(\beta_k(V_m))_{k\geq 1}$ which is assumed to be decreasing.

Theorem 5.2 *Let P be a Markov kernel on $(\mathbb{X}, \mathcal{X})$ satisfying Condition* (**S**)*. Moreover assume that there exists a collection $\{V_i\}_{i=0}^m$ of Lyapunov functions satisfying Conditions (45). Then the following assertions hold.*

(i) $\forall i \in \{0, \ldots, m\},\ \forall k \geq 1,\ \beta_k(V_i) < \infty.$

(ii) If the sequence $(\beta_k(V_m))_{k\geq 1}$ is decreasing, then

$$\forall n \geq 1, \quad \beta_n(V_m) \leq \frac{C_m\, \nu(V_0)}{n^m} \quad \textit{with} \quad C_m := 2^{\frac{m(m+1)}{2}-1}. \tag{50}$$

(iii) If the sequence $\big(\beta_k(V_m)\big)_{k\geq 1}$ is decreasing and if $\mu(V_0) := \sum_{k=1}^{+\infty} \beta_k(V_0) < \infty$, then

$$\forall n \geq 1, \quad \beta_n(V_m) \leq \frac{D_m\, \mu(V_0)}{n^{m+1}} \quad \textit{with} \quad D_m := 2^{\frac{(m+1)(m+2)}{2}+1}. \tag{51}$$

Lemma 5.1 *Assume that P satisfies Condition* (**S**)*. Let V and W be two Lyapunov functions such that*

$$(P - T)V \leq V - W \quad \textit{where} \quad T(\cdot) := \nu(\cdot)1_S. \tag{52}$$

Then the following assertions hold.

(a) $\forall k \geq 1,\ \beta_k(V) < \infty.$

(b) The sequence $\big(\beta_k(V)\big)_{k\geq 1}$ is decreasing.

(c) If the sequence $\big(\beta_k(W)\big)_{k\geq 1}$ is decreasing, then we have for every $k \geq 1$ and $\varepsilon \in \{0, 1\}$

$$\beta_k(W) \leq \frac{\nu(V)}{k} \quad \textit{and} \quad \beta_{2k-\varepsilon}(W) \leq \frac{\beta_k(V)}{k}.$$

(d) If $\mu(V) := \sum_{k=1}^{+\infty} \beta_k(V) < \infty$ and if the sequence $\big(\beta_k(W)\big)_{k\geq 1}$ is decreasing, then

$$\forall n \geq 1, \quad \beta_n(W) \leq \frac{16\, \mu(V)}{n^2}.$$

Proof. Note that $W \leq V$ from (52) and $P - T \geq 0$. Next we deduce from (52) that $\forall j \geq 1,\ (P - T)^j V \leq (P - T)^{j-1}(V - W)$. Then (14) gives

$$\forall j \geq 1, \quad \beta_{j+1}(V) \leq \beta_j(V) - \beta_j(W) \leq \beta_j(V) \quad \text{in } [0, +\infty].$$

Using $\beta_1(V) = \nu(V) < \infty$, Assertion (a) is obtained by induction, and Assertion (b) is then obvious. Next rewrite the previous inequalities as

$$\forall j \geq 1, \quad 0 \leq \beta_j(W) \leq \beta_j(V) - \beta_{j+1}(V) \tag{53}$$

and suppose that $(\beta_j(W))_{j\geq 1}$ is decreasing. Then it follows from (53) that

$$\forall k \geq 1, \quad k\, \beta_k(W) \leq \sum_{j=1}^{k} \beta_j(W) \leq \beta_1(V) - \beta_{k+1}(V) \leq \nu(V),$$

from which we deduce the first inequality in Assertion (c). Moreover (53) also gives

$$\forall k\ge 1,\ \forall \varepsilon\in\{0,1\}\quad k\,\beta_{2k-\varepsilon}(W)\le \sum_{j=k}^{2k-\varepsilon}\beta_j(W)\le \beta_k(V)-\beta_{2k-\varepsilon+1}(V)\le \beta_k(V), \tag{54}$$

from which we deduce the second inequality in Assertion (c). Finally, to prove Assertion (d), note that for every $\ell\ge 1$ and every $\varepsilon\in\{0,1\}$

$$\ell\,\beta_{2\ell-\varepsilon}(V)\le \sum_{j=\ell}^{2\ell-\varepsilon}\beta_j(V)\le \mu(V)<\infty \tag{55}$$

since $(\beta_j(V))_{j\ge1}$ is decreasing (Assertion (b)). Let $n\ge1$ and write $n=2(2\ell-\varepsilon_1)-\varepsilon_2$ with $\ell\ge1$ and $(\varepsilon_1,\varepsilon_2)\in\{0,1\}^2$. Then it follows from (54) and (55) that

$$\beta_n(W)\le \frac{\beta_{2\ell-\varepsilon_1}(V)}{2\ell-\varepsilon_1}\le \frac{\mu(V)}{\ell(2\ell-1)}\le \frac{\mu(V)}{\ell^2}=\frac{16\,\mu(V)}{(n+2\varepsilon_1+\varepsilon_2)^2}\le \frac{16\,\mu(V)}{n^2}.$$

□

Proof of Theorem 5.2. Assertion (a) of Lemma 5.1 applied with $V:=V_0$ and $W:=V_1$ proves that: $\forall k\ge1,\ \beta_k(V_0)<\infty$. Then Assertion (i) of Theorem 5.2 holds since $V_i\le V_0$. Now let us prove by induction on the positive integer m that Property (50) holds. If $m=1$, then the first inequality in Assertion (c) of Lemma 5.1 applied with $V:=V_0$ and $W:=V_1$ provides

$$\forall n\ge1,\quad \beta_n(V_1)\le\frac{\nu(V_0)}{n}.$$

Hence (50) holds with $C_1=1$ when $m=1$. Now suppose that (50) holds for some $m\ge1$. Let $\{V_i\}_{i=0}^{m+1}$ be a collection of Lyapounov functions such that

$$\forall i\in\{0,\dots,m\},\quad (P-T)V_i\le V_i-V_{i+1}$$

and finally such that the sequence $(\beta_k(V_{m+1}))_{k\ge1}$ is decreasing. Note that Assertion (b) of Lemma 5.1 applied with $V:=V_m$ and $W:=V_{m+1}$ ensures that the sequence $(\beta_k(V_m))_{k\ge1}$ is decreasing. Consequently we have

$$\forall k\ge1,\quad \beta_k(V_m)\le\frac{C_m\,\nu(V_0)}{k^m}\quad\text{with}\quad C_m:=2^{\frac{m(m+1)}{2}-1} \tag{56}$$

from the induction hypothesis. Next let $n\ge1$ and write $n=2k-\varepsilon$ with $k\ge1$ and $\varepsilon\in\{0,1\}$. Then the second inequality in Assertion (c) of Lemma 5.1 applied with $V:=V_m$ and $W:=V_{m+1}$ gives

$$\beta_n(V_{m+1})\le\frac{\beta_k(V_m)}{k} \tag{57}$$

so that $\beta_n(V_{m+1})\le C_m\nu(V_0)/k^{m+1}$ from (56). Hence

$$\beta_n(V_{m+1})\le\frac{2^{m+1}\,C_m\,\nu(V_0)}{(n+\varepsilon)^{m+1}}\le\frac{C_{m+1}\,\nu(V_0)}{n^{m+1}}\quad\text{with}\quad C_{m+1}=2^{m+1}\,C_m=2^{\frac{(m+1)(m+2)}{2}-1}.$$

We have proved Assertion (ii) of Theorem 5.2.

The proof of Assertion (iii) of Theorem 5.2 follows the same induction procedure. Indeed, if $m = 1$, then Assertion (d) of Lemma 5.1 applied with $V := V_0$ and $W := V_1$ provides

$$\forall n \geq 1, \quad \beta_n(V_1) \leq \frac{16\,\mu(V_0)}{n^2}.$$

Hence (51) holds with $D_1 = 16$ when $m = 1$. Now, assume that (51) is true at order m for some $m \geq 1$, and consider a collection $\{V_i\}_{i=0}^{m+1}$ of Lyapunov functions as in the above induction proof. Then, writing $n \geq 1$ as $n = 2k - \varepsilon$ with $k \geq 1$ and $\varepsilon \in \{0, 1\}$, we deduce from (57) and from the induction hypothesis that

$$\beta_n(V_{m+1}) \leq \frac{\beta_k(V_m)}{k} \leq \frac{D_m\,\mu(V_0)}{k^{m+2}} \quad \text{with} \quad D_m := 2^{\frac{(m+1)(m+2)}{2}+1}.$$

Hence

$$\beta_n(V_{m+1}) \leq \frac{2^{m+2}\,D_m\,\mu(V_0)}{(n+\varepsilon)^{m+2}} \leq \frac{D_{m+1}\,\mu(V_0)}{n^{m+2}} \quad \text{with} \quad D_{m+1} = 2^{m+2}\,D_m.$$

This proves (51).

□

Under Conditions (45) the smallest function V_m can be replaced by $1_{\mathbb{X}}$. Moreover for every $j = 1, \ldots, m-1$ the sequence $\big(\beta_k(V_j)\big)_{k\geq 1}$ is decreasing from Assertion (b) of Lemma 5.1. This allows us to deduce computable bounds for the error terms ε_n and ε_{n,V_j} in (22a)-(22b) when applied to $W := V_j$.

Corollary 5.1 *Let P be a Markov kernel on $(\mathbb{X}, \mathcal{X})$ satisfying Condition* (**S**) *and Conditions (45) with respect to some collection $\{V_i\}_{i=0}^{m}$ of Lyapunov functions. Then the following assertions hold with the positive constants C_j and D_j defined in Theorem 5.2.*

(a) If $m \geq 2$, then $\mu(1_{\mathbb{X}}) < \infty$, and Estimate (24) *holds with*

$$\forall n \geq 1, \quad \varepsilon_n \leq \frac{C_m\,\nu(V_0)}{m-1}\,\frac{1}{n^{m-1}}. \tag{58}$$

Moreover, if $m \geq 3$, then for every $j = 2, \ldots, m-1$ we have $\pi(V_j) \leq \mu(V_j) < \infty$, and Estimates (22a)-(22b) *hold with $W := V_j$ and*

$$\forall n \geq 1, \quad \varepsilon_{n,V_j} \leq \frac{C_j\,\nu(V_0)}{j-1}\,\frac{1}{n^{j-1}}. \tag{59}$$

(b) If $m \geq 1$ and $\mu(V_0) < \infty$, then Estimate (24) *holds with*

$$\forall n \geq 1, \quad \varepsilon_n \leq \frac{D_m\,\mu(V_0)}{m}\,\frac{1}{n^m}. \tag{60}$$

Moreover, if $m \geq 2$, then for every $j = 1, \ldots, m-1$ Estimates (22a)-(22b) *hold with $W := V_j$ and*

$$\forall n \geq 1, \quad \varepsilon_{n,V_j} \leq \frac{D_j\,\mu(V_0)}{j}\,\frac{1}{n^j}. \tag{61}$$

Note that, using the triangle inequality, any quantitative error bounds on $|P^n(x,A) - \pi(1_A)|$ as in [AF05, AFV15] can be combined with (58) or (60) to approximate the value of $P^n(x,A)$ for any $A \in \mathcal{X}$ and to control the error.

