Tweedie-type stability estimates for the invariant probability measures of perturbed Markov chains under drift conditions

Loïc Hervé, James Ledoux

Tweedie-type stability estimates for the invariant probability measures of perturbed Markov chains under drift conditions.

Loïc HERVÉ, and James LEDOUX *

version: Thursday 24^{th} November, 2022 – 11:09

Abstract

Given a perturbed version P_θ of a Markov kernel P_{θ_0} with respective invariant probability measures π_θ and π_{θ_0}, we provide estimates of the W-weighted norm $\|\pi_\theta - \pi_{\theta_0}\|_W$ for some Lyapunov function W. We follow Tweedie's approach proposed in a seminal paper on the truncation-augmentation scheme for approximating the invariant probability measure of discrete-state Markov kernels. But the novelty here is that the state space for the Markov kernels and the form of the perturbation are general, and that the intermediate term $\|\pi_\theta - P_\theta^n(x,\cdot)\|_W$ usually involved to control the error norm $\|\pi_\theta - \pi\|_W$ is replaced with $\|\pi_\theta - \widetilde{\mu}_n^{(\theta)}\|_W$, where $\widetilde{\mu}_n^{(\theta)}$ is an alternative probability measure which has been introduced in a recent work for approximating π_θ under a minorization condition. The interest is that the estimates of $\|\pi_\theta - \widetilde{\mu}_n^{(\theta)}\|_W$ turn out to be much more accurate and practicable than for $\|\pi_\theta - P_\theta^n(x,\cdot)\|_W$ under geometric or polynomial drift conditions. Moreover we do not need to resort to the use of techniques related to the existence of an atom for P_{θ_0}. This study is performed for geometrically or polynomially ergodic Markov kernels, and compared with prior works when applied to the standard truncation-augmentation scheme for discrete-state Markov kernels.

AMS subject classification : 60J05

Keywords : Invariant probability measure; Rate of convergence; Perturbed Markov kernels; Drift conditions; Small set; Truncation-augmentation approximation.

1 Introduction

In this paper, we study the sensitivity of the invariant probability measure of a Markov kernel when replaced with a perturbed version. Before presenting the general framework and the main results of this work, we discuss the classical example of the truncated-augmented scheme for approximating the invariant probability measure of an infinite stochastic matrix. Let $P := (P(i,j))_{(i,j)\in\mathbb{N}^2}$ be an infinite stochastic matrix and for every $k \geq 1$ let P_k be a linear augmentation (e.g. in the first or the last column) of the $(k+1)\times(k+1)$ north-west corner truncation of P. Let π (resp. π_k) be the invariant probability measure of P on $\mathbb{N}$ (resp. of P_k on $B_k := \{0,\ldots,k\}$). For the sake of simplicity, the natural extension on $\mathbb{N}$ of

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

the finite matrix P_k and of its invariant probability measure are still denoted by P_k and π_k in this introductory section. If P satisfies some minorization and geometric drift conditions with respect to (w.r.t.) some Lyapunov function $V : \mathbb{N} \to [1, +\infty)$, Tweedie proved in [Twe98, Th. 3.2] that, for the first-column augmentation, we have $\lim_k \|\pi_k - \pi\|_V = 0$ where $\|\cdot\|_V$ stands for the V-weighted total variation norm, see the definition in (6). Tweedie's proof is based on the $V-$geometrical ergodicity property of both P and P_k, that is: there exist $\rho \in (0,1)$ and $C \in (0, +\infty)$ such that

$$\forall k \in \mathbb{N}^* \cup \{\infty\},\ \forall n \geq 0, \quad \sup_{|f|\leq V} \sup_{x\in\mathbb{N}} \frac{\left|({P_k}^n f)(x) - \pi_k(f)1_{\mathbb{X}}\right|}{V(x)} \leq C\,\rho^n \tag{1}$$

with the convention $P_\infty := P$, $\pi_\infty := \pi$. Explicit bounds of $\|\pi_k - \pi\|_V$ are not stated in [Twe98, Th. 3.2] since the use of the rate ρ and constant C in (1) are unlikely to be of practical value, see [Twe98, p. 526]. Although progress on finding computable bounds for ρ and C has been made (e.g. see [MT94, Bax05, HL22a, and references therein]), this issue remains a difficult problem. A favourable but very specific case is when P is stochastically monotone and satisfies the following geometric drift condition w.r.t. some finite set S

$$\exists \delta \in (0,1),\ b > 0, \quad PV \leq \delta\, V + b\, 1_S. \tag{2}$$

In this case and assuming that $S := \{0\}$, Tweedie proved in [Twe98, Th. 4.2, Eq. (46)] that

$$\forall n \geq 1, \quad \|\pi_k - \pi\|_{TV} \leq \frac{4\,b}{1-\delta}\,\delta^n + n\,\eta_k \quad \text{with} \quad \eta_k = \frac{2b}{(1-\delta)\,V(k)} \tag{3}$$

where $\|\cdot\|_{TV}$ denotes the total variation norm (see (7)) and π_k in (3) is the invariant probability measure of the last-column augmentation P_k of P. Under the same assumptions on P, such a bound is proved to hold in [Liu10, Th. 5.2] for any arbitrary augmented truncation approximation. Moreover Liu shows in [Liu10] that Tweedie's approach can also be used to get an explicit bound for $\|\pi_k - \pi\|_{TV}$ when P is assumed to be stochastically monotone and polynomially ergodic. Specifically, if P satisfies a minorization condition with respect to some finite atom S and if the following polynomial drift condition introduced in [JR02]

$$\exists \alpha < 1,\ b, c > 0, \quad PV \leq V - c\,V^\alpha + b\,1_S \tag{4}$$

holds with $c = 1$ and $S = \{0\}$, then we have (see [Liu10, Th. 5.1])

$$\forall n \geq 1, \quad \|\pi_k - \pi\|_{TV} \leq \frac{8V(1)}{(1-\alpha)^{\frac{\alpha}{1-\alpha}}}\frac{1}{n^{\frac{\alpha}{1-\alpha}}} + n\,\xi_k \quad \text{with} \quad \xi_k = \frac{2b}{V(k+1)^\alpha}. \tag{5}$$

In fact, this estimate only requires that P be dominated by a stochastically monotone Markov kernel Q satisfying the above drift conditions. Mention that this kind of estimates has been obtained in [Mas16] for the last-column augmentation of block-monotone Markov chains under the subgeometric drift condition introduced in [DFMS04].

The purpose of this work is to use the recent work [HL22b] in order to extend Tweedie-type estimate (5) to perturbed Markov kernels $\{P_\theta\}_{\theta\in\Theta}$ (not necessarily obtained by truncation) defined on a general state space and satisfying uniform (w.r.t. $\theta \in \Theta$) minorization and polynomial drift conditions. In particular, we do not assume that P_θ is stochastically monotone, nor do we assume that the small set of the minorization condition is an atom. Moreover, all the estimates are expressed in the $V-$weighted total variation norm for some $V \geq 1$, which obviously dominates the total variation norm. Application to truncation of discrete Markov kernels is discussed in Section 3. Note that, although the case of geometrically ergodic Markov kernels P can be similarly developed (see Appendix A), it is omitted here since it has been proved in [LL18, HL22a] that a better approach can be used to derive very simple explicit estimates of $\|\pi_\theta - \pi\|_V$ from $\int \|P_\theta(x,\cdot) - P(x,\cdot)\|_V \, \pi_\theta(dx)$ where π_θ is the invariant probability measure of P_θ, see Remark 2.2.

Our general context is the following one. Let $(\mathbb{X}, \mathcal{X})$ be a measurable space and let $\{P_\theta\}_{\theta\in\Theta}$ be a family of transition kernels on $(\mathbb{X}, \mathcal{X})$, where Θ is an open subset of some metric space. 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 $\int_{\mathbb{X}} f d\mu$. We assume that $\{P_\theta\}_{\theta\in\Theta}$ satisfies the following minorization condition

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

Assumption $(\boldsymbol{S_\Theta})$ means that S is a small-set for the whole family $\{P_\theta\}_{\theta\in\Theta}$ and that the associated positive measure is the same for all $\theta \in \Theta$, namely ν. We also suppose that, for every $\theta \in \Theta$, there exists a unique $P_\theta-$invariant probability measure π_θ on $(\mathbb{X}, \mathcal{X})$ such that $\pi_\theta(1_S) > 0$. Moreover, for every $\theta \in \Theta$ and for every $n \geq 1$ we consider the following probability measure $\widetilde{\mu}_n^{(\theta)}$ on $(\mathbb{X}, \mathcal{X})$ introduced in [HL20, HL22b]

$$\widetilde{\mu}_n^{(\theta)} := \frac{1}{\mu_n^{(\theta)}(1_{\mathbb{X}})}\, \mu_n^{(\theta)} \quad \text{where} \quad \mu_n^{(\theta)} := \sum_{k=1}^{n} \nu \circ R_\theta^{k-1} \quad \text{and} \quad R_\theta := P_\theta - \nu(\cdot) 1_S.$$

Now let $W : \mathbb{X} \to [1, +\infty)$ be a measurable function such that

$$\forall \theta \in \Theta, \quad \pi_\theta(W) < \infty \quad \text{and} \quad \exists C \in (0, +\infty),\ \forall \theta \in \Theta, \quad P_\theta W \leq CW. \tag{$\boldsymbol{W_\Theta}$}$$

Under Assumption $(\boldsymbol{S_\Theta})$ and the first condition in $(\boldsymbol{W_\Theta})$ we know from [HL22b, Ths. 2.1-3.1] that for every $\theta \in \Theta$ we have $\lim_n \|\pi_\theta - \widetilde{\mu}_n^{(\theta)}\|_W = 0$, where $\|\cdot\|_W$ stands for the W-weighted total variation norm, see (6). Now let θ_0 be fixed in Θ. To have a good understanding of the next results, P_{θ_0} has to be viewed as the unperturbed Markov kernel with unknown invariant probability measure π_{θ_0}, while P_θ for $\theta \neq \theta_0$ are the perturbed Markov kernels with known or computable invariant probability measure π_θ. The goal is to approximate π_{θ_0} by π_θ.