Proof. As previously mentioned, the function V_m in (45) can be replaced by $1_{\mathbb{X}}$. Moreover recall that the sequence $(\beta_k(1_{\mathbb{X}}))_{k\geq 1}$ is decreasing from (18). Hence it follows from (50) that

$$\forall n \geq 1, \quad \beta_n(1_{\mathbb{X}}) \leq \frac{C_m\, \nu(V_0)}{n^m}. \tag{62}$$

If $m \geq 2$, then Condition (ii) of Theorem 2.1 is fulfilled thanks to (62). Then Inequality (58) is deduced from

$$\forall n \geq 1, \quad \varepsilon_n = \sum_{k=n+1}^{+\infty} \beta_k(1_{\mathbb{X}}) \leq C_m \nu(V_0) \sum_{k=n+1}^{+\infty} \frac{1}{k^m} \leq C_m \nu(V_0) \int_n^{+\infty} \frac{dt}{t^m} = \frac{C_m\, \nu(V_0)}{(m-1)\, n^{m-1}}.$$

Now assume that $\{V_i\}_{i=0}^m$ satisfies Conditions (45) with $m \geq 3$. Let $j \in \{2, \ldots, m-1\}$. The sequence $(\beta_k(V_j))_{k\geq 1}$ is decreasing from Assertion (b) of Lemma 5.1, and obviously $\{V_i\}_{i=0}^j$ also satisfies Conditions (45). Then it follows from (50) that

$$\forall n \geq 1, \quad \beta_n(V_j) \leq \frac{C_j\, \nu(V_0)}{n^j} \quad \text{with} \quad C_j := 2^{\frac{j(j+1)}{2}-1}. \tag{63}$$

Thus $\pi(V_j) \leq \mu(V_j) < \infty$ since $j \geq 2$, and (59) follows from comparison sums/integrals as above. Finally assume that $\mu(V_0) < \infty$ and $m \geq 1$. We deduce from (51) that

$$\forall n \geq 1, \quad \beta_n(1_{\mathbb{X}}) \leq \frac{D_m\, \mu(V_0)}{n^{m+1}}. \tag{64}$$

Then (60) can be derived from comparison sums/integrals. Next assume that $m \geq 2$, and let $j \in \{1, \ldots, m-1\}$. Then Property (61) can be established by using as above the family $\{V_i\}_{i=0}^j$ and the fact that the sequence $(\beta_k(V_j))_{k\geq 1}$ is decreasing, then by applying (51) to V_j (in place of V_m), and finally by using again comparison sums/integrals. □

5.2 Applications

Let P be a Markov kernel on $(\mathbb{X}, \mathcal{X})$ satisfying Condition $(\mathbf{S})$ for some $S \in \mathcal{X}$ and $\nu \in \mathcal{M}_*^+$. For $m \geq 1$ let us introduce the following condition: there exists a collection $\{V_i\}_{i=0}^m$ of Lyapunov functions such that

$$\forall i \in \{0, \ldots, m-1\}, \quad \begin{cases} \forall x \in \mathbb{X}, & V_{i+1}(x) \leq V_i(x) \\ \forall x \in S^c, & (PV_i)(x) \leq V_i(x) - V_{i+1}(x) \\ \forall x \in S, & (PV_i)(x) \leq \big(V_i(x) - V_{i+1}(x)\big) + \nu(V_i). \end{cases} \tag{65}$$

Note that the second condition in (65) implies that $V_{i+1} \leq V_i$ on S^c, so that the first condition may be replaced with $V_{i+1} \leq V_i$ on S. Also note that the term $(V_i(x) - V_{i+1}(x))$ in the third condition of (65) is non-negative. Finally observe that Condition (65) rewrites in a more concise form as follows

$$\forall i \in \{0, \ldots, m-1\}, \quad \begin{cases} V_{i+1} \leq V_i \\ PV_i \leq \big(V_i - V_{i+1}\big) + \nu(V_i)\, 1_S. \end{cases}$$

However, as in the previous section, it is convenient to separate the conditions on S^c and S respectively.

Corollary 5.2 *Assume that P satisfies Condition* (**S**) *and that there exists a collection $\{V_i\}_{i=0}^m$ of Lyapunov functions satisfying Conditions (65). Then P satisfies Conditions (45) w.r.t. $\{V_i\}_{i=0}^m$, so that all the assertions of Theorem 5.2 and Corollary 5.1 hold.*

Proof. Prove that (65) implies (45), so that Theorem 5.2 and Corollary 5.1 apply. We have

$$\begin{aligned}\forall i \in \{0,\dots,m-1\}, \quad (P-T)V_i &= 1_{S^c}\big(PV_i - \nu(V_i)1_S\big) + 1_S\big(PV_i - \nu(V_i)1_S\big) \\ &= 1_{S^c}\, PV_i + 1_S\big(PV_i - \nu(V_i)1_S\big) \\ &\le 1_{S^c}\big(V_i - V_{i+1}\big) + 1_S\big(PV_i - \nu(V_i)1_S\big) \\ &= V_i - V_{i+1} + 1_S\big(PV_i - V_i + V_{i+1} - \nu(V_i)1_S\big) \\ &\le V_i - V_{i+1}.\end{aligned}$$

This gives (45). □

Corollary 5.3 (Atomic case) *Let P be a Markov kernel on $(\mathbb{X},\mathcal{X})$ satisfying Condition* (**S**) *with an atom S and with $\nu(\cdot)$ defined by $\nu(\cdot) := P(a_0,\cdot)$ for $a_0 \in S$. Moreover assume that there exists a collection $\{V_i\}_{i=0}^m$ of Lyapunov functions such that*

$$\forall i \in \{0,\dots,m-1\}, \quad \begin{cases} \forall x \in S, & V_{i+1}(x) \le V_i(x) \\ \forall x \in S^c, & (PV_i)(x) \le V_i(x) - V_{i+1}(x). \end{cases} \tag{66}$$

Then P satisfies Conditions (65) (so (45)) w.r.t. $\{V_i\}_{i=0}^m$, so that all the assertions of Theorem 5.2 and Corollary 5.1 hold (recall that (26) is an alternative bound to (22b) in the atomic case).

Proof. Prove that (66) implies (65) so that Corollary 5.2 applies. First note that the second condition of (66) ensures that $V_{i+1} \le V_i$ on S^c, thus the first condition in (65) is satisfied. The second condition in (66) and (65) are the same. Finally, for every $i = 0,\dots,m-1$ and every $x \in S$ we have $PV_i(x) = \nu(V_i)$ since S in an atom, so that

$$\forall x \in S, \quad PV_i(x) - \big(V_i(x) - V_{i+1}(x)\big) - \nu(V_i) = V_{i+1}(x) - V_i(x) \le 0.$$

This proves the third condition of (65).

□

Now we apply the two previous corollaries under the following subgeometric drift condition

$$\exists \alpha \in [0,1), \exists c_1 > 0,\ \forall x \in S^c, \quad (PV)(x) \le V(x) - c_1\, V(x)^\alpha \tag{$\boldsymbol{Sub_{\alpha,S^c}}$}$$

where V is some Lyapunov function. We begin with the atomic case. For any $\alpha \in [0,1)$ define the integer $m \equiv m_\alpha \ge 1$ by

$$m := \big\lfloor (1-\alpha)^{-1} \big\rfloor. \tag{67}$$

Corollary 5.4 (Atomic case) *Let P be a Markov kernel on $(\mathbb{X},\mathcal{X})$ satisfying Conditions* (**S**) *and* ($\boldsymbol{Sub_{\alpha,S^c}}$) *with an atom S and with $\nu(\cdot)$ defined by $\nu(\cdot) := P(a_0,\cdot)$ for $a_0 \in S$. Assume that PV is bounded on S. Then all the assertions of Theorem 5.2 and Corollary 5.1 hold with the positive integer $m \equiv m(\varepsilon,\alpha,\eta_0)$ defined in* (67) *and with functions $V_0, V_1, \dots, V_m$ specified in the proof.*

To prove Corollary 5.4 we use the following lemma which is based on [JR02, Lem. 3.5].

Lemma 5.2 *Let $S \in \mathcal{X}$, and let W be a Lyapunov function such that PW is bounded on S. Let $0 < \theta_2 < \theta_1 < 1$ be such that*

$$\exists c > 0,\ \forall x \in S^c, \quad (PW^{\theta_1})(x) \leq W(x)^{\theta_1} - c\, W(x)^{\theta_2}.$$

Then

$$\exists c' > 0,\ \forall x \in S^c, \quad (PW^{\theta_2})(x) \leq W(x)^{\theta_2} - c'\, W(x)^{\theta_3} \qquad \textit{with} \quad \theta_3 = 2\theta_2 - \theta_1.$$

Proof. The hypothesis of Lemma 5.2 writes as $PW^{\theta_1} \leq W^{\theta_1} - c\left(W^{\theta_1}\right)^{\theta_2/\theta_1}$ on S^c. It follows from [JR02, Lem. 3.5] that

$$\forall \eta \in (0,1],\ \exists c' > 0, \quad PW^{\eta\theta_1} \leq W^{\eta\theta_1} - c'\left(W^{\theta_1}\right)^{\frac{\theta_2}{\theta_1}+\eta-1} \quad \text{on } S^c.$$

Setting $\eta := \theta_2/\theta_1$ this gives

$$PW^{\theta_2} \leq W^{\theta_2} - c'\, W^{2\theta_2-\theta_1} \quad \text{on } S^c.$$

$\square$

Proof of Corollary 5.4. Note that if $\theta_2 = 0$ then Lemma 5.2 does not apply since $P1_{\mathbb{X}} = 1_X$ and this would give $c' = 0$. Let $\alpha_1 := 1 - 1/m \in [0,1)$ with m given in (67). Note that $\alpha_1 \leq \alpha$. Then it follows from ($\boldsymbol{Sub_{\alpha,S^c}}$) that

$$PV \leq V - c_1\, V^{\alpha_1} \quad \text{on } S^c. \tag{68}$$

Note that we can choose $c_1 < 1$ in (68).