In Section 2, under Assumptions $(\boldsymbol{S_\Theta})$-$(\boldsymbol{W_\Theta})$ we introduce the following quantities

$$\forall n \geq 1,\ \widetilde{\varepsilon}_{n,\Theta,W} := \sup_{\theta\in\Theta} \|\pi_\theta - \widetilde{\mu}_n^{(\theta)}\|_W \ \text{ and } \ \forall \theta \in \Theta,\ \forall x \in \mathbb{X},\ \Delta_{\theta,W}(x) := \|P_\theta(x,\cdot) - P_{\theta_0}(x,\cdot)\|_W,$$

and we prove in Theorem 2.1 that $\lim_{\theta\to\theta_0} \|\pi_\theta - \pi_{\theta_0}\|_W = 0$ provided that the two following conditions hold:

$$\lim_{n\to+\infty} \widetilde{\varepsilon}_{n,\Theta,W} = 0 \quad \text{and} \quad \forall x \in \mathbb{X}, \quad \lim_{\theta\to\theta_0} \Delta_{\theta,W}(x) = 0.$$

Under these assumptions we prove in Corollary 2.1 that the real numbers $\gamma_\theta := \pi_\theta(\Delta_{\theta,1_{\mathbb{X}}})$ and $\gamma_{\theta,W} := \pi_\theta(\Delta_{\theta,W})$ converge to 0 when $\theta \to \theta_0$. Moreover we give a rate of convergence for $\|\pi_\theta - \pi_{\theta_0}\|_W$ depending on $\widetilde{\varepsilon}_{n,\Theta,W}$, γ_θ and $\gamma_{\theta,W}$. In practice, γ_θ and $\gamma_{\theta,W}$ are supposed to be computable for $\theta \neq \theta_0$, so that the error bounds for $\|\pi_\theta - \pi_{\theta_0}\|_W$ obtained in Corollary 2.1 is relevant whenever the rate of convergence of the sequence $\{\widetilde{\varepsilon}_{n,\Theta,W}\}_{n\geq 1}$ is known. This last question can be investigated from [HL22b] under geometrical or polynomial drift conditions. Then, assuming that the whole family $\{P_\theta\}_{\theta\in\Theta}$ satisfies these polynomial drift conditions in a uniform way in $\theta \in \Theta$, accurate and computable bounds for $\widetilde{\varepsilon}_{n,\Theta,W}$ are addressed in Section 3. When applied to the above described truncation framework with an atomic small-set as in [Twe98, Liu10], the error bounds for $\|\pi_k - \pi\|_{TV}$ are proved to be quite similar to Liu's estimate (5) under the polynomial drift condition (4) on P. Under geometric drift condition (2) on P, the estimates of $\|\pi_k - \pi\|_V$ are quite similar to Tweedie's estimate (3), see Appendix A.

The novelty of this work is to use the intermediate quantity $\|\pi_\theta - \widetilde{\mu}_n^{(\theta)}\|_W$ in place of $\|\pi_\theta - P_\theta^n(x,\cdot)\|_W$ to control the error norm $\|\pi_\theta - \pi_{\theta_0}\|_W$. This control is studied under standard drifts conditions. The benefit of this approach is that the bounds on $\|\pi_\theta - \widetilde{\mu}_n^{(\theta)}\|_W$ obtained in [HL22b] are in general much more accurate and practicable than those obtained in the literature for $\|\pi_\theta - P_\theta^n(x,\cdot)\|_W$, see [HL22b, Sec. 6]. Finally recall that neither atomic (except when we compare with Liu's work) nor stochastic monotonicity assumptions are required in this paper. In the context of truncation approximation, the convergence of $\{\pi_k\}_{k\geq 1}$ to π has been studied for a long time, see [Twe98, IGL22, IG22, and references therein]. The results related to this note are briefly discussed in Remark 3.4.

Notations. If $W : \mathbb{X} \to [1,+\infty)$ is a measurable function and if $(\lambda_1,\lambda_2) \in (\mathcal{M}^+)^2$ is such that $\lambda_i(W) < \infty$ for $i = 1,2$, then the W-weighted total variation norm $\|\lambda_1 - \lambda_2\|_W$ is defined by

$$\|\lambda_1 - \lambda_2\|_W := \sup_{|f|\leq W} \big|\lambda_1(f) - \lambda_2(f)\big|. \tag{6}$$

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, which can also be defined by

$$\|\lambda_1 - \lambda_2\|_{TV} = 2 \sup_{A\in\mathcal{X}} |\lambda_1(1_A) - \lambda_2(1_A)|. \tag{7}$$

Recall that a non-negative kernel $R(x,dy)$, $x \in \mathbb{X}$, on $(\mathbb{X},\mathcal{X})$ is said to be submarkovian if for every $x \in \mathbb{X}$ we have $R(x,\mathbb{X}) \leq 1$. We denote by R its functional action defined by

$$\forall x \in \mathbb{X}, \quad (Rf)(x) := \int_{\mathbb{X}} f(y)\, R(x,dy),$$

where $f : \mathbb{X} \to \mathbb{R}$ is any $R(x,\cdot)-$integrable function. For every $n \geq 1$ the $n-$th iterate kernel of $R(x,dy)$ is denoted by $R^n(x,dy)$, $x \in \mathbb{X}$, and R^n stands for its functional action. As usual R^0 is the identity map I by convention.

2 Main results

Let $\{P_\theta\}_{\theta\in\Theta}$ be a family of transition kernels on $(\mathbb{X},\mathcal{X})$, where Θ is an open subset of some metric space. We assume that the family $\{P_\theta\}_{\theta\in\Theta}$ satisfies Assumption $(\boldsymbol{S_\Theta})$, and that for every $\theta \in \Theta$ there exists a unique $P_\theta-$invariant probability measure π_θ on $(\mathbb{X},\mathcal{X})$ such that $\pi_\theta(1_S) > 0$. For every $\theta \in \Theta$ we consider the following recursive sequence $\{\beta_k^{(\theta)}\}_{k\geq 1} \in (\mathcal{M}^+)^{\mathbb{N}}$

introduced in [HL20, HL22b]

$$\beta_1^{(\theta)}(\cdot) := \nu(\cdot) \quad \text{and} \quad \forall n \geq 2, \quad \beta_n^{(\theta)}(\cdot) := \nu\big(P_\theta^{n-1}\,\cdot\,\big) - \sum_{k=1}^{n-1} \nu\big(P_\theta^{n-k-1} 1_S\big)\, \beta_k^{(\theta)}(\cdot). \tag{8}$$

We know from [HL20, HL22b] that

$$0 < \mu^{(\theta)}(1_{\mathbb{X}}) := \sum_{k=1}^{+\infty} \beta_k^{(\theta)}(1_{\mathbb{X}}) < \infty, \quad \pi_\theta(1_S) = \mu^{(\theta)}(1_{\mathbb{X}})^{-1} \quad \text{and} \quad \pi_\theta := \mu^{(\theta)}(1_{\mathbb{X}})^{-1}\, \mu^{(\theta)} \tag{9}$$

where $\mu^{(\theta)} \in \mathcal{M}_*^+$ is defined by

$$\mu^{(\theta)} := \sum_{k=1}^{+\infty} \beta_k^{(\theta)}. \tag{10}$$

Moreover, for every $\theta \in \Theta$ and for every $n \geq 1$, let us define $\mu_n^{(\theta)} \in \mathcal{M}_*^+$ and the probability measure $\widetilde{\mu}_n^{(\theta)}$ on $(\mathbb{X}, \mathcal{X})$ by:

$$\mu_n^{(\theta)} := \sum_{k=1}^{n} \beta_k^{(\theta)} \quad \text{and} \quad \widetilde{\mu}_n^{(\theta)} := \mu_n^{(\theta)}(1_{\mathbb{X}})^{-1}\, \mu_n^{(\theta)}. \tag{11}$$

2.1 Basic estimates

For any measurable function $W : \mathbb{X} \to [1, +\infty)$ satisfying Assumption $(\boldsymbol{W_\Theta})$ and for any fixed $\theta_0 \in \Theta$, let us introduce the following quantities

$$\forall n \geq 1, \quad \widetilde{\varepsilon}_{n,\Theta,W} = \sup_{\theta \in \Theta} \|\pi_\theta - \widetilde{\mu}_n^{(\theta)}\|_W \tag{12}$$

$$\forall \theta \in \Theta,\ \forall x \in \mathbb{X}, \quad \Delta_{\theta,W}(x) := \|P_\theta(x,\cdot) - P_{\theta_0}(x,\cdot)\|_W \tag{13}$$

together with the two following conditions:

$$\lim_{n \to +\infty} \widetilde{\varepsilon}_{n,\Theta,W} = 0 \tag{$\boldsymbol{\mathcal{E}_W}$}$$

$$\forall x \in \mathbb{X}, \quad \lim_{\theta \to \theta_0} \Delta_{\theta,W}(x) = 0. \tag{$\boldsymbol{\Delta_W}$}$$