- If $\alpha_1 = 0$ (i.e. $\alpha \in [0, 1/2)$), then Conditions (66) of Corollary 5.3 hold with $m = 1, V_0 = c_1^{-1}V$ and $V_1 = 1_{\mathbb{X}}$. Note that $1_{\mathbb{X}} = V_1 \leq V_0$.
- If $\alpha_1 = 1/2$ (i.e. $\alpha \in [1/2, 2/3)$), then we deduce from (68) and Lemma 5.2 that

  $$\exists c_2 > 0, \quad PV^{\alpha_1} \leq V^{\alpha_1} - c_2\, V^{\alpha_2} \quad \text{on } S^c \tag{69}$$

  with $\alpha_2 := 2\alpha_1 - 1 = 0$. Again note that we can choose $c_2 < 1$. Then the procedure stops, and Conditions (66) of Corollary 5.3 hold with $m = 2, V_0 = c_1^{-1}c_2^{-1}V$, $V_1 = c_2^{-1}V^{\alpha_1}$ and $V_2 = 1_{\mathbb{X}}$. Note that $1_{\mathbb{X}} = V_2 \leq V_1 \leq V_0$.
- If $\alpha_1 > 1/2$, then Lemma 5.2 can be repeated recursively to provide inequalities of the form $PV^{\alpha_{i-1}} \leq V^{\alpha_{i-1}} - c_i\, V^{\alpha_i}$ on S^c with $c_i < 1$ and

  $$\alpha_i = 2\alpha_{i-1} - \alpha_{i-2} = (\alpha_1 - 1)\, i + 1.$$

  Actually Lemma 5.2 can only be repeated until the value $i = m$ since $\alpha_m = 0$ and $\alpha_i < 0$ for $i > m$. Then Assumptions (66) of Corollary 5.3 hold with

  $$V_0 = \Big[\prod_{k=1}^{m} c_k\Big]^{-1} V, \quad \forall 1 \leq i \leq m-1: \ V_i = \Big[\prod_{k=i+1}^{m} c_k\Big]^{-1} V^{\alpha_i}, \quad V_m = 1_{\mathbb{X}}.$$

  Note that $1_{\mathbb{X}} = V_m \leq \cdots \leq V_0$.

Then the conclusions of Corollary 5.4 follows from Corollary 5.3.

□

Now we consider the general case where P satisfies Conditions (**S**) and ($\boldsymbol{Sub_{\alpha,S^c}}$). Using Corollary 5.2, we prove that the procedure in the atomic case (Corollary 5.4) extends to the non atomic case provided that Condition ($\boldsymbol{Sub_{\alpha,S^c}}$) can be modified thanks to Lemma 5.2 in order to fulfil the third condition in (65). To that effect, assume that PV is bounded on S. Then

$$\forall \varepsilon \in (0, \nu(1_{\mathbb{X}})),\ \exists \eta_0 \equiv \eta_0(\varepsilon) \in (0,1],\ \forall \eta \in (0,\eta_0],\ \forall x \in S, \quad (PV^\eta)(x) \le V(x)^\eta + \nu(V^\eta) - \varepsilon. \tag{70}$$

Indeed we have

$$\forall x \in S, \quad (PV^\eta)(x) - V(x)^\eta - \nu(V^\eta) \le (\sup_S PV)^\eta - 1 - \nu(1_{\mathbb{X}})$$

from Jensen's inequality and $1_{\mathbb{X}} \le V^\eta$. Then (70) follows from the following property

$$\exists \eta_0 \in (0,1],\ \forall \eta \in (0,\eta_0], \quad (\sup_S PV)^\eta - 1 \le \nu(1_{\mathbb{X}}) - \varepsilon$$

which holds since $(\sup_S PV)^\eta \to 1$ when $\eta \to 0$. Next, if $\eta_0 \ge 1-\alpha$, define the positive integer $m \equiv m(\varepsilon, \alpha, \eta_0)$ as follows

$$m := \left\lfloor \frac{\eta_0}{1-\alpha} \right\rfloor. \tag{71}$$

Corollary 5.5 *Assume that P satisfies Conditions* (**S**) *and* ($\boldsymbol{Sub_{\alpha,S^c}}$) *for some $S \in \mathcal{X}$, $\nu \in \mathcal{M}_*^+$, $\alpha \in [0,1)$ and for some Lyapunov function V. Moreover assume that V and PV are bounded on S. Let $\varepsilon \in (0, \nu(1_{\mathbb{X}}))$, and assume that the real number η_0 given in (70) is such that $\eta_0 \ge 1-\alpha$. Then all the assertions of Theorem 5.2 and Corollary 5.1 hold with the positive integer $m \equiv m(\varepsilon, \alpha, \eta_0)$ defined in* (71) *and with functions $V_0, V_1, \ldots, V_m$ specified in the proof.*

Proof. Note that the third condition in (65) associated with Condition ($\boldsymbol{Sub_{\alpha,S^c}}$) may fail, that is the inequality $PV \le V - c_1 V^\alpha + \nu(V)$ on S may be false. To initialize the procedure, apply [JR02, Lem. 3.5] from ($\boldsymbol{Sub_{\alpha,S^c}}$) with the exponent η_0 given in (70), that is:

$$\exists c_{\eta_0} > 0,\ \forall x \in S^c, \quad (PV^{\eta_0})(x) \le V(x)^{\eta_0} - c_{\eta_0} V(x)^{\alpha+\eta_0-1}. \tag{72}$$

If $\alpha + \eta_0 - 1 < 0$, then Inequality (72) cannot be used to apply Corollary 5.2 since the function V_1 in Conditions (65) must take its values in $[a, +\infty)$ for some $a > 0$. Now assume that $\alpha + \eta_0 - 1 \ge 0$ and prove that the third condition in (65) associated with (72) is satisfied. Let $M_1 := \sup_S V$ and $M_2 := \sup_S PV$. Recall that $\varepsilon \in (0, \nu(1_{\mathbb{X}}))$ and note that c_{η_0} in (72) can be chosen such that $c_{\eta_0} M_1^{\alpha+\eta_0-1} \le \varepsilon$ (up to reduce the value of c_{η_0}). Then we have from (70)

$$\forall x \in S, \quad (PV^{\eta_0})(x) - V(x)^{\eta_0} + c_{\eta_0} V(x)^{\alpha+\eta_0-1} - \nu(V^{\eta_0}) \le 0 \tag{73}$$

Now, starting from (72)-(73), iterate Lemma 5.2 as many times as possible. Namely, let

$$\widehat{V} := V^{\eta_0} \quad \text{and} \quad \widehat{\alpha}_1 := 1 - \frac{1}{m}$$

with m defined in (71). Note that $m = \lfloor (1-\widehat{\alpha})^{-1} \rfloor$ with $\widehat{\alpha} = 1-(1-\alpha)/\eta_0$, and that $\widehat{\alpha}_1 \le \widehat{\alpha}$. Also set $\widehat{c}_1 = c_{\eta_0}$. Then

$$\begin{cases} \forall x \in S^c, & (P\widehat{V})(x) \le \widehat{V}(x) - \widehat{c}_1\, \widehat{V}(x)^{\widehat{\alpha}_1} \\ \forall x \in S, & (P\widehat{V})(x) \le \widehat{V}(x) - \widehat{c}_1\, \widehat{V}(x)^{\widehat{\alpha}_1} + \nu(\widehat{V}) \end{cases} \tag{74}$$

from (72)-(73) and $\widehat{\alpha}_1 \le \widehat{\alpha}$. Then, starting from (74) and iterating Lemma 5.2, we can proceed exactly as in the proof of Corollary 5.4, provided that the third condition in (65) holds at each step (this was automatically fulfilled in the atomic case). More precisely, at each step, Lemma 5.2 provides an inequality of the form

$$P\widehat{V}^{\widehat{\alpha}_{i-1}} \le \widehat{V}^{\widehat{\alpha}_{i-1}} - \widehat{c}_i\, \widehat{V}^{\widehat{\alpha}_i} \quad \text{on } S^c \tag{75}$$

with some $\widehat{c}_i > 0$ and with

$$\widehat{\alpha}_i = 2\widehat{\alpha}_{i-1} - \widehat{\alpha}_{i-2} = (\widehat{\alpha}_1 - 1)\, i + 1.$$

This procedure can be repeated only until the value $i = m$ since $\widehat{\alpha}_m = 0$ and $\widehat{\alpha}_i < 0$ for $i > m$, but we have moreover to check that the third condition in (65) associated with (75) holds. To verify this last point, note that $\widehat{\alpha}_{i-1} \le 1$ and that

$$P\widehat{V}^{\widehat{\alpha}_{i-1}} - \widehat{V}^{\widehat{\alpha}_{i-1}} - \nu(\widehat{V}^{\widehat{\alpha}_{i-1}}) = PV^{\eta_i} - V^{\eta_i} - \nu(V^{\eta_i}) \quad \text{with} \quad \eta_i := \eta_0\widehat{\alpha}_{i-1} \in (0, \eta_0]$$

from $\widehat{V} := V^{\eta_0}$. It then follows from (70) and $\widehat{V}^{\widehat{\alpha}_{i-1}} \ge 1_{\mathbb{X}}$ that

$$\forall x \in S, \quad (P\widehat{V}^{\widehat{\alpha}_{i-1}})(x) - \widehat{V}^{\widehat{\alpha}_{i-1}}(x) + \widehat{c}_i\, \widehat{V}^{\widehat{\alpha}_i}(x) - \nu(\widehat{V}^{\widehat{\alpha}_{i-1}}) \le \widehat{c}_i\, V^{\eta_0}(x) - \varepsilon \le 0 \tag{76}$$

since $\widehat{c}_i$ in (75) can be chosen such that $\widehat{c}_i M_1^{\eta_0} \le \varepsilon$ (recall that $M_1 := \sup_S V$). Then Conditions (65) of Corollary 5.2 hold with

$$V_0 = (\widehat{c}_1\widehat{c}_2\cdots\widehat{c}_m)^{-1}\widehat{V},\ V_1 = (\widehat{c}_2\cdots\widehat{c}_m)^{-1}\,\widehat{V}^{\alpha_1}, \ldots, V_{m-1} = \widehat{c}_m^{\,-1}\widehat{V}^{\alpha_{m-1}},\ V_m = 1_{\mathbb{X}}$$

(note that $1_{\mathbb{X}} = V_m \le \cdots \le V_0$). Then the conclusions of Corollary 5.5 follow from Corollary 5.2 □

Remark 5.1 *In practice, for the choice of $\varepsilon \in (0, \nu(1_{\mathbb{X}}))$ in Corollary 5.5, a trade-off must be made with respect to Condition (70) versus the resulting positive constant $\nu(V_0)$ and $\mu(V_0)$ in (58) and (60). Indeed, the smaller ε is, the larger η_0 in (70) will be, so the larger m in (71) will be. However, the smaller ε is, the larger $[\prod_{i=1}^m \widehat{c}_i]^{-1}$ will be in the above definition of V_0, so that the larger constants $\nu(V_0)$ and $\mu(V_0)$ in (58) and (60) will be.*

The following proposition shows that Condition (70) can be simplified under general conditions on $\mathbb{X}$, S, P and V.