Theorem 2.1 *Assume that $\{P_\theta\}_{\theta \in \Theta}$ satisfies Assumption $(\boldsymbol{S_\Theta})$, and that for every $\theta \in \Theta$ there exists a unique $P_\theta-$invariant probability measure π_θ on $(\mathbb{X}, \mathcal{X})$ such that $\pi_\theta(1_S) > 0$. Let $W : \mathbb{X} \to [1, +\infty)$ be a measurable function satisfying Assumption $(\boldsymbol{W_\Theta})$. Then the following inequalities hold for for every $n \geq 2$*

$$\textit{for } U \in \{1_{\mathbb{X}}, W\}, \quad \|\mu_n^{(\theta)} - \mu_n^{(\theta_0)}\|_U \leq \sum_{k=1}^{n-1} \min\left\{\mu_k^{(\theta_0)}(\Delta_{\theta,U}),\, \mu_k^{(\theta)}(\Delta_{\theta,U})\right\} \tag{14a}$$

$$\|\pi_\theta - \pi_{\theta_0}\|_W \leq 2\,\widetilde{\varepsilon}_{n,\Theta,W} + \frac{\|\mu_n^{(\theta)} - \mu_n^{(\theta_0)}\|_W}{\mu_n^{(\theta)}(1_{\mathbb{X}})} + \frac{\mu_n^{(\theta_0)}(W)\,\|\mu_n^{(\theta_0)} - \mu_n^{(\theta)}\|_{TV}}{\mu_n^{(\theta)}(1_{\mathbb{X}})\,\mu_n^{(\theta_0)}(1_{\mathbb{X}})}. \tag{14b}$$

If moreover Assumptions $(\boldsymbol{\Delta_W})$ and $(\boldsymbol{\mathcal{E}_W})$ hold, then

$$\lim_{\theta \to \theta_0} \|\pi_\theta - \pi_{\theta_0}\|_W = 0. \tag{15}$$

Remark 2.1 *If $\{P_\theta\}_{\theta\in\Theta}$ satisfies Assumption ($\boldsymbol{S_\Theta}$) and if each P_θ admits a unique invariant probability measure π_θ on $(\mathbb{X},\mathcal{X})$ such that $\pi_\theta(1_S)>0$, then*

$$\|\pi_\theta-\pi_{\theta_0}\|_{TV}\le 2\,\widetilde{\varepsilon}_{n,\Theta,1_{\mathbb{X}}}+2\,\mu_n^{(\theta)}(1_{\mathbb{X}})^{-1}\,\|\mu_n^{(\theta)}-\mu_n^{(\theta_0)}\|_{TV} \tag{16}$$

since $W=1_{\mathbb{X}}$ obviously satisfies Assumption ($\boldsymbol{W_\Theta}$), so that (14b) with $W=1_{\mathbb{X}}$ gives (16). In particular both estimates (16) and (14b) hold under the assumptions of Theorem 2.1.

The proof of Theorem 2.1 is based on the two next lemmas. Lemma 2.1 below is classical, e.g. see [Twe98]. Its proof is recalled for convenience.

Lemma 2.1 *Let $W:\mathbb{X}\to[1,+\infty)$ be a measurable function. For $i=0,1$ let $R_i(x,dy)$, $x\in\mathbb{X}$, be two non-negative submarkovian kernels on $(\mathbb{X},\mathcal{X})$ such that*

$$\forall i\in\{0,1\},\ \forall x\in\mathbb{X},\quad \int_{\mathbb{X}} W(y)\,R_i(x,dy)\le W(x), \tag{17}$$

and define

$$\forall x\in\mathbb{X},\quad \Delta_W(x)=\|R_1(x,\cdot)-R_0(x,\cdot)\|_W.$$

Then we have for every measurable function $f:\mathbb{X}\to\mathbb{R}$ such that $|f|\le W$

$$\forall n\ge 1,\ \forall x\in\mathbb{X},\quad \big|(R_1^nf)(x)-(R_0^nf)(x)\big|\le\min\Big\{\sum_{j=0}^{n-1}\big(R_0^j\Delta_W\big)(x),\sum_{j=0}^{n-1}\big(R_1^j\Delta_W\big)(x)\Big\}. \tag{18}$$

Proof. For $n=1$ Inequality (18) holds since we have for every measurable function $f:\mathbb{X}\to\mathbb{R}$ such that $|f|\le W$

$$\big|(R_1f)(x)-(R_0f)(x)\big|\le\Delta_W(x)$$

from the definition of the W-weighted total variation norm. Next proceed by induction. Assume that (18) holds for some $n\ge1$. Let $g:\mathbb{X}\to\mathbb{R}$ be any measurable function such that $|g|\le W$. Then

$$\begin{aligned}\big|(R_1^{n+1}g)(x)-(R_0^{n+1}g)(x)\big| &\le \big|\big(R_1^n(R_1-R_0)g\big)(x)\big|+\big|\big((R_1^n-R_0^n)R_0g\big)(x)\big|\\ &\le \int_{\mathbb{X}}\big|(R_1g)(y)-(R_0g)(y)\big|\,R_1^n(x,dy)+\sum_{j=0}^{n-1}\big(R_1^j\Delta_W\big)(x)\\ &\le \int_{\mathbb{X}}\Delta_W(y)\,R_1^n(x,dy)+\sum_{j=0}^{n-1}\big(R_1^j\Delta_W\big)(x)\ =\ \sum_{j=0}^{n}\big(R_1^j\Delta_W\big)(x)\end{aligned}$$

using the triangular inequality, the fact that $|R_0g|\le R_0W\le W$ by hypothesis (17) and the induction assumption, and finally the definition of Δ_W. Exchanging the role of R_0 and R_1 in the previous inequality gives (18) at order $n+1$.

□

For any ν-integrable function $f:\mathbb{X}\to\mathbb{R}$, we set

$$Tf:=\nu(f)\,1_S\quad\text{with } S\in\mathcal{X}\text{ and }\nu\in\mathcal{M}_*^+\text{ given in }(\boldsymbol{S_\Theta}). \tag{19}$$

Note that for every $\theta\in\Theta$ we have $0\le T\le P_\theta$ from the positivity of ν and from ($\boldsymbol{S_\Theta}$). Define the submarkovian kernel $R_\theta=P_\theta-T$. We know from [HL22b, Prop. 2.1] that

$$\forall\theta\in\Theta,\ \forall k\ge1,\quad \beta_k^{(\theta)}=\nu\circ R_\theta^{k-1} \tag{20}$$

with the convention $R_\theta^0=I$. Moreover recall that $\beta_1^{(\theta)}=\beta_1^{(\theta_0)}=\nu$.

Lemma 2.2 *Under the assumptions of Theorem 2.1, Property (14a) holds and we have*

$$\forall\theta \in \Theta,\ \forall n \geq 2, \quad \|\widetilde{\mu}_n^{(\theta)} - \widetilde{\mu}_n^{(\theta_0)}\|_W \leq a_n \, \|\mu_n^{(\theta)} - \mu_n^{(\theta_0)}\|_W + b_n \, \|\mu_n^{(\theta_0)} - \mu_n^{(\theta)}\|_{TV} \tag{21a}$$

$$\textit{with} \quad a_n := \frac{1}{\mu_n^{(\theta)}(1_{\mathbb{X}})} \quad \textit{and} \quad b_n := \frac{\mu_n^{(\theta_0)}(W)}{\mu_n^{(\theta)}(1_{\mathbb{X}})\, \mu_n^{(\theta_0)}(1_{\mathbb{X}})}. \tag{21b}$$

Proof. It is sufficient to prove (14a) with $U = W$, see Remark 2.1. Then we have for every measurable function $f : \mathbb{X} \to \mathbb{R}$ such that $|f| \leq W$

$$\begin{aligned}
\forall n \geq 2, \quad \big|\mu_n^{(\theta)}(f) - \mu_n^{(\theta_0)}(f)\big| \quad &\leq \quad \sum_{k=2}^{n} \big|\beta_k^{(\theta)}(f) - \beta_k^{(\theta_0)}(f)\big| \quad \text{(from (11))} \\
&= \sum_{k=2}^{n} \Big| \int_{\mathbb{X}} (R_\theta^{k-1} f)(x) - (R_{\theta_0}^{k-1} f)(x)\, d\nu(x)\Big| \quad \text{(from (20))} \\
&\leq \quad \sum_{k=1}^{n-1} \int_{\mathbb{X}} \big|(R_\theta^{k} f)(x) - (R_{\theta_0}^{k} f)(x)\big|\, d\nu(x) \\
&\leq \quad \sum_{k=1}^{n-1} \sum_{j=0}^{k-1} \nu\big(R_\theta^{j} \Delta_{\theta,W}\big)
\end{aligned}$$

from (18) in Lemma 2.1 applied to $R_0 = R_{\theta_0}$ and $R_1 = R_\theta$, observing moreover that $\Delta_{\theta,W}$ in (13) is also given by

$$\forall x \in \mathbb{X}, \quad \Delta_{\theta,W}(x) = \|R_\theta(x,\cdot) - R_{\theta_0}(x,\cdot)\|_W.$$