Proposition 5.3 *Assume that any one of the two following conditions holds:*

(a) $\mathbb{X}$ is discrete and S is finite.

(b) $\mathbb{X}$ is a metric space, S is compact and the functions V and PV^η ($\forall \eta \in (0, 1]$) are continuous on S.

Then Corollary 5.5 applies with Condition (70) replaced by the following simpler one

$$\exists \eta_0 \in (0,1],\ \forall x \in S, \quad (PV^{\eta_0})(x) < V(x)^{\eta_0} + \nu(V^{\eta_0}). \tag{77}$$

Proof. First observe that the proof of Corollary 5.5 is still valid when Condition (76) holds with some $\varepsilon_i > 0$ for $i = 1, \ldots, m$ (in place of $\varepsilon > 0$): then $\widehat{c}_i$ in (75) has to be chosen such that $\widehat{c}_i M_1^{\eta_0} \leq \varepsilon_i$, and the function V_0 is defined as in the previous proof from such $\widehat{c}_i$. Consequently, under the conditions (a) or (b), we have to prove that (77) implies that

$$\forall \eta \in (0, \eta_0],\ \forall x \in S, \qquad (PV^{\eta})(x) < V(x)^{\eta} + \nu(V^{\eta}). \tag{78}$$

We use the notations introduced in the proof of Proposition 4.1. Recall that, for any $x \in S$, $\sigma_x(\cdot) = P(x, \cdot) - \nu(\cdot)$ is a non-negative measure on $(\mathbb{X}, \mathcal{X})$ from Assumption (**S**), and that $\sigma_x(1_{\mathbb{X}}) = 1 - \nu(1_{\mathbb{X}})$ does not depend on x. We set $\sigma := 1 - \nu(1_{\mathbb{X}})$. If $\sigma = 0$ (atomic case), then σ_x is null, thus for every $\eta \in (0, 1]$ we have $(PV^{\eta})(x) = \nu(V^{\eta})$, so that (78) is obvious. Now assume that $\sigma > 0$, and note that (78) is equivalent to

$$\forall \eta \in (0, \eta_0],\ \forall x \in S, \quad \sigma_x(V^{\eta}) < V(x)^{\eta}.$$

Define the following probability measure on $(\mathbb{X}, \mathcal{X})$: $\widetilde{\sigma}_x(\cdot) = \sigma_x(\cdot)/\sigma$. Let us prove that

$$\forall \eta \in (0, \eta_0),\ \forall x \in S, \quad \sigma_x(V^{\eta_0}) < V(x)^{\eta_0} \implies \sigma_x(V^{\eta}) < V(x)^{\eta}. \tag{79}$$

Assume that $\sigma_x(V^{\eta_0}) < V(x)^{\eta_0}$. It follows from Jensen's inequality that

$$\widetilde{\sigma}_x(V^{\eta}) = \widetilde{\sigma}_x\big((V^{\eta_0})^{\eta/\eta_0}\big) \leq \big[\widetilde{\sigma}_x(V^{\eta_0})\big]^{\eta/\eta_0}.$$

Then we deduce from the definition of $\widetilde{\sigma}_x$ and from $\sigma_x(V^{\eta_0}) < V(x)^{\eta_0}$ that

$$\frac{\sigma_x(V^{\eta})}{\sigma} \leq \frac{\big(\sigma_x(V^{\eta_0})\big)^{\eta/\eta_0}}{\sigma^{\eta/\eta_0}} < \frac{V(x)^{\eta}}{\sigma^{\eta/\eta_0}}$$

hence

$$\sigma_x(V^{\eta}) < \frac{\sigma}{\sigma^{\eta/\eta_0}}\, V(x)^{\eta} < V(x)^{\eta}$$

since $0 < \sigma < 1$ and $0 < \eta/\eta_0 < 1$. This proves (79). Therefore (77) implies (78).

□

We conclude this section by presenting a result on the approximation of P^n by the submarovian kernel T_n given in (13a) under the subgeometric drift conditions (45).

Theorem 5.3 *Assume that P satisfies the assumptions of Theorem 5.1 with some $m \geq 1$. Then we have for every measurable function $f : \mathbb{X} \to \mathbb{R}$ such that $|f| \leq V_m$:*

$$\forall x \in \mathbb{X}, \quad \sum_{k=0}^{+\infty} (k+1)^{m-1} \big|(P^k f)(x) - (T_k f)(x)\big| \leq \vartheta_{m-1} V_0(x) \tag{80}$$

with ϑ_{m-1} defined in (47).

Proof. If $m=1$, then (13a) and the positivity of $P-T$ give for every measurable function $f:\mathbb{X}\to\mathbb{R}$ such that $|f|\le V_1$ (see the proof of Theorem 5.1)

$$\sum_{k=0}^{+\infty}\big|(P^kf)(x)-(T_kf)(x)\big|\le\sum_{k=0}^{+\infty}\big((P-T)^k|f|\big)(x)\le\sum_{k=0}^{+\infty}\big((P-T)^kV_1\big)(x)\le V_0(x).$$

This proves (80) for $m=1$. Inequality (80) for $m\ge 1$ easily follows by induction from the following fact: if $(P-T)V_m\le V_m-V_{m+1}$, then we have

$$\forall k\ge 0,\quad (P-T)^kV_{m+1}\le (P-T)^kV_m-(P-T)^{k+1}V_m,$$

from which we deduce that for every measurable function $f:\mathbb{X}\to\mathbb{R}$ such that $|f|\le V_{m+1}$ (see the proof of Theorem 5.1)

$$\begin{aligned}\sum_{k=0}^{+\infty}(k+1)^m\left|P^kf-T_kf\right| &\le \sum_{k=0}^{+\infty}(k+1)^m\,(P-T)^kV_{m+1}\\ &\le \sum_{j=0}^{m-1}C_m^j\sum_{k=0}^{+\infty}k^j\,(P-T)^kV_m\\ &\le \sum_{j=0}^{m-1}C_m^j\sum_{k=0}^{+\infty}(k+1)^j\,(P-T)^kV_{j+1}\\ &\le \left(\sum_{j=0}^{m-1}C_m^j\vartheta_j\right)V_0 \qquad \text{(from induction hypothesis)}\\ &\le \vartheta_mV_0.\end{aligned}$$

□

6 Examples

Excepted in Subsection 6.1.1, the focus is on standard non-atomic examples from the literature on the rate of convergence of Markov chains. Moreover, in Subsections 6.3-6.4, we only deal with the error bound in total variation distance (24) to make comparison easier with previous works.

6.1 Birth-and-Death Markov chains (geometric case)

6.1.1 Atomic case

Let us introduce the following example with $\mathbb{X}=\mathbb{N}$ and a transition kernel P specified by

$$\begin{gathered}a:=P(0,0)\in(0,1),P(0,1)=1-a\\ \forall n\ge 1,\ P(n,n-1):=p,\quad P(n,n):=r,\quad P(n,n+1):=q\\ \text{with } p,q,r\in[0,1] \text{ such that } p+r+q=1 \text{ and } 0<q<p.\end{gathered}\tag{81}$$

Set

$$\delta:=r+2\sqrt{pq}=1-(\sqrt{p}-\sqrt{q})^2\in(0,1),\ \gamma:=\sqrt{p/q}\in(1,+\infty)\ \text{ and }\ V\equiv V_\gamma:=\{\gamma^n\}_{n\in\mathbb{N}}.$$

Then P satisfies Conditions (**S**) and *($\boldsymbol{D_{S^c}}$)* with the atom $S = \{0\}$ and with $\nu := P(0, \cdot)$ (see [HL14, Prop. 4.1]). We deduce from Corollary 4.1 that Estimates (22a)-(22b) (also see (26)) hold with $W := V$ and

$$\mu(V) \leq \frac{a + (1-a)\gamma}{1-\delta} \quad \text{and} \quad \forall n \geq 1,\ \varepsilon_n \leq \varepsilon_{n,V} \leq \frac{a + (1-a)\gamma}{1-\delta}\, \delta^n. \tag{82}$$

If $r := 0$, $a := p$, the rate $\delta = 2\sqrt{pq}$ is well-known for the V-weighted operator norm of $P^n - \pi(\cdot)1_{\mathbb{X}}$, which is related to the V-geometrical ergodicity of P, see [MT94, LT96, Bax05, 8.1 and 8.4]. The simple and explicit error bound (82) holds for any Birth-and-Death Markov chain satisfying (81). If $r > 0$, then (82) provides a rate of convergence in Estimates (22a)-(22b) with $W := V$, which may be better than the rate of convergence of the V-weighted operator norm of $P^n - \pi(\cdot)1_{\mathbb{X}}$, see [HL14, Prop. 4.1, (4.9a)]. This is due to the existence of eigenvalues λ of P such that $\delta < |\lambda| < 1$ (see Remark 4.4). Note that this fact is not inconsistent since the stationary distribution π is approximated in two different ways.

It turns out that, in the geometric and atomic case, the approximation of π in (22a)-(22b) and in (26) erases the effect of possible eigenvalues λ of P such that $\delta < |\lambda| < 1$, where δ is the real number in *($\boldsymbol{D_{S^c}}$)*. By contrast, the next non-atomic instances show that the rate of convergence in (22a)-(22b) may be only $\mathrm{O}(\delta^{\alpha_0 n})$ for some $\alpha_0 \in (0,1)$ rather than $\mathrm{O}(\delta^n)$ (see Corollary 4.2): this could correspond to the case when P admits some eigenvalues λ in the annulus $\{z \in \mathbb{C} : \delta < |z| < 1\}$ (e.g. see the end of Subsection 6.1.2).

6.1.2 Non atomic case

Now, assume that P is specified by

$$\begin{gathered} a := P(0,0) = P(1,0) \in (0,1) \text{ and } P(0,1) = P(1,2) = 1-a, \\ \forall n \geq 2,\ P(n,n-1) := p,\ P(n,n) := r,\ P(n,n+1) := q \quad \text{with } 0 < q < p \text{ and } p + q \leq 1. \end{gathered}$$

Set $S := \{0,1\}$ which is not an atom. Condition (**S**) is satisfied with $\nu := a\delta_0$, and P satisfies ($\boldsymbol{D_{S^c}}$) with $\delta := r + 2\sqrt{pq} \in (0,1)$ and with $V := \{\gamma^n\}_{n\in\mathbb{N}}$ (e.g. see [HL14, Sect. 4.1]). Let us illustrate Corollary 4.2 in this case. Note that: $\forall \alpha \in (0,1]$, $\forall n \in \mathbb{N}$, $V(n)^\alpha = \gamma^{\alpha n}$.

1. Condition (35) holds with $\alpha_0 = 1$ if and only if $1 - a \leq \delta\gamma^{-1}$. Consequently, under this last condition, Corollary 4.2 applies with $\alpha_0 = 1$ (that is Theorem 4.2).

2. When p, q, r are chosen such that $1 - a > \delta\gamma^{-1}$, (35) is fulfilled with $\alpha_0 < 1$. Indeed, we have
$$\sup_{i=0,1} \left[(PV^\alpha)(i) - \delta^\alpha V(i)^\alpha\right] - \nu(V^\alpha) \leq 0 \iff \alpha \leq \frac{\ln(1-a)}{\ln(\delta\gamma^{-1})}.$$
Thus Corollary 4.2 applies with $\alpha_0 = \ln(1-a)/\ln(\delta\gamma^{-1})$. When $1 - a = P(1,2) \to 1$, we obtain that $\alpha_0 \to 0$, so that the rate δ^{α_0} of Corollary 4.2 converges to 1. This comment should be compared with that of [HL14, Sect. 4.2]) on a very similar model (the two models only differ on the fact that $a_{-2} := P(n, n-2) > 0$ for $n \geq 2$) for which the so-called second eigenvalue related to the V-geometrical ergodicity of P tends to one when $P(1,2) \to 1$, i.e. the spectral gap tends to zero, see [HL14, Table 1] for details.

6.2 Random walk on a half line (subgeometric case)

Let $(X_n)_{n\geq 0}$ be the so-called $\mathbb{N}$-valued random walk on a half line, defined as follow:

$$\forall n \geq 1, \quad X_n = \max\big(0, X_{n-1} + W_n\big) \tag{83}$$

where X_0 is an $\mathbb{N}$-valued random variable (r.v.) and where $(W_n)_{n\geq 0}$ is an independent and identically distributed sequence of $\mathbb{Z}$-valued r.v., assumed to be independent of X_0. The common probability distribution of $(W_n)_{n\geq 0}$ is denoted by: $\forall j \in \mathbb{Z},\ p_j := \mathbb{P}(W_1 = j)$. Thus, $(X_n)_{n\geq 0}$ is a Markov chain with transition kernel P on $\mathbb{X} = \mathbb{N}$ given by

$$\begin{aligned}\forall f \in \mathcal{B}, \forall i \in \mathbb{N}, \quad (Pf)(i) &= \sum_{j\in\mathbb{Z}} f\big(\max(0, i+j)\big)\, p_j = f(0)\mathbb{P}(W < -i) + \sum_{j\geq -i} f(i+j)\, p_j \\ &= f(0)\mathbb{P}(W < -i) + \sum_{j\geq 0} f(j)\, p_{j-i}.\end{aligned}$$

We assume that $p_j > 0$ for at least one $j \geq 1$, so that $\mathbb{P}(W > 0) > 0$. For every finite set $S \subset \mathbb{N}$, P satisfies Condition $(\mathbf{S})$ with

$$\nu := \sum_{j=0}^{+\infty} \nu_j\, \delta_j \quad \text{with} \quad \nu_j := \min_{i\in S} p_{j-i} \tag{84}$$

provided that $\nu(1_{\mathbb{N}}) > 0$. Let $m_0 \geq 2$ be any integer. For every $q \in \{1, \ldots, m_0\}$ we define

$$\forall i \in \mathbb{N}, \quad V_q(i) = (1+i)^q \tag{85}$$

and we simply write V for V_{m_0}. Let us assume that W has a moment of order m_0 and has a negative expectation, that is:

$$\mathbb{E}[\,|W^{m_0}|\,] = \sum_{j\in\mathbb{Z}} |j|^{m_0}\, p_j \ < \infty \qquad \text{and} \qquad \mathbb{E}[W] = \sum_{j\in\mathbb{Z}} j\, p_j \ < 0. \tag{86}$$

Under these moment conditions (86), P satisfies Condition $(\boldsymbol{Sub_{\alpha,S^c}})$ with $V := V_{m_0}$, $\alpha = 1 - 1/m_0$ for some finite set $S \subset \mathbb{N}$, e.g. see [DMPS18]. Therefore, to apply Corollary 5.5, we have to find $\eta_0 \in (0, 1]$ such that (see Proposition 5.3)

$$\forall i \in S, \quad (PV^{\eta_0})(i) < V(i)^{\eta_0} + \nu(V^{\eta_0}) = (1+i)^{m_0\eta_0} + \sum_{j=0}^{+\infty} \nu_j\, (1+j)^{m_0\eta_0}. \tag{87}$$

For a given probability distribution $\{p_j\}_{j\in\mathbb{Z}}$, the study of the numeric function ψ defined by

$$\forall \eta \in [0, 1], \quad \psi(\eta) := \max_{i\in S} \big[(PV^\eta)(i) - V(i)^\eta\big] - \nu(V^\eta)$$

gives η_0 satisfying (87). If $\eta_0 \geq 1/m_0$, then the assertions of Theorem 5.2 and Corollary 5.1 apply with $m = \lfloor \eta_0 m_0 \rfloor$ due to Corollary 5.5.

In the next proposition we present additional assumptions under which the assumptions of Corollary 5.5 are fulfilled.

Proposition 6.1 *Assume that moment conditions (86) holds with $m_0 = 3$. Let $\tau \in (0,1)$ and let $s \equiv s(\tau) \geq 2$ be an integer such that*

$$\forall i > s, \quad \sum_{j \geq -i} j p_j \leq \tau\,\mathbb{E}[W] \quad \text{and} \quad \frac{\mathbb{P}(W < -i)}{(1+i)^{m_0-1}} + \sum_{j\geq -i}\sum_{k=2}^{m_0} \frac{C_{m_0}^k j^k p_j}{(1+i)^{k-1}} \;\leq\; -\,\frac{\tau\, m_0\,\mathbb{E}[W]}{2}. \quad (88)$$

Moreover assume that

$$\{p_j\}_{j=0}^{+\infty} \text{ is non-increasing and: } \quad \forall j \in \{-s,\dots,-1\},\ p_j \geq p_0. \quad (89)$$

Then the following assertions hold with $S := \{0,\dots,s\}$ and $\nu(\cdot)$ given by

$$\nu := \sum_{j\geq 0} p_j\,\delta_j. \quad (90)$$

(i) If $\mathbb{E}[W^2] \leq \min_{i\in S\setminus\{0\}} \big[-(2+i)\,\mathbb{E}[W] + \frac{\mathbb{P}(W\geq 0)}{3i}\big]$, then the assertions of Theorem 5.2 and Corollary 5.1 apply with $m = 3$.

(ii) If $\forall i \in i \in S\setminus\{0\}$, $C_2(i) - \mathbb{P}(W\geq 0) < -2\,i\,\mathbb{E}[W]$, where

$$\forall i \in S\setminus\{0\}, \quad C_2(i) := -\,\mathbb{P}\big(-i \leq W < 0\big) + \sum_{j=-\infty}^{-2} p_j\,(1+j)^2 - \sum_{j=-\infty}^{-i-2} p_j\,\big(1+i+j\big)^2,$$

then the assertions of Theorem 5.2 and Corollary 5.1 apply with $m = 2$.

For every $\tau \in (0,1)$, the inequalities in (88) hold for i large enough since $\sum_{j\geq -i} jp_j \to E[W]$ when $i \to +\infty$ and the other term in (88) converges to 0 when $i\to+\infty$ (note that $-\tau\, m_0\,\mathbb{E}[W] > 0$). Assumption (89) is introduced in order to rewrite the positive measure ν in (84) as in (90), so that

$$\forall q = 1,\dots,m_0, \quad (PV_q)(0) - \nu(V_q) = \mathbb{P}(W < 0). \quad (91)$$

Also note that, although Condition (89) is restrictive, it is not inconsistent with the condition $\mathbb{E}[W] < 0$. The condition in (i) is satisfied when $\mathbb{E}[W^2] \leq -\,3\,\mathbb{E}[W]$.

The proof of Proposition 6.1 is based on the two following lemmas proved in Annex B. Note that these two lemmas may be relevant under alternative assumptions on $\{p_j\}_{j\in\mathbb{Z}}$ since Condition (89) is not assumed.

Lemma 6.2 *Assume that moment conditions (86) hold for some $m_0 \geq 2$. Let $\tau \in (0,1)$ and let $s \geq 2$ the integer in (88). Set $S := \{0,\dots,s\}$. Then we have (with $V := V_{m_0}$)*

$$\forall i \in S^c, \quad (PV)(i) \leq V(i) - c\,V(i)^{1-\frac{1}{m_0}} \qquad \text{with } \ c := -\,\frac{\tau\, m_0\,\mathbb{E}[W]}{2} > 0.$$

Lemma 6.3 *Assume that moment Conditions (86) hold for some $m_0 \geq 2$. Then we have for every $q \in \{1,\dots,m_0\}$*

$$\forall i \in \mathbb{N}, \qquad (PV_q)(i) = (PV_q)(0) + \mathbb{E}\big[(1+i+W)^q\big] - \mathbb{E}\big[(1+W)^q\big] + C_q(i) \qquad (92)$$

with V_q defined in (85), with $\nu(\cdot)$ given in (84), and with $C_q(i)$ defined by $C_q(0) := 0$ and

$$\forall i \in \mathbb{N}\setminus\{0\}, \quad C_q(i) := -\,\mathbb{P}\big(-i \leq W < 0\big) + \sum_{j=-\infty}^{-2} p_j\,(1+j)^q - \sum_{j=-\infty}^{-i-2} p_j\,\big(1+i+j\big)^q.$$

Proof of Proposition 6.1. Recall that, in the present context, we just have to find $\eta_0 \in (0,1]$ such that Condition (87) holds with $S := \{0,\dots,s\}$ and $\nu(\cdot)$ given in (90).

Under the assumption of Assertion (i), we prove that Inequality (87) holds with $\eta_0 = 1$. Indeed, apply (92) with $q = m_0 = 3$. First note that for every $i \in S \setminus \{0\}$ we have $C_3(i) < 0$ since

$$\forall j \leq -i-2, \quad 1+j \leq 1+i+j \qquad \text{and} \qquad \forall j \in (-i-2,-2], \quad 1+j \leq 0.$$

Recall that $C_3(0) = 0$. Moreover we have

$$\mathbb{E}\big[(1+i+W)^3\big] - \mathbb{E}\big[(1+W)^3\big] - (1+i)^3 = -1 + 3\,i\,(2+i)\,\mathbb{E}[W] + 3\,i\,\mathbb{E}[W^2].$$

Hence it follows from (92) and from (91) that for every $i \in S \setminus \{0\}$

$$\begin{aligned}(PV_3)(i) - V_3(i) - \nu(V_3) &= \mathbb{P}(W<0) - 1 + 3\,i\,(2+i)\,\mathbb{E}[W] + 3\,i\,\mathbb{E}[W^2] + C_3(i) \\ &< \mathbb{P}(W<0) - 1 + 3\,i\,(2+i)\,\mathbb{E}[W] + 3\,i\,\mathbb{E}[W^2] \qquad (93)\end{aligned}$$

since $C_3(i) < 0$. Finally $(PV_3)(0) - V_3(0) - \nu(V_3) = \mathbb{P}(W<0) - 1 < 0$, and for every $i \in S \setminus \{0\}$ we have $(PV_3)(i) - V_3(i) - \nu(V_3) < 0$ under the condition of Assertion (i). Thus, Inequality (87) holds with $V := V_3$ and $\eta_0 = 1$. Thus the assumptions of Corollary 5.5 hold with $m = 3$.