Then (14a) follows from (20) and from the definition of the W-weighted total variation norm (moreover exchange the role of θ and θ_0 in the previous inequality to obtain the complete form of (14a)). Finally let us prove (21a). We have for every measurable function $f : \mathbb{X} \to \mathbb{R}$ such that $|f| \leq W$

$$\begin{aligned}
\big|\widetilde{\mu}_n^{(\theta)}(f) - \widetilde{\mu}_n^{(\theta_0)}(f)\big| \quad &\leq \quad \frac{1}{\mu_n^{(\theta)}(1_{\mathbb{X}})} \big|\mu_n^{(\theta)}(f) - \mu_n^{(\theta_0)}(f)\big| + |\mu_n^{(\theta_0)}(f)| \left| \frac{1}{\mu_n^{(\theta)}(1_{\mathbb{X}})} - \frac{1}{\mu_n^{(\theta_0)}(1_{\mathbb{X}})} \right| \\
&\leq \quad \frac{1}{\mu_n^{(\theta)}(1_{\mathbb{X}})} \|\mu_n^{(\theta)} - \mu_n^{(\theta_0)}\|_W + \mu_n^{(\theta_0)}(W)\, \frac{\big|\mu_n^{(\theta_0)}(1_{\mathbb{X}}) - \mu_n^{(\theta)}(1_{\mathbb{X}})\big|}{\mu_n^{(\theta)}(1_{\mathbb{X}})\, \mu_n^{(\theta_0)}(1_{\mathbb{X}})} \\
&\leq \quad \frac{1}{\mu_n^{(\theta)}(1_{\mathbb{X}})} \|\mu_n^{(\theta)} - \mu_n^{(\theta_0)}\|_W + \frac{\mu_n^{(\theta_0)}(W)}{\mu_n^{(\theta)}(1_{\mathbb{X}})\, \mu_n^{(\theta_0)}(1_{\mathbb{X}})} \|\mu_n^{(\theta_0)} - \mu_n^{(\theta)}\|_{TV}
\end{aligned}$$

from which we deduce (21a).

□

Proof of Theorem 2.1. Property (14a) has been proved in Lemma 2.2. Next we have

$$\forall\theta \in \Theta,\ \forall n \geq 1, \quad \|\pi_\theta - \pi_{\theta_0}\|_W \leq \|\pi_\theta - \widetilde{\mu}_n^{(\theta)}\|_W + \|\widetilde{\mu}_n^{(\theta)} - \widetilde{\mu}_n^{(\theta_0)}\|_W + \|\widetilde{\mu}_n^{(\theta_0)} - \pi_{\theta_0}\|_W.$$

Then (14b) follows from the definition of $\widetilde{\varepsilon}_{n,\Theta,W}$ in (12) and from (21a). Now prove (15) under the additional assumptions $(\boldsymbol{\Delta_W})$-$(\boldsymbol{\mathcal{E}_W})$. We have for every $\theta \in \Theta$ and every $n \geq 2$

$$\begin{aligned}
\|\pi_\theta - \pi_{\theta_0}\|_W \quad &\leq \quad 2\,\widetilde{\varepsilon}_{n,\Theta,W} + (a_n + b_n)\, \|\mu_n^{(\theta)} - \mu_n^{(\theta_0)}\|_W \\
&\leq \quad 2\,\widetilde{\varepsilon}_{n,\Theta,W} + \nu(1_{\mathbb{X}})^{-1}\big(1 + \nu(1_{\mathbb{X}})^{-1}\mu_n^{(\theta_0)}(W)\big)\, \|\mu_n^{(\theta)} - \mu_n^{(\theta_0)}\|_W \\
&\leq \quad 2\,\widetilde{\varepsilon}_{n,\Theta,W} + \nu(1_{\mathbb{X}})^{-1}\big(1 + \nu(1_{\mathbb{X}})^{-1}\mu_n^{(\theta_0)}(W)\big)\, (n-1)\, \mu^{(\theta_0)}(\Delta_{\theta,W})
\end{aligned} \tag{22}$$

from $\mu_n^{(\theta)}(1_{\mathbb{X}}) \geq \mu_1^{(\theta)}(1_{\mathbb{X}}) = \beta_1^{(\theta)}(1_{\mathbb{X}}) = \nu(1_{\mathbb{X}})$, and from (14a) and $\mu_k^{(\theta_0)} \leq \mu^{(\theta_0)}$. Recall that $\mu^{(\theta_0)} \in \mathcal{M}^+$ and that

$$\forall \theta \in \Theta, \quad \Delta_{\theta,W} \leq 2CW \tag{23}$$

from ($\boldsymbol{W_\Theta}$). Then, under Assumption ($\boldsymbol{\Delta_W}$), we obtain that

$$\lim_{\theta \to \theta_0} \mu^{(\theta_0)}(\Delta_{\theta,W}) = 0$$

from Lebesgue's theorem since $\mu^{(\theta_0)}(W) < \infty$ (use $\pi_{\theta_0}(W) < \infty$ and (9)). It follows from (22) that

$$\forall n \geq 2, \quad \limsup_{\theta \to \theta_0} \|\pi_\theta - \pi_{\theta_0}\|_W \leq 2\, \widetilde{\varepsilon}_{n,\Theta,W}.$$

Then Assumption ($\boldsymbol{\mathcal{E}_W}$) gives (15) since n is arbitrarily large in the previous inequality. □

2.2 Convergence rates

In practice the stationary distribution π_{θ_0} of P_{θ_0} is supposed to be unknown and to be not directly computable. By contrast the stationary distribution π_θ of the perturbed transition kernel P_θ is expected to be computable and then to provide an approximation of π_{θ_0}. In this context, Inequalities (14a)-(14b) of Theorem 2.1 can be used to obtain an explicit control of $\|\pi_\theta - \pi_{\theta_0}\|_W$, provided that the functions $\Delta_{\theta,W}(\cdot)$ and $\Delta_{\theta,1_{\mathbb{X}}}(\cdot)$ defined in (13) are also supposed to be computable, so that the real numbers $\gamma_{\theta,W} := \pi_\theta(\Delta_{\theta,W})$ and $\gamma_\theta := \pi_\theta(\Delta_{\theta,1_{\mathbb{X}}})$ introduced below are available. The bound in (14b) also depends on the term $\widetilde{\varepsilon}_{n,\Theta,W}$ defined in (12), which is investigated in the next section under a polynomial drift condition, see also Remark 2.2.

Corollary 2.1 *Assume that the assumptions of Theorem 2.1 hold and that there exists an integer $n^* \geq 2$ such that*

$$\forall n \geq n^*, \quad \varepsilon_{n,\Theta,1_{\mathbb{X}}} := \sup_{\theta\in\Theta} \left[\mu^{(\theta)}(1_{\mathbb{X}}) - \mu_n^{(\theta)}(1_{\mathbb{X}})\right] \leq \frac{1}{2}. \tag{24}$$

Then the following inequalities hold for every $\theta \in \Theta$ and for every $n \geq n^*$:

$$\|\pi_\theta - \pi_{\theta_0}\|_{TV} \leq 2\, \widetilde{\varepsilon}_{n,\Theta,1_{\mathbb{X}}} + 4(n-1)\, \gamma_\theta \tag{25a}$$

$$\|\pi_\theta - \pi_{\theta_0}\|_W \leq 2\, \widetilde{\varepsilon}_{n,\Theta,W} + 2(n-1)\big(\gamma_{\theta,W} + \ 2\pi_{\theta_0}(W)\gamma_\theta\big) \tag{25b}$$

$$\textit{with} \quad \gamma_\theta := \pi_\theta(\Delta_{\theta,1_{\mathbb{X}}}) \quad \textit{and} \quad \gamma_{\theta,W} := \pi_\theta(\Delta_{\theta,W}). \tag{25c}$$

If moreover Assumptions ($\boldsymbol{\Delta_W}$) and ($\boldsymbol{\mathcal{E}_W}$) hold, then we have

$$\lim_{\theta \to \theta_0} \gamma_\theta = \lim_{\theta \to \theta_0} \gamma_{\theta,W} = 0.$$

Proof. It follows from the definition (24) of n^* that

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

since $\mu^{(\theta)}(1_{\mathbb{X}}) = 1/\pi_\theta(1_S) \geq 1$, see (9). Hence

$$\forall n \geq n^*,\ \forall \theta \in \Theta, \quad \frac{1}{\mu_n^{(\theta)}(1_{\mathbb{X}})} \leq 2\, \pi_\theta(1_S).$$

Moreover Property (14a) and $\mu_k^{(\theta)} \leq \mu^{(\theta)}$ give for $U \in \{1_{\mathbb{X}}, W\}$

$$\|\mu_n^{(\theta)} - \mu_n^{(\theta_0)}\|_U \leq (n-1)\mu^{(\theta)}(\Delta_{\theta,U}).$$