Now prove Assertion (ii). Apply (92) with $q = 2$. We have

$$\mathbb{E}\big[(1+i+W)^2\big] - \mathbb{E}\big[(1+W)^2\big] - (1+i)^2 = -1 + 2i\mathbb{E}[W].$$

It follows from (92) with $q = 2$ that

$$\forall i \in S, \quad (PV_2)(i) - V_2(i) - \nu(V_2) = \mathbb{P}(W<0) - 1 + 2i\mathbb{E}[W] + C_2(i). \qquad (94)$$

Thus $(PV_2)(0) - V_2(0) - \nu(V_2) = \mathbb{P}(W<0) - 1 < 0$. Moreover for every $i \in S \setminus \{0\}$ we have $(PV_2)(i) - V_2(i) - \nu(V_2) < 0$ under the condition of Assertion (ii). Thus, Property (87) holds with $\eta_0 = 2/3$ since ${V_3}^{2/3} = V_2$. Thus the assumptions of Corollary 5.5 hold with $m = \lfloor \eta_0 m_0 \rfloor = 2$. □

Remark 6.1 *Under the assumptions of Proposition 6.1, alternatives to Assertions (i) or (ii) can be obtained thanks to (79). For instance assume that $\mathbb{E}[W^2] \leq -4\mathbb{E}[W]$. Then we easily deduce from (93) that $(PV_3)(i) - V_3(i) - \nu(V_3) < 0$ for $i \in S \setminus \{1\}$. The last condition is not guaranteed for $i = 1$, but we know from (94) that $(PV_2)(1) - V_2(1) - \nu(V_2) < 0$ provided that $C_2(1) - \mathbb{P}(W \geq 0) < -2\,\mathbb{E}[W]$. Finally it follows from (79) applied with $V := V_3$, $\eta_0 = 1$ and $\eta = 2/3$ and from the above inequality concerning V_3 that we also have $(PV_2)(i) - V_2(i) - \nu(V_2) < 0$ for $i \in S \setminus \{1\}$. Consequently, if $\mathbb{E}[W^2] \leq -4\mathbb{E}[W]$ and if $C_2(1) - \mathbb{P}(W \geq 0) < -2\,\mathbb{E}[W]$, then the assumptions of Corollary 5.5 hold with $m = 2$.*

6.3 Metropolis-Hastings algorithm for the standard Gaussian distribution (geometric case)

We are interested in the $\mathbb{R}$-valued Markov chain used in the Metropolis-Hastings algorithm to simulate the standard Gaussian distribution $\pi := \mathcal{N}(0,1)$ (the target distribution) with the Gaussian distribution $\mathcal{N}(x,1)$ as the proposal distribution. The iterates of the transition kernel are used in [MT94, MT96, RT99, RT00],[Bax05, Sect. 8.2] to approximate π, while the ergodic averages are used in [RR97, RT99]. In this part, the focus is on geometric rate of convergence, so that we do not discuss the results in [RR97, RT99] with ergodic averages. The transition kernel P of the Metropolis-Hastings Markov chain is reversible and positive in the sense given in [Bax05, Th. 1.3]. But, here these additional properties are not used. We simply apply Theorem 4.2 or Corollary 4.2 according to whether $\alpha_0 = 1$ or $\alpha_0 < 1$ in (35).

Let $\mathbb{X} := \mathbb{R}$ and $r, d > 0$ be two positive scalars. Set $V_r(x) := e^{r|x|}$ for any $x \in \mathbb{X}$ and $S_d := [-d, d]$. Let us denote the function PV_r/V_r by $\lambda(\cdot, r)$. We know from the computation in [Bax05, Sect. 8.2] that P satisfies ($\boldsymbol{D_{S^c}}$) with

$$\begin{aligned}
\delta_{d,r} &:= \lambda(d,r) = \max_{|x|\geq d} \lambda(x,r) \\
\lambda(x,r) &= e^{r^2/2}\big[\,\Phi(-r) - \Phi(-r-x)\,\big] + \frac{1}{\sqrt{2}} e^{(x-r)^2/4}\,\Phi\big((r-x)/\sqrt{2}\,\big) + \\
&\quad e^{r^2/2-2rx}\big[\,\Phi(-x+r) - \Phi(-2x+r)\,\big] + \frac{1}{\sqrt{2}} e^{(x^2-6rx+r^2)/4}\,\Phi\big((r-3x)/\sqrt{2}\,\big) \\
&\quad + \Phi(0) + \Phi(-2x) - \frac{1}{\sqrt{2}} e^{x^2/4}\big[\,\Phi(-x/\sqrt{2}) + \Phi\big(-3x/\sqrt{2}\,\big)\,\big]
\end{aligned}$$

where Φ denotes the standard Gaussian distribution function. Moreover P satisfies ($\mathbf{S}$) with the minorization measure $\nu_d(dx) = \big(e^{-d^2}/\sqrt{2\pi}\big)e^{-x^2}\,1_{[-d,d]}(x)\,dx$ (see [Bax05, p. 727]). Note that

$$\nu_d(1_{\mathbb{X}}) = \sqrt{2}\,e^{-d^2}\big[\,\Phi(\sqrt{2}d) - \Phi(0)\,\big].$$

An easy computation gives

$$\nu_d(V_r) = \frac{\sqrt{2}\,e^{-d^2}}{\sqrt{\pi}} \int_0^d e^{rx-x^2}\,dx = \sqrt{2}\,e^{-d^2+r^2/4}\big[\,\Phi\big(\sqrt{2}(d-r/2)\big) - \Phi\big(-r/\sqrt{2}\,\big)\,\big].$$

Finally we deduce from [Bax05, p. 726] that Condition ($\boldsymbol{D_S}$) involves the following term

$$\max_{|x|\leq d}\big[(PV_r)(x) - \delta_{d,r}\,V_r(x)\big] - \nu_d(V_r) \;=\; PV_r(0) - \delta_{d,r}V_r(0) - \nu_d(V_r) = \lambda(0,r) - \delta_{d,r} - \nu_d(V_r).$$

The best rate of convergence in [Bax05, 5th line in Tab. 2] is obtained when $d := 1.1$ and $r := 0.16$. Thus, we get $\delta_{1.1,0.16} = 0.9744$ and $\nu_{1.1}(V_{0.16}) \approx 0.1997$. It follows that Condition ($\boldsymbol{D_S}$) holds since

$$\lambda(0, 0.16) - \delta_{1.1,0.16} - \nu_{1.1}(V_{0.16}) \approx 0.0942 - 0.1997 < 0.$$

Therefore, Theorem 4.2 applies and provides the following estimate from (24) and (32)

$$\forall n \geq 1, \quad \|\pi - \widetilde{\mu}_n\|_{TV} \leq 2\,\frac{\nu_{1.1}(V_{0.16})}{1-\delta_{1.1,0.16}} \times {\delta_{1.1,0.16}}^n \approx 15.61 \times 0.9744^n.$$

Note that the above rate 0.9744^n for the total variation norm is slightly better than 0.9747^n obtained in [Bax05, 5th line of Tab. 2] for the $V_{0.16}$-weighted operator norm of $P^n - \pi(\cdot)1_{\mathbb{X}}$, which is related to the $V_{0.16}$-geometrical ergodicity of P. Moreover the specific properties of the transition kernel P involved in [Bax05, Th. 1.3] are not used here. Anyway observe that the multiplicative constant 15.61 is not too large.

The rate of convergence reported in [RT99, Section 5] for $\|\pi - P^n(x,\cdot)\|_{TV}$ is obtained for $d := 1.15$ and $r := 0.48$. We get from these parameters

$$\delta_{1.15,0.48} \approx 0.9353, \quad \nu_{1.15}(1_{\mathbb{X}}) \approx 0.1688 \quad \text{and} \quad \nu_{1.15}(V_{0.48}) \approx 0.2131$$

so that Condition $(\boldsymbol{D_S})$ does not hold since $\lambda(0,0.48) - \delta_{1.15,0.48} - \nu_{1.15}(V_{0.48}) \approx 0.0925 > 0$. In such as case, we can apply Corollary 4.2. Indeed, we deduce from [Bax05, p. 726] that $PV_{0.48}$ is bounded on $[-1.15, 1.15]$

$$K_{1.15,0.48} := \sup_{|x|\le 1.15} (PV_{0.48})(x) = PV_{0.48}(1.15) = e^{1.15\times 0.48}\,\lambda(1.15, 0.48) \approx 1.6244.$$

Let $\sigma := 1 - \nu_{1.15}(1_{\mathbb{X}}) = 0.8312$. It follows from Proposition 4.1 that for every $\alpha \in (0,1]$ and for every $x \in [-1.15, 1.15]$

$$\begin{aligned}(P{V_{0.48}}^{\alpha})(x) - \nu_{1.15}({V_{0.48}}^{\alpha}) &\le \frac{\sigma}{\sigma^{\alpha}}\left[(PV_{0.48})(x) - \nu_{1.15}(V_{0.48})\right]^{\alpha} \\ &\le \frac{\sigma}{\sigma^{\alpha}}\left[K_{1.15,0.48} - \nu_{1.15}(V_{0.48})\right]^{\alpha}.\end{aligned}$$

Hence we have with $\alpha_0 = 0.31$

$$(P{V_{0.48}}^{\alpha_0})(x) - \nu_{1.15}({V_{0.48}}^{\alpha_0}) - {\delta_{1.15,0.48}}^{\alpha_0}\, V_{0.48}(x)^{\alpha_0} \le \frac{\sigma}{\sigma^{\alpha_0}}\, 1.6244^{\alpha_0} - 0.9353^{\alpha_0} \le 0.$$

Therefore it follows from Corollary 4.2 that the estimate (24) is

$$\forall n \ge 1, \quad \|\pi - \widetilde{\mu}_n\|_{TV} \le 2\,\frac{\nu_{1.15}({V_{0.48}}^{\alpha_0})}{1 - {\delta_{1.15,0.48}}^{\alpha_0}}\, {\delta_{1.15,0.48}}^{\alpha_0 n} \approx 17.65 \times 0.9795^n.$$

In this case, we get better bounds in total variation norm than that provided in [RT99, Table 4]. Our conclusions are summarized in Table 1. Note that, in Table 1, a factor 2 is applied to the estimates in [Ros95, RT99] since our definition of total variation norm is twice that used in these papers. For each $r \in \{0.16, 0.48\}$, the best value of d using the bound (24) is provided in the last row. Finally, we also report in the last row of the table, the best bound derived from (24) tuning the parameters (r, d).

r	d	Method	rate	α_0	Bound	Bound $n = 500$	Bound $n = 650$
0.16	1.11	[Bax05, Tab. 2, Th. 1.3]	0.9747^n				
		(24) in Th 3.1	0.9744^n	1	15.61×0.9744^n	3.68e-05	7.54e-07
	1.39	(24) in Th 3.1	0.9634^n	1	5.78×0.9634^n	4.60e-08	1.71e-10
0.48	1.15	[Ros95] in [RT99, Table 4]	0.991^n			0.092	0.024
		[RT99, Table 4]	0.983^n			0.02	0.002
		(24) in Th 3.1	0.9795^n	0.31	17.65×0.9795^n	0.0005	2.48e-05
	1.06	(24) in Th 3.1	0.9784^n	0.397	20.13×0.9784^n	0.0004	1.43e-05
0.36	1.10	(24) in Th 3.1	0.951^n	1	8.96×0.9510^n	1.12e-10	6.02e-14

Table 1: The results of [RT99, Bax05] and the estimates from (24)

The following more accurate minorization measure is used in [Bax05, Section 8.4]

$$\nu_d(dx) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-(|x|+d)^2/2} dx & \text{if } |x| \le d \\ \frac{1}{\sqrt{2\pi}} e^{-d|x|-x^2} dx & \text{if } |x| \ge d. \end{cases}$$

Choose $d := 1.1$ and $r := 0.22$ as in [Bax05, Section 8.4]. We get

$$\begin{aligned} \nu_d(V_r) &= 2\, e^{((r-d)^2-d^2)/2} \big[\, \Phi(2d-r) - \Phi(d-r)\, \big] + \sqrt{2} e^{(d-r)^2/4} \big[\, 1 - \Phi\big((3d-r)/\sqrt{2}\big)\, \big] \\ &\approx 0.2916. \end{aligned}$$

Given that P satisfies ($\boldsymbol{D_{S^c}}$) with $\delta_{d,r} = \lambda(d,r) \approx 0.9664$, Condition ($\boldsymbol{D_S}$) holds since

$$\max_{|x|\le 1.1} \big[(PV_r)(x) - \delta_{d,r}\, V_r(x)\big] - \nu_d(V_r) = \lambda(0,r) - \delta_{d,r} - \nu_d(V_r) \approx 0.1307 - 0.2916 < 0.$$

We deduce from Theorem 4.2 and (24) that

$$\forall n \ge 1, \quad \|\pi - \widetilde{\mu}_n\|_{TV} \le 2\, \frac{\nu_{1.1}(V_{0.22})}{1 - \delta_{1.1,0.22}} \times {\delta_{1.1,0.22}}^n \approx 17.37 \times 0.9664^n. \tag{95}$$

The rate of convergence 0.9664^n is slightly better than 0.9667^n obtained in [Bax05, 4th line of Table 3]. Anyway observe that the above multiplicative constant 17.37 is not too large. If we tune the values of the parameters (d,r) to derive the best rate of convergence from (24) in Th 3.1, then we obtain for $(d,r) = (1.2, 0.43)$ that $\alpha_0 = 1$ and

$$\forall n \ge 1, \quad \|\pi - \widetilde{\mu}_n\|_{TV} \le 2\, \frac{\nu_{1.2}(V_{0.43})}{1 - \delta_{1.2,0.43}} \times {\delta_{1.2,0.43}}^n \approx 8.46 \times 0.9344^n.$$

The earlier bound (95) is improved by this last approximation of π.

6.4 Gaussian autoregressive Markov chain (geometric case)

We consider an autoregressive Gaussian Markov chain on $\mathbb{X} = \mathbb{R}$ associated with Gaussian transition kernel $P(x,\cdot) = \mathcal{N}(\theta x, 1-\theta^2)$ with $\theta \in (-1,1)$, that is

$$\forall x \in \mathbb{R}, \quad P(x,dy) = \frac{1}{\sqrt{2\pi(1-\theta^2)}} \exp\left(-\frac{(y-\theta x)^2}{2(1-\theta^2)}\right) dy.$$

The P-invariant distribution is $\pi = \mathcal{N}(0,1)$ for any $\theta \in (-1,1)$. This Markov model is also known as contracting normals if introduced as a component of a two-component Gibbs sampler. The convergence of the ergodic averages to π is studied in [RR97, RT99]. The convergence of the iterates is investigated in [Ros95, RT99, QH21], [Bax05, Sect. 8.3]. Set $V(x) := 1 + x^2$ and $S := [-d,d]$ as in these works. Then, if $d > 1$, we know from the computations in [Bax05, Sect. 8.3] that P satisfies ($\boldsymbol{D_{S^c}}$) with

$$\delta_{d,\theta} = \theta^2 + 2\, \frac{1-\theta^2}{1+d^2} < 1.$$

Moreover P satisfies (**S**) with

$$\nu_{d,\theta}(dy) = \min_{x\in[-d,d]} \frac{1}{\sqrt{2\pi(1-\theta^2)}} \exp\left(- \frac{(y-\theta\, x)^2}{2(1-\theta^2)}\right) 1_{[-d,d]}(y)\, dy.$$

We know from the formula given in [Bax05, p. 728] that

$$\nu_{d,\theta}(1_{\mathbb{X}}) = 2\bigg(\Phi\bigg(\frac{(1+|\theta|)\,d}{\sqrt{1-\theta^2}}\bigg) - \Phi\bigg(\frac{|\theta|\,d}{\sqrt{1-\theta^2}}\bigg)\bigg).$$

Moreover, setting $c := (2\pi(1-\theta^2))^{-1/2}$ and $W(y) := y^2$, we obtain that

$$\nu_{d,\theta}(W) = c\bigg(\int_{-d}^{0} y^2 \exp\Big(-\frac{(y-|\theta|\,d)^2}{2(1-\theta^2)}\Big)\,dy + \int_{0}^{d} y^2 \exp\Big(-\frac{(y+|\theta|\,d)^2}{2(1-\theta^2)}\Big)\,dy\bigg),$$

from which we deduce that

$$\nu_{d,\theta}(V) = \nu_{d,\theta}(1_{\mathbb{X}}) + \nu_{d,\theta}(W).$$

Finally, we know from [Bax05, p. 728] that

$$\sup_{x\in[-d,d]} (PV)(x) - \delta_{d,\theta} V(x) = \frac{2(1-\theta^2)\,d^2}{1+d^2}.$$

Then, it is easily checked that for $\theta \in \{0.5, 0.75, 0.9\}$, Theorem 4.2 does not apply. Let us use Corollary 4.2. We deduce from the formula given in [Bax05, p. 728] that

$$K_{d,\theta} := \sup_{x\in[-d,d]} (PV)(x) = 2 + \theta^2(d^2-1).$$

so that we must find $\alpha_0 \in (0,1]$ so that Inequality (35) is satisfied. Since the procedure is as in the Metropolis-Hastings example, the details are omitted. In Table 2, we report the error term (24) of Corollary 4.2 as well as the rates of convergence in [Bax05, Table 4, Th 1.3] which provides the best estimation in the V-weighted operator norm of $P^n - \pi(\cdot)1_{\mathbb{X}}$ among all the methods compared in [Bax05, Sect. 8.3]. The results of [RT99] which are expressed in total variation norm are also reported. Recall that a factor 2 is applied to the estimates in [Ros95, RT99]. Our rates for the convergence in total variation norm of $\widetilde{\mu}_n - \pi$ are slightly better except when $\theta = 1/2$. If $\theta = 1/2$, the rate of convergence is known to be $(1/2)^n$. Thus all these upper bounds are not sharp. Such a gap supports the idea that minorization-drift conditions may be not well suited for obtaining sharp upper bounds for the approximation of π (see [QH21, and references therein] for such a discussion for the convergence of $P(x,\cdot) - \pi$ to 0 when $\mathbb{X} = \mathbb{R}^q$ with large q).

θ	d	Method	Rate	α_0	Bound	Bound $n=45$	$n=60$	Bound $n=1000$
0.5	1.5	[Bax05, Table 4, Th 1.3]	0.897^n					
		(24) in Th 3.1	0.892^n	0.336	7.59×0.892^n	0.044	0.008	
	$\sqrt{3}$	[RT99, Table 4]	0.846^n			0.020	0.001	
		[Ros95] in [RT99, Table 4]	0.881^n			0.022	0.0017	
		(24) in Th 3.1	0.894^n	0.239	6.28×0.894^n	0.04	0.0074	
	1.6	(24) in Th 3.1	0.891^n	0.290	6.89×0.891^n	0.038	0.0067	
0.75	1.2	[Bax05, Table 4, Th 1.3]	0.9847^n					
		(24) in Th 3.1	0.9844^n	0.191	22.64×0.9844^n			
	$\sqrt{3}$	[RT99, Table 4]	0.992^n					0.014
		[Ros95] in [RT99, Table 4]	0.993^n					0.004
		(24) in Th 3.1	0.991^n	0.036	11.30×0.991^n			0.0016
	1.3	(24) in Th 3.1	0.9834^n	0.141	17.13×0.9834^n			9.1e-07
0.9	1.1	[Bax05, Table 4, Th 1.3]	0.99948^n					
		(24) in Th 3.1	0.99947^n	0.029	87.49×0.99947^n			
	1.14	(24) in Th 3.1	0.99944^n	0.022	66.70×0.99944^n			

Table 2: The results of [Bax05] and [Ros95, RT99] and the estimates from (24)

References

[AF05] C. Andrieu and G. Fort. Explicit control of subgeometric ergodicity. Technical Report 05:17, 2005.

[AFV15] C. Andrieu, G. Fort, and M. Vihola. Quantitative convergence rates for subgeometric Markov chains. *J. Appl. Probab.*, 52(2):391–404, 2015.

[Bax05] P. H. Baxendale. Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.*, 15(1B):700–738, 2005.

[Del17] B. Delyon. Convergence rate of the powers of an operator. Applications to stochastic systems. *Bernoulli*, 23(4A):2129–2180, 2017.

[DFMS04] R. Douc, G. Fort, E. Moulines, and P. Soulier. Practical drift conditions for subgeometric rates of convergence. *Ann. Appl. Probab.*, 14(3):1353–1377, 2004.

[DMPS18] R. Douc, E. Moulines, P. Priouret, and P. Soulier. *Markov chains*. Springer, 2018.

[Hen06] H. Hennion. Quasi-compactness and absolutely continuous kernels, applications to Markov chains. https://arxiv.org/abs/math/0606680, June 2006.