Then, applying (14b) and the two last inequalities, we obtain that for every $n \geq n^*$ and for every $\theta \in \Theta$

$$\begin{aligned}\|\pi_\theta - \pi_{\theta_0}\|_W &\leq 2\,\widetilde{\varepsilon}_{n,\Theta,W} + 2(n-1)\Big(\pi_\theta(1_S)\,\mu^{(\theta)}(\Delta_{\theta,W}) + 2\pi_\theta(1_S)\pi_{\theta_0}(1_S)\mu^{(\theta_0)}(W)\mu^{(\theta)}(\Delta_{\theta,1_{\mathbb{X}}})\Big)\\ &\leq 2\,\widetilde{\varepsilon}_{n,\Theta,W} + 2(n-1)\big(\pi_\theta(\Delta_{\theta,W}) + 2\pi_{\theta_0}(W)\pi_\theta(\Delta_{\theta,1_{\mathbb{X}}})\big)\end{aligned}$$

from (9). This proves (25b). Similarly, starting from (16) (in place of (14b)), we obtain (25a). Next assume that Assumptions *($\boldsymbol{\Delta_W}$)* and *($\boldsymbol{\mathcal{E}_W}$)* hold. Note that

$$\gamma_{\theta,W} = \pi_\theta(\Delta_{\theta,W}) \leq \big|\pi_\theta(\Delta_{\theta,W}) - \pi_{\theta_0}(\Delta_{\theta,W})\big| + \pi_{\theta_0}(\Delta_{\theta,W}) \leq 2C\|\pi_\theta - \pi_{\theta_0}\|_W + \pi_{\theta_0}(\Delta_{\theta,W})$$

since $\Delta_{\theta,W} \leq 2CW$, see (23). Hence $\lim_{\theta\to\theta_0}\gamma_{\theta,W} = 0$ from (15) and from Lebesgue's theorem with respect to the probability measure π_{θ_0} (recall that $\pi_{\theta_0}(W) < \infty$ from Assumption ($\boldsymbol{W_\Theta}$)). Finally we have $\lim_{\theta\to\theta_0}\gamma_\theta = 0$ since $\gamma_\theta \leq \gamma_{\theta,W}$. □

Remark 2.2 *At this stage of the exposition, when $\{P_\theta\}_{\theta\in\Theta}$ is a family of perturbed geometrically or polynomially ergodic Markov kernels, then the bounds (25a)-(25b) of Corollary 2.1 combined with the estimates of $\widetilde{\varepsilon}_{n,\Theta,W}$ derived from [HL22b] can be used to control the error term $\|\pi_\theta - \pi_{\theta_0}\|_W$ for some suitable functions W linked to the drift conditions. In this work we are mainly concerned with the polynomial case, see Section 3. Indeed, in the case of $V-$geometrically ergodic Markov kernels, the norm $\|\pi_\theta - \pi_{\theta_0}\|_W$ with $W = V^{\alpha_0}$ for some suitable $\alpha_0 \in (0,1]$ can be simply and efficiently controlled from $\gamma_{\theta,W} = \int_{\mathbb{X}} \|P_{\theta_0}(x,\cdot) - P_\theta(x,\cdot)\|_W\, \pi_\theta(dx)$ combining a spectral approach with solutions to Poisson's equation (see [LL18, Th. 2] for irreducible and positive recurrent discrete Markov kernels and [HL22a, Th. 6.1] in a general context). So, we are not going in that direction. What can be done using Corollary 2.1 in the context of perturbed $V-$geometrically ergodic Markov kernels is postponed in Appendix A for completeness. In particular we show in Appendix A that our general estimates are very close to Tweedie's (see (2)) when applied to discrete truncation issues.*

3 Applications to polynomially ergodic Markov kernels

Under Assumption ($\boldsymbol{S_\Theta}$), let us introduce the following condition for some positive integer m: there exists a collection $\{V_i\}_{i=0}^m$ of Lyapunov functions (i.e. $V_i : \mathbb{X} \to [1, +\infty)$ is measurable) with $V_m = 1_{\mathbb{X}}$ and $\forall\theta \in \Theta,\ \forall x \in \mathbb{X},\ (P_\theta V_0)(x) < \infty$, such that

$$\forall\theta \in \Theta,\ \forall i \in \{0,\ldots,m-1\},\quad (P_\theta - T)V_i \leq V_i - V_{i+1} \quad \text{with} \quad T(\cdot) := \nu(\cdot)1_S. \tag{26}$$

Since $P_\theta - T \geq 0$ from ($\boldsymbol{S_\Theta}$), the properties (26) imply that

$$V_m = 1_{\mathbb{X}} \leq V_{m-1} \leq \cdots \leq V_1 \leq V_0.$$

Moreover we have $\nu(V_0) < \infty$. For any positive integer j define

$$C_j := 2^{\frac{j(j+1)}{2}-1} \quad \text{and} \quad D_j := 2^{\frac{(j+1)(j+2)}{2}+1}. \tag{27}$$

The next estimates for the terms $\widetilde{\varepsilon}_{n,\Theta,1_{\mathbb{X}}}$ and $\widetilde{\varepsilon}_{n,\Theta,V_j}$ defined in (12) and used in Estimates (25a)-(25b) are obtained from the bounds in [HL22b, Eq. (24), (27) and Cor. 5.1] applied to P_θ, which only depend on $\nu(V_0)$ with $\nu \in \mathcal{M}_*^+$ (independent of $\theta \in \Theta$) given in $(\boldsymbol{S_\Theta})$ and on the positive constants $\mu^{(\theta)}(V_j)$. Hence, for $j = 0, \ldots, m$ we define (a priori in $[0, +\infty]$)

$$\vartheta_j := \sup_{\theta \in \Theta} \mu^{(\theta)}(V_j).$$

Complements on these constants are provided in Remark 3.1.

Theorem 3.1 *Assume that the family $\{P_\theta\}_{\theta\in\Theta}$ satisfies Assumption $(\boldsymbol{S_\Theta})$ and Conditions (26) for some $m \geq 1$ with respect to some collection $\{V_i\}_{i=0}^m$ of Lyapunov functions with $V_m = 1_{\mathbb{X}}$ and $\forall \theta \in \Theta$, $\forall x \in \mathbb{X}$, $(P_\theta V_0)(x) < \infty$. Also assume that for every $\theta \in \Theta$ there exists a unique P_θ−invariant probability measure π_θ on $(\mathbb{X}, \mathcal{X})$ such that $\pi_\theta(1_S) > 0$. Then the following assertions hold.*

(a) If $m \geq 2$, then Estimate (25a) holds with

$$\forall n \geq n^*, \quad \widetilde{\varepsilon}_{n,\Theta,1_{\mathbb{X}}} \leq \frac{2\, C_m\, \nu(V_0)}{m-1} \, \frac{1}{n^{m-1}}$$
$$\text{where} \quad n^* = \max\Big(2\,, \big\lfloor \big(2(m-1)^{-1} C_m \nu(V_0)\big)^{1/(m-1)} \big\rfloor + 1\Big). \tag{28}$$

(b) If $m \geq 3$, then for every $j = 2, \ldots, m-1$ and every $\theta \in \Theta$ we have $\pi_\theta(V_j) \leq \mu^{(\theta)}(V_j) < \infty$. Moreover we have $\vartheta_j < \infty$, and Estimate (25b) holds with $W = V_j$ and

$$\forall n \geq n^*, \quad \widetilde{\varepsilon}_{n,\Theta,V_j} \leq \frac{C_j\, \nu(V_0)}{j-1} \, \frac{1}{n^{j-1}} + 2\, \vartheta_j \, \frac{C_m\, \nu(V_0)}{m-1} \, \frac{1}{n^{m-1}}.$$

(c) If $m \geq 1$ and $\vartheta_0 < \infty$, then Estimate (25a) holds with

$$\forall n \geq n^{**}, \quad \widetilde{\varepsilon}_{n,\Theta,1_{\mathbb{X}}} \leq \frac{2 D_m \vartheta_0}{m} \, \frac{1}{n^m}$$
$$\text{where} \quad n^{**} = \max\Big(2\,, \big\lfloor \big(2m^{-1} D_m \vartheta_0\big)^{1/m} \big\rfloor + 1\Big). \tag{29}$$

(d) If $m \geq 2$ and $\vartheta_0 < \infty$, then for every $j = 1, \ldots, m-1$ we have $\vartheta_j \leq \vartheta_0$, and Estimate (25b) holds with $W = V_j$ and

$$\forall n \geq n^{**}, \quad \widetilde{\varepsilon}_{n,\Theta,V_j} \leq \frac{D_j\, \vartheta_0}{j} \, \frac{1}{n^j} + 2\, \vartheta_j \, \frac{D_m\, \vartheta_0}{m} \, \frac{1}{n^m}.$$

Remark 3.1 *Note that for $j = 0, \ldots, m-1$ the second condition of $(\boldsymbol{W_\Theta})$ holds with $W = V_j$ from (26) and $\nu(V_j) < \infty$. Also mention that n^* in (28) (resp. n^{**} in (29)) satisfies (24) under the assumptions of Assertion (a) (resp. Assertion (c)) since for every $\theta \in \Theta$ the quantity $\varepsilon_n := \mu^{(\theta)}(1_{\mathbb{X}}) - \mu_n^{(\theta)}(1_{\mathbb{X}})$ satisfies from [HL22b, Cor. 5.1]*

$$\varepsilon_n \leq \frac{C_m\, \nu(V_0)}{m-1} \, \frac{1}{n^{m-1}} \quad \text{in Assertion } (a) \quad \text{and} \quad \varepsilon_n \leq \frac{D_m \vartheta_0}{m} \, \frac{1}{n^m} \quad \text{in Assertion } (c).$$

Moreover the following bound for the constant ϑ_j can be used in Assertions (b) of Theorem 3.1 (see the proof [HL22b, Cor. 5.1]):

$$\vartheta_j = \sup_{\theta\in\Theta} \sum_{k=1}^{+\infty} \beta_k^{(\theta)}(V_j) \le C_j\, \nu(V_0) \sum_{k=1}^{+\infty} \frac{1}{k^j}. \tag{30}$$

Also recall that $\mu^{(\theta)}(V_j) = \mu^{(\theta)}(1_{\mathbb{X}})\pi_\theta(V_j) = \pi_\theta(V_j)/\pi_\theta(1_S)$ from (9). In particular Condition $\vartheta_0 < \infty$ in Assertions (c)-(d) of Theorem 3.1 holds if, and only if, $a := \inf_{\theta\in\Theta} \pi_\theta(1_S) > 0$ and $L := \sup_{\theta\in\Theta} \pi_\theta(V_0) < \infty$, in which case we have $\vartheta_0 \le L/a$. Moreover, in Assertions (a) and (c), we have $\lim_{\theta\to\theta_0} \|\pi_\theta - \pi_{\theta_0}\|_{TV} = 0$ from Corollary 2.1, provided that Assumption ($\boldsymbol{\Delta_W}$) holds with $W = 1_{\mathbb{X}}$. Similarly, in Assertion (b) and (d) we have $\lim_{\theta\to\theta_0} \|\pi_\theta - \pi_{\theta_0}\|_{V_j} = 0$ provided that Assumption ($\boldsymbol{\Delta_W}$) holds with $W = V_j$ (with the condition on j given in Assertion (b) and (d) respectively). The smaller j is, the larger the function $W = V_j$ is in (25b), but the worse bound of $\widetilde{\varepsilon}_{n,\Theta,V_j}$ is. Consequently in practice, for a given measurable function $f : \mathbb{X} \to \mathbb{R}$ such that f/V_1 is bounded, the best bound for $|\pi_\theta(f) - \pi_{\theta_0}(f)|$ which can be derived from Assertions (b) or (d) is obtained by choosing the greatest integer $j \le m$ (i.e. the smallest function V_j) such that $|f|/V_j$ is bounded. Of course, if f is bounded (i.e. f/V_m is bounded), then the best bound for $|\pi_\theta(f) - \pi_{\theta_0}(f)|$ is provided by (a) (or by (c) if $\vartheta_0 < \infty$). Also observe that the bound of $\widetilde{\varepsilon}_{n,\Theta,V_j}$ in Assertion (b) does not apply to $j = 1$. By contrast, under the condition $\vartheta_0 < \infty$, the bound of $\widetilde{\varepsilon}_{n,\Theta,V_j}$ in Assertion (d) applies to $j = 1$.

Application to truncation-augmentation of discrete Markov kernels

Let $P := (P(x,y))_{(x,y)\in\mathbb{N}^2}$ be a Markov kernel on $\mathbb{X} := \mathbb{N}$. For any $k \ge 1$ let $B_k := \{0, \ldots, k\}$. We assume that there exists a finite subset $S \subset \mathbb{N}$ and $\nu \in \mathcal{M}_*^+$ with finite support $\mathrm{Supp}(\nu) \subset \mathbb{N}$ such that

$$\forall x \in \mathbb{N},\ \forall A \subset \mathbb{N}, \quad P(x, A) \ge \nu(1_A)\, 1_S(x). \tag{$\boldsymbol{S}$}$$

Moreover we assume that there exists an unbounded and non-decreasing sequence $V := (V(x))_{x\in\mathbb{N}}$ with $V(0) = 1$ such that

$$M := \sup_{x\in S} (PV)(x) < \infty \tag{$\boldsymbol{M}$}$$

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

In the present context, Inequality ($\boldsymbol{DJ_{S^c}}$) is the polynomial drift condition introduced in [JR02] and is nothing else than Inequality (4) in the introduction. This condition was used with $c = 1$ and $S = \{0\}$ in [Liu10, Th. 5.1]. Recall that this polynomial drift condition has been generalized in [DFMS04] to cover general subgeometric rates of the convergence of the iterates P^n to $\pi(\cdot)\, 1_{\mathbb{X}}$, see also [DMPS18, Sect. 17.2]. Finally P is assumed to have a unique invariant probability measure π such that $\pi(1_S) > 0$.

Now, for $k \ge 1$ set ${B_k}^c := \mathbb{N} \setminus B_k$ and let us consider the k-th truncated and arbitrary augmented matrix P_k of the $(k+1) \times (k+1)$ north-west corner truncation of P:

$$\forall (x,y) \in {B_k}^2, \quad P_k(x,y) := P(x,y) + P(x, {B_k}^c)\, \psi_{x,k}(y) \tag{31}$$

where $\psi_{x,k}(\cdot)$ is some probability measure on B_k. When $\psi_{x,k}(\cdot) \equiv \psi_k(\cdot)$ only depends on k then this is referred as to a linear augmentation. When $\psi_{x,k}(\cdot) = \delta_0(\cdot)$ or $\psi_{x,k}(\cdot) = \delta_k(\cdot)$

then we obtain the first/last column linear augmentation used in [Twe98]. The goal here is to prove that, if P satisfies Assumptions $(\boldsymbol{S})$-$(\boldsymbol{M})$-$(\boldsymbol{DJ_{S^c}})$, then its invariant probability measure π can be approximated by the P_k-invariant probability measure π_k, with an explicit error control in function of the integer k. Since P is an infinite matrix, we first define the following extended Markov kernel $\widehat{P}_k$ of P_k on $\mathbb{N}$:

$$\forall (x,y) \in \mathbb{N}^2, \quad \widehat{P}_k(x,y) := \begin{cases} P_k(x,y) & \text{if } (x,y) \in {B_k}^2 \\ 1 & \text{if } y=0 \text{ and } x>k \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, if π_k is a P_k-invariant probability measure on B_k, then we define the extended probability measure $\widehat{\pi}_k$ on $\mathbb{N}$ by

$$\forall x \in \mathbb{N}, \quad \widehat{\pi}_k(\{x\}) := \begin{cases} \pi_k(\{x\}) & \text{if } x \in B_k \\ 0 & \text{if } x \notin B_k. \end{cases} \tag{32}$$

For every $k \geq 1$, let us introduce

$$\forall x \in \mathbb{N}, \quad \Delta_k(x) := \|P(x,\cdot) - \widehat{P}_k(x,\cdot)\|_{TV} \tag{33}$$

and, if π_k is a P_k-invariant probability measure on B_k, define

$$\gamma_k := \sum_{x \in B_k} \pi_k(x)\, \Delta_k(x). \tag{34}$$

Finally let $k_0 \in \mathbb{N}$ be the smallest integer such that

$$S \subset B_{k_0} \quad \text{and} \quad \text{Supp}(\nu) \subset B_{k_0}. \tag{35}$$

The main focus here is on the comparison of our results with Liu's work [Liu10], so that the finite set S in Assumptions $(\boldsymbol{S})$-$(\boldsymbol{M})$-$(\boldsymbol{DJ_{S^c}})$ is assumed to be an atom in the next Theorem 3.2 which is based on [HL22b, Cor. 5.4], see Remark 3.3 for the non-atomic case. More specifically assume that $P := (P(x,y))_{(x,y)\in\mathbb{N}^2}$ satisfies Assumptions $(\boldsymbol{S})$-$(\boldsymbol{M})$-$(\boldsymbol{DJ_{S^c}})$ with S supposed to be an atom satisfying Condition (35). Next define the following positive integer

$$m := \lfloor (1-\alpha)^{-1} \rfloor \tag{36}$$

with $\alpha \in [0,1)$ given in $(\boldsymbol{DJ_{S^c}})$ and where $\lfloor \cdot \rfloor$ denotes the integer part function. We know from [HL22b, Subs. 5.2] that the single Markov kernel P satisfies Condition (26) with respect to the following family $\{V_i\}_{i=0}^m$ of Lyapunov functions

$$V_0 = \Big[\prod_{k=1}^{m} c_k\Big]^{-1} V, \qquad \forall 1 \leq i \leq m-1: \ V_i = \Big[\prod_{k=i+1}^{m} c_k\Big]^{-1} V^{\alpha_i}, \qquad V_m = 1_{\mathbb{X}} \tag{37}$$

with $c_1 := c \in (0,1)$ (we can choose $c \in (0,1)$ in $(\boldsymbol{DJ_{S^c}})$) and some explicit $\{c_i\}_{i=2}^m \in (0,1)^{m-1}$, and with $0 < \alpha_{m-1} < \cdots < \alpha_2 < \alpha_1 < 1$ recursively defined by $\alpha_0 = 1$, $\alpha_1 := 1 - 1/m \in [0,1)$, and

$$\forall i = 2,\ldots,m-1, \quad \alpha_i = 2\alpha_{i-1} - \alpha_{i-2} = (\alpha_1 - 1)\, i + 1,$$

see [HL22b, Cor. 5.4 and its proof] for details. We have

$$\forall i = 0,\dots,m-1, \quad \lim_{x\to+\infty} \big[V_i(x) - V_{i+1}(x)\big] = 0$$

since $\alpha_{i+1} < \alpha_i$ and $V(x) \nearrow +\infty$ when $x \to +\infty$ by hypothesis. Hence there exists a positive integer k_1 such that

$$\forall i = 0,\dots,m-1,\ \forall x \in \mathbb{N} \cup [k_1,+\infty), \quad V_i(x) - V_{i+1}(x) \geq V_i(0). \tag{38}$$