[Hen07] H. Hennion. Quasi-compactness and absolutely continuous kernels. *Probab. Theory Related Fields*, 139:451–471, 2007.

[HL14] L. Hervé and J. Ledoux. Spectral analysis of Markov kernels and aplication to the convergence rate of discrete random walks. *Adv. Appl. Probab.*, 46(4):1036–1058, 2014.

[HL20a] L. Hervé and J. Ledoux. Additional material on V-geometrical ergodicity of Markov kernels via finite-rank approximations. hal-02410491v4, 2020.

[HL20b] L. Hervé and J. Ledoux. V-geometrical ergodicity of Markov kernels via finite-rank approximations. *Electron. Commun. Probab.*, 25(23):1–12, March 2020.

[HM11] M. Hairer and J. C. Mattingly. Yet another look at Harris' ergodic theorem for Markov chains. In Robert Dalang, Marco Dozzi, and Francesco Russo, editors, *Seminar on Stochastic Analysis, Random Fields and Applications VI*, pages 109–117, Basel, 2011. Springer Basel.

[JR02] S. F. Jarner and G. O. Roberts. Polynomial Convergence Rates of Markov Chains. *Ann. Appl. Probab.*, 12(1):224 – 247, 2002.

[LC14] Manuel E. Lladser and Stephen R. Chestnut. Approximation of sojourn-times via maximal couplings: motif frequency distributions. *J. Math. Biol.*, 69(1):147–182, 2014.

[LT96] R. B. Lund and R. L. Tweedie. Geometric convergence rates for stochastically ordered Markov chains. *Math. Oper. Res.*, 21(1):182–194, 1996.

[MT93] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability.* Springer-Verlag London Ltd., London, 1993.

[MT94] S. P. Meyn and R. L. Tweedie. Computable bounds for geometric convergence rates of Markov chains. *Ann. Probab.*, 4:981–1011, 1994.

[MT96] K. L. Mengersen and R. L. Tweedie. Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.*, 24(1):101–121, 1996.

[MT09] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability.* Cambridge University Press, 2th edition, 2009.

[Num84] E. Nummelin. *General irreducible Markov chains and nonnegative operators.* Cambridge University Press, Cambridge, 1984.

[QH21] Q. Qin and J. P. Hobert. On the limitations of single-step drift and minorization in Markov chain convergence analysis. *Ann. Appl. Probab.*, 31(4):1633–1659, 2021.

[Ros95] J. S. Rosenthal. Minorization conditions and convergence rates for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.*, 90(430):558–566, 1995.

[Ros02] J. S. Rosenthal. Quantitative convergence rates of Markov chains: a simple account. *Electron. Comm. Probab.*, 7:123–128, 2002.

[RR97] G. O. Roberts and J. S. Rosenthal. Geometric ergodicity and hybrid Markov chains. *Electron. Commun. Probab.*, 2:13–25, 1997.

[RR04] Gareth O. Roberts and Jeffrey S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probab. Surv.*, 1:20–71 (electronic), 2004.

[RT99] G. O. Roberts and R. L. Tweedie. Bounds on regeneration times and convergence rates for Markov chains. *Stochastic Process. Appl.*, 80(2):211–229, 1999.

[RT00] G. O. Roberts and R. L. Tweedie. Rates of convergence of stochastically monotone and continuous time Markov models. *J. Appl. Probab.*, 37(2):359–373, 2000.

A Proof of Proposition 2.1

Recall that $0 \le T \le P$ is deduced from the positivity of ν and from (**S**). That $0 \le T_n \le P^n$ follows from $0 \le T \le P$. The equality in (13a) for $n = 1$ is obvious from the definition of T. Now assume that this equality holds for some $n \ge 1$. Then

$$P^{n+1} - T_{n+1} := (P-T)^{n+1} \quad = \quad (P-T)(P^n - T_n) = P^{n+1} - PT_n - TP^n + TT_n$$

from which we deduce that, for every $f \in \mathcal{B}$

$$\begin{aligned} T_{n+1} f &= PT_n f + TP^n f - TT_n f \\ &= \sum_{k=1}^{n} \beta_k(f) P^{n-k+1} 1_S + \Big(\beta_1(P^n f) - \sum_{k=1}^{n} \beta_k(f) \nu(P^{n-k} 1_S) \Big) 1_S \\ &= \sum_{k=1}^{n} \beta_k(f) P^{n+1-k} 1_S + \beta_{n+1}(f) 1_S \end{aligned}$$

with $\beta_{n+1}(\cdot)$ defined in (11). This provides the equality in (13a) by induction. Next we obtain that for every $n \ge 1$

$$P^n - T_n := (P-T)^n = (P^{n-1} - T_{n-1})(P-T) = P^n - P^{n-1}T - T_{n-1}P + T_{n-1}T$$

so that

$$T_n - T_{n-1}P = P^{n-1}T - T_{n-1}T = (P^{n-1} - T_{n-1})T.$$

Formula (13b) is proved. Now it follows from (11) and (13a) that

$$\beta_n(f) = \nu\big(P^{n-1} f\big) - \sum_{k=1}^{n-1} \beta_k(f)\, \nu\big(P^{n-k-1} 1_S\big) = \nu\big(P^{n-1} f - T_{n-1} f\big).$$

This gives the two first equalities in (14), from which the last one is easily deduced. Finally note that $\beta_1(\cdot) = \nu(\cdot)$ is a positive measure on $(\mathbb{X}, \mathcal{X})$, so that for every $n \ge 1$ $\beta_n(\cdot)$ is a linear combination of non-negative measures on $(\mathbb{X}, \mathcal{X})$ (by induction). That β_n is a finite non-negative measure follows from (14) since $0 \le P^{n-1} - T_{n-1} \le P^{n-1}$

B Proofs of Lemmas 6.2 and 6.3

Proof of Lemma 6.2. Using the definition of P and (88) we have for every $i > s$

$$\begin{aligned} (PV)(i) &= \mathbb{P}(W < -i) + \sum_{j \ge -i} (1+i+j)^{m_0}\, p_j \\ &= \mathbb{P}(W < -i) + \sum_{j \ge -i} \sum_{k=0}^{m_0} C_{m_0}^k\, (1+i)^{m_0-k}\, j^k p_j \\ &\le V(i) + \tau m_0 \mathbb{E}[W](1+i)^{m_0-1} + \left(\frac{\mathbb{P}(W < -i)}{(1+i)^{m_0-1}} + \sum_{j \ge -i} \sum_{k=2}^{m_0} \frac{C_{m_0}^k j^k p_j}{(1+i)^{k-1}} \right) (1+i)^{m_0-1} \\ &\le V(i) + \tau m_0 \mathbb{E}[W]\, (1+i)^{m_0-1} - \frac{\tau\, m_0\, \mathbb{E}[W]}{2}\, (1+i)^{m_0-1} \\ &\le V(i) + \frac{\tau\, m_0\, \mathbb{E}[W]}{2}\, (1+i)^{m_0-1}. \end{aligned}$$

$\square$

Proof of Lemma 6.3. Let $q \in \{1, \dots, m_0\}$ and $\ell \in S \setminus \{0\}$. Then

$$\begin{aligned}
(PV_q)(\ell) &= \mathbb{P}(W < -\ell) + \sum_{j \geq -\ell} (1+\ell+j)^q \, p_j \\
&= \mathbb{P}(W < -\ell) + \sum_{k=0}^{q} C_q^k \sum_{j \geq -\ell} (\ell+j)^k \, p_j \\
&= \mathbb{P}(W < -\ell) + \sum_{j \geq -\ell+1} (\ell+j)^q \, p_j + \sum_{k=0}^{q-1} C_q^k \sum_{j \geq -\ell} (\ell+j)^k \, p_j \\
&= \mathbb{P}(W < -\ell) - \mathbb{P}(W < -\ell+1) + (PV_q)(\ell-1) + \sum_{k=0}^{q-1} C_q^k \sum_{j \geq -\ell} (\ell+j)^k \, p_j. \qquad (96)
\end{aligned}$$

Next

$$\sum_{k=0}^{q-1} C_q^k \sum_{j \geq -\ell} (\ell+j)^k \, p_j = \sum_{k=0}^{q-1} C_q^k \, \mathbb{E}\big[(\ell+W)^k\big] - \sum_{k=0}^{q-1} C_q^k \sum_{j=-\infty}^{-\ell-1} (\ell+j)^k \, p_j. \qquad (97)$$

Now let $i \in \mathbb{N} \setminus \{0\}$. We have

$$\begin{aligned}
\sum_{\ell=1}^{i} \sum_{k=0}^{q-1} C_q^k \, \mathbb{E}\big[(\ell+W)^k\big] &= \sum_{\ell=1}^{i} \Big(\mathbb{E}\big[(1+\ell+W)^q\big] - \mathbb{E}\big[(\ell+W)^q\big] \Big) \\
&= \mathbb{E}\big[(1+i+W)^q\big] - \mathbb{E}\big[(1+W)^q\big]. \qquad (98)
\end{aligned}$$

Moreover we have for any $i \in \mathbb{N} \setminus \{0\}$.

$$\begin{aligned}
D_q(i) := -\sum_{\ell=1}^{i} \sum_{k=0}^{q-1} C_q^k \sum_{j=-\infty}^{-\ell-1} (\ell+j)^k \, p_j &= -\sum_{j=-\infty}^{-2} p_j \sum_{\ell=1}^{\min(i,-j-1)} \sum_{k=0}^{q-1} C_q^k \, (\ell+j)^k \\
&= -\sum_{j=-\infty}^{-2} p_j \sum_{\ell=1}^{\min(i,-j-1)} \big[(1+\ell+j)^q - (\ell+j)^q\big] \\
&= \sum_{j=-\infty}^{-2} p_j \left[(1+j)^q - \big(1+\min(i,-j-1)+j\big)^q\right] \\
&= \sum_{j=-\infty}^{-2} p_j \, (1+j)^q - \sum_{j=-\infty}^{-i-2} p_j \, \big(1+i+j\big)^q. \qquad (99)
\end{aligned}$$

Then it follows from (96) and from (97)-(98)-(99) that

$$\begin{aligned}
\forall i \in S \setminus \{0\}, \quad (PV_q)(i) - (PV_q)(0) &= \sum_{\ell=1}^{i} \big[(PV_q)(\ell) - (PV_q)(\ell-1)\big] \\
&= \mathbb{E}\big[(1+i+W)^q\big] - \mathbb{E}\big[(1+W)^q\big] + C_q(i)
\end{aligned}$$

with $C_q(i)$ defined in Lemma 6.3. This gives (92) for every $i \in \mathbb{N} \setminus \{0\}$. Equality (92) is obvious for $i = 0$ since $C_q(0) = 0$ by definition. $\square$