Now assume that P has a unique invariant probability measure π such that $\pi(1_S) > 0$ and that for every $k \geq k_{\max} := \max(k_0,k_1)$, the matrix P_k in (31) admits a unique P_k-invariant probability measure π_k on B_k such that $\pi_k(1_S) > 0$. Then the next Lemmas 3.1-3.3 show that $\widehat{\pi}_k$ defined in (32) is the unique $\widehat{P}_k-$invariant probability measure, and that the whole family $\{P_\theta\}_{\theta\in\Theta}$ with

$$\Theta := \big([k_{\max},+\infty) \cap \mathbb{N}\big) \cup \{\infty\}, \quad P_\infty := P, \quad \forall \theta \geq k_{\max},\ P_\theta := \widehat{P}_k, \tag{39}$$

satisfies Assumption ($\boldsymbol{S_\Theta}$) with S,ν given in ($\boldsymbol{S}$) and Condition (26) w.r.t. the Lyapunov functions $\{V_i\}_{i=0}^m$ defined in (37). Moreover Assumption ($\boldsymbol{\Delta_W}$) is fulfilled with $W = V_0$. Accordingly all the conclusions of Theorem 3.1 hold with $\pi_{\theta_0} = \pi_\infty = \pi$ and $\pi_\theta = \widehat{\pi}_k$ for $k \geq k_{\max}$. This provides an explicit control for $\|\widehat{\pi}_k - \pi\|_{TV}$ or for $\|\widehat{\pi}_k - \pi\|_{V_j}$ according to the value of m in (36), which only depends on $\alpha \in [0,1)$ in ($\boldsymbol{DJ_{S^c}}$). The next statement only focusses on the error bound in total variation distance $\|\widehat{\pi}_k - \pi\|_{TV}$ to fit the framework of [Liu10].

Theorem 3.2 *Assume that $P := (P(x,y))_{(x,y)\in\mathbb{N}^2}$ satisfies Assumptions ($\boldsymbol{S}$)-($\boldsymbol{M}$)-($\boldsymbol{DJ_{S^c}}$) with S supposed to be an atom and to satisfy Condition (35). Let k_0 and k_1 be given in (35) and (38) respectively, and set $k_{\max} := \max(k_0,k_1)$. Moreover assume that P has a unique invariant probability measure π such that $\pi(1_S) > 0$ and that for every $k \geq k_{\max}$ (up to pick a larger integer $k_{\max}$) the matrix P_k admits a unique P_k-invariant probability measure π_k on B_k such that $\pi_k(1_S) > 0$. Finally let m be defined by (36) and let V_0 be the Lyapunov function in (37). Then the following assertions holds.*

(a) For every $k \geq k_{\max}$, $\widehat{\pi}_k$ defined in (32) is the unique $\widehat{P}_k-$invariant probability measure.

(b) The sequence $\{\gamma_k\}_{k\geq 1}$ defined in (34) satisfies: $\lim_k \gamma_k = 0$.

(c) If $m \geq 2$, defining C_m as in (27) and the integer $n^* \geq 2$ as in (28), then we have:

$$\forall k \geq k_{\max},\ \forall n \geq n^*, \quad \|\widehat{\pi}_k - \pi\|_{TV} \leq \frac{4\,C_m\,\nu(V_0)}{m-1}\,\frac{1}{n^{m-1}} + 4(n-1)\,\gamma_k. \tag{40}$$

(d) If $m \geq 1$ and $L := \sup_{k\in\Theta} \widehat{\pi}_k(V_0) < \infty$, $a = \inf_{k\in\Theta} \pi_k(1_S) > 0$, then

$$\forall k \geq k_{\max},\ \forall n \geq n^{**}, \quad \|\widehat{\pi}_k - \pi\|_{TV} \leq \frac{4D_m\vartheta_0}{m}\,\frac{1}{n^m} + 4(n-1)\,\gamma_k \tag{41}$$

with $\vartheta_0 = L/a$ and where D_m and the integer $n^{**} \geq 2$ are defined as in (27) and (29).

The P_k−invariant probability measure π_k on B_k is supposed to be computable since P_k is a finite matrix. In this case the exact value of the real numbers $\gamma_k := \pi_k(\Delta_k)$ in (34) is computable from Formula (42) below. A bound of γ_k is given in Proposition 3.4. Finally note that, if the positive measure $\mu^{(\theta)}$ in (10) with Θ given here in (39) is known, then ϑ_0 in Assertion (d) is given by $\vartheta_0 := \sup_{\theta\in\Theta} \mu^{(\theta)}(V_0)$, provided that this quantity is finite, see Theorem 3.1 and Remark 3.1.

Proof. Assertion (a) is proved in Lemma 3.1 below. Note that we have $\forall k \geq k_0,\ \widehat{\pi}_k(1_S) = \pi_k(1_S)$ since $S \subset B_{k_0}$ from Assumption (35). Moreover we deduce from Lemma 3.1 that $\gamma_k = \widehat{\pi}_k(\Delta_k)$, thus $\gamma_k \equiv \pi_\theta(\Delta_{\theta,1_{\mathbb{X}}}) = \gamma_\theta$ using the notations from (39) and Corollary 2.1, see (25c). Then the conclusions (40) and (41) follow from (25a) combined with Assertions (a) and (c) of Theorem 3.1. Indeed Lemma 3.3 below shows that the whole family $\{P_\theta\}_{\theta\in\Theta}$ given in (39) satisfies the assumptions of Theorem 3.1, see also Remark 3.1 concerning the constant ϑ_0 in Assertion (d). Moreover we know from Theorem 3.1 that Assumption $(\boldsymbol{\mathcal{E}_W})$ holds with $W = 1_{\mathbb{X}}$, and Lemma 3.2 below shows that Assumption $(\boldsymbol{\Delta_W})$ holds with $W = 1_{\mathbb{X}}$. This proves Assertion (b) due to Corollary 2.1. □

Lemma 3.1 *Let $P := (P(x,y))_{(x,y)\in\mathbb{N}^2}$ be a Markov kernel on $\mathbb{N}$, let $k \geq 1$, and let P_k be the stochastic matrix P_k given in (31). If P_k admits a unique invariant probability measure π_k on B_k, then $\widehat{\pi}_k$ defined in (32) is the unique $\widehat{P}_k$−invariant probability measure.*

Proof. We deduce from the definitions of $\widehat{P}_k$ and $\widehat{\pi}_k$ that

$$\forall y \in {B_k}^c, \quad \sum_{x\in\mathbb{N}} \widehat{P}_k(x,y)\,\widehat{\pi}_k(\{x\}) = 0 = \widehat{\pi}_k(\{y\}).$$

Thus

$$\begin{aligned}\forall y \in B_k, \quad \sum_{x\in\mathbb{N}} \widehat{P}_k(x,y)\,\widehat{\pi}_k(\{x\}) &= \sum_{x\in B_k} \widehat{P}_k(x,y)\,\widehat{\pi}_k(\{x\}) \\ &= \sum_{x\in B_k} P_k(x,y)\,\pi_k(\{x\}) = \pi_k(\{y\}) = \widehat{\pi}_k(\{y\})\end{aligned}$$

using successively the definitions of $\widehat{\pi}_k$ and $\widehat{P}_k$, the P_k−invariance of π_k, and again the definition of $\widehat{\pi}_k$. We have proved that $\widehat{\pi}_k$ is a $\widehat{P}_k$−invariant probability measure. To prove the uniqueness, consider any $\widehat{P}_k$−invariant probability measure $\widehat{\eta} = (\widehat{\eta}(\{x\}))_{x\in\mathbb{N}}$. Then

$$\forall y \in {B_k}^c, \quad \widehat{\eta}(\{y\}) = \sum_{x\in\mathbb{N}} \widehat{P}_k(x,y)\,\widehat{\eta}(\{x\}) = 0$$

from the definition of $\widehat{P}_k$. Thus

$$\forall y \in B_k, \quad \widehat{\eta}(\{y\}) = \sum_{x\in\mathbb{N}} \widehat{P}_k(x,y)\,\widehat{\eta}(\{x\}) = \sum_{x\in B_k} \widehat{P}_k(x,y)\,\widehat{\eta}(\{x\}) = \sum_{x\in B_k} P_k(x,y)\,\widehat{\eta}(\{x\})$$

from the definition of $\widehat{P}_k$. Thus $\eta := (\widehat{\eta}(\{x\}))_{x\in B_k}$ is a P_k−invariant probability measure on B_k. This proves that $\widehat{\eta} = \widehat{\pi}_k$. □

Lemma 3.2 *Let $P := (P(x,y))_{(x,y)\in\mathbb{N}^2}$ be a Markov kernel on $\mathbb{N}$. For any $k \geq 1$, let $\Delta_k(\cdot)$ be given in (33). Then we have*

$$\forall k \geq 1,\ \forall x \in \mathbb{N}, \quad \Delta_k(x) = 2\big(1_{B_k}(x)P(x,{B_k}^c) + 1_{{B_k}^c}(x)P(x,\mathbb{N}^*)\big). \tag{42}$$

Moreover we have $\forall x \in \mathbb{N},\ \lim_k \Delta_k(x) = 0$, that is Assumption $(\boldsymbol{\Delta_W})$ holds with $W = 1_{\mathbb{X}}$.

Proof. For any $x \in B_k$ we have

$$\Delta_k(x) = \sum_{y\in\mathbb{N}} \big|P(x,y) - \widehat{P}_k(x,y)\big| = P(x, {B_k}^c) \sum_{y\in B_k} \psi_{x,k}(y) + P(x, {B_k}^c) = 2P(x, {B_k}^c) \quad (43)$$

from the definitions of $\widehat{P}_k$, P_k in (31) using that $\psi_{x,k}(B_k) = 1$. For any $x \in {B_k}^c$, we obtain from the definition of $\widehat{P}_k$ that

$$\Delta_k(x) = (1 - P(x,0)) + \sum_{y\in\mathbb{N}^*} P(x,y) = 2P(x,\mathbb{N}^*).$$

Thus Equality (42) holds for any $x \in \mathbb{N}$. Finally, for any $x \in \mathbb{N}$, the convergence to 0 of the sequence $\{\Delta_k(x)\}_{k\geq 1}$ easily follows from (42) and the convergence of $\sum_{y\in\mathbb{N}} P(x,y)$. □

Lemma 3.3 *If* $P := (P(x,y))_{(x,y)\in\mathbb{N}^2}$ *satisfies the assumptions of Theorem 3.2, then the family* $\{P_\theta\}_{\theta\in\Theta}$ *given in (39) satisfies the assumptions of Theorem 3.1.*

Proof. Let $k \geq k_0$. For every $x \in S$ and every $A \subset \mathbb{N}$ we have

$$\widehat{P}_k(x,A) \geq \sum_{y\in A\cap B_k} \widehat{P}_k(x,y) \geq \sum_{y\in A\cap B_k} P(x,y) = P(x, A\cap B_k) \geq \nu(A\cap B_k) = \nu(A)$$

using successively $x \in S \subset B_{k_0} \subset B_k$ and the definitions of $\widehat{P}_k$ and P_k, Assumption ($\boldsymbol{S}$), and finally $\mathrm{Supp}(\nu) \subset B_{k_0} \subset B_k$. This proves that the family $\{P_\theta\}_{\theta\in\Theta}$ in (39) satisfies Assumption ($\boldsymbol{S_\Theta}$) with S, ν given in ($\boldsymbol{S}$). Now let us prove that $\{P_\theta\}_{\theta\in\Theta}$ satisfies Condition (26) with respect to the family $\{V_i\}_{i=0}^m$ defined in (37). That the Markov kernel $P_\infty = P$ satisfies (26) w.r.t. $\{V_i\}_{i=0}^m$ is discussed before Theorem 3.2. Next, we have to prove that, for every $k \geq k_{\max}$, the Markov kernel $\widehat{P}_k$ satisfies (26) with respect to the same family $\{V_i\}_{i=0}^m$.

Let $W := (W(x))_{x\in\mathbb{N}}$ be any non-decreasing sequence with $W(0) \geq 1$. Let $k \geq 1$. We have

$$\begin{aligned}
\forall x \in B_k, \quad (\widehat{P}_k W)(x) &= \sum_{y\in B_k} P(x,y)\, W(y) + P(x,{B_k}^c) \sum_{y\in B_k} \psi_{x,k}(y) W(y) \\
&\leq \sum_{y\in B_k} P(x,y) W(y) + P(x,{B_k}^c) \Big[W(k) \sum_{y\in B_k} \psi_{x,k}(y) \Big] \\
&= \sum_{y\in B_k} P(x,y) W(y) + \sum_{y\in {B_k}^c} P(x,y)\, W(k) \\
&\leq \sum_{y\in\mathbb{N}} P(x,y) W(y) = (PW)(x) \qquad (44)
\end{aligned}$$

since for any $(y,z) \in B_k \times {B_k}^c$, $W(y) \leq W(k) \leq W(z)$ and since $\psi_{x,k}(\cdot)$ is a probability measure on B_k. Moreover we have

$$\forall x \in {B_k}^c, \quad (\widehat{P}_k W)(x) = W(0).$$

Note that for every $i = 0, \ldots, m-1$ the function $W = V_i$ is non-decreasing and such that $V_i(0) \geq 1$ since $V_i \geq V_m = 1_{\mathbb{X}}$. Let $k \geq k_1$. Then applying the previous inequalities to $W = V_i$ for any $i = 0, \ldots, m-1$ provides

$$\forall x \in B_k, \quad (\widehat{P}_k V_i)(x) \leq (PV_i)(x) \leq V_i(x) - V_{i+1}(x) + \nu(V_i) 1_S(x) \qquad (45)$$

since P satisfies (26), and

$$\forall x \in {B_k}^c, \quad (\widehat{P}_k V_i)(x) = V_i(0) \leq V_i(x) - V_{i+1}(x) \leq V_i(x) - V_{i+1}(x) + \nu(V_i)1_S(x) \tag{46}$$

from (38). This proves that $\widehat{P}_k$ satisfies (26).

Finally note that we have $\forall x \in \mathbb{X},\ (PV_0)(x) < \infty$ from $(\boldsymbol{M})$ and $(\boldsymbol{DJ_{S^c}})$. Then we obtain that $\forall k \geq k_{\max},\ \forall x \in \mathbb{X},\ (\widehat{P}_k V_0)(x) < \infty$ from (45)-(46) applied with $i = 0$. □

Remark 3.2 *Assertions (b) and (d) of Theorem 3.1 apply too under the assumptions of Theorem 3.2. Recall that the bounds (30) can be used for Assertions (b) of Theorem 3.1. Moreover set*

$$\Delta_{k,V_j}(x) = \|P(x,\cdot) - \widehat{P}_k(x,\cdot)\|_{V_j}.$$

Note that the term $\gamma_{\theta,V_j} := \pi_\theta(\Delta_{\theta,V_j})$ in Estimates (25b) is given here by

$$\gamma_{k,V_j} = \widehat{\pi}_k(\Delta_{k,V_j}) = \sum_{x\in B_k} \pi(\{x\})\Delta_{k,V_j}(x).$$

For completeness let us prove that $\forall x \in \mathbb{N},\ \lim_k \Delta_{k,V_j}(x) = 0$, so that $\lim_k \gamma_{k,V_j} = 0$ due to Corollary 2.1. Obviously it is sufficient to prove that $\forall x \in \mathbb{N},\ \lim_k \Delta_{k,V_0}(x) = 0$ since $V_j \leq V_0$. From the definition of $\widehat{P}_k$ and (31), we have for every $x \in B_k$

$$\begin{aligned}
\Delta_{k,V_0}(x) &= \sum_{y\in\mathbb{N}} \big|P(x,y) - \widehat{P}_k(x,y)\big|V_0(y) \\
&= P(x,{B_k}^c) \sum_{y\in B_k} \psi_{x,k}(y)V_0(y) + \sum_{y\in {B_k}^c} P(x,y)V_0(y) \\
&\leq P(x,{B_k}^c)\, V_0(k) + \sum_{y\in {B_k}^c} P(x,y)V_0(y) \\
&\leq \sum_{z\in {B_k}^c} P(x,z)V_0(z) + \sum_{y\in {B_k}^c} P(x,y)V_0(y) \ \leq 2 \sum_{y\in {B_k}^c} P(x,y)V_0(y)
\end{aligned} \tag{47}$$

since V_0 is non-decreasing and $\psi_{x,k}(B_k) = 1$. Moreover for any $x \in {B_k}^c$ we have

$$\Delta_{k,V_0}(x) = P(x,\mathbb{N}^*)V_0(x) + \sum_{y\in\mathbb{N}^*} P(x,y)V_0(y).$$

Now fix $x \in \mathbb{N}$. Then it follows from (47) applied to any $k > x$ that $\lim_k \Delta_{k,V_0}(x) = 0$ since $\sum_{y\in\mathbb{N}} P(x,y)V_0(y) = (PV_0)(x) < \infty$ from $(\boldsymbol{M})$ and $(\boldsymbol{DJ_{S^c}})$.

Remark 3.3 *The non-atomic case can be addressed in a similar way using [HL22b, Cor. 5.5], but in this case the analogue of the integer m and of the Lyapunov functions V_i require more preparation. More specifically, assume that $P := (P(x,y))_{(x,y)\in\mathbb{N}^2}$ satisfies Assumptions $(\boldsymbol{S})$-$(\boldsymbol{M})$-$(\boldsymbol{DJ_{S^c}})$ with some finite set S satisfying Condition (35). First of all, the biggest function V_0 is of the form $V_0 = c_0 V^{\eta_0}$ for some constants $c_0 > 0$ and $\eta_0 \in (0,1]$ from [HL22b, Cor. 5.5]. Note that the case $\eta_0 < 1$ is possible when S is not an atom. Then we define the integer $m := \lfloor \eta_0(1-\alpha)^{-1} \rfloor$. Observe that, if $\eta_0 < 1$, then m is not necessary positive, so that we have to assume that $\eta_0 \geq 1-\alpha$ (thus $m \geq 1$) to continue the construction of the Lyapunov functions V_i. From [HL22b, Cor. 5.5], if $\eta_0 \geq 1-\alpha$, then the single Markov kernel P satisfies*

Condition (26) w.r.t. some family $\{V_i\}_{i=0}^m$ of Lyapunov functions defined as in (37), but with V^{η_0} in place of V (see the proof of [HL22b, Cor. 5.5] for details on the construction of $\{V_i\}_{i=0}^m$ and use [HL22b, Cor. 5.2] to obtain that P satisfies (26) w.r.t. $\{V_i\}_{i=0}^m$). Consequently, when the assumptions of Theorem 3.2 holds with a non-atomic finite set S, then all the conclusions of Theorem 3.2 remain true with respect to the family $\{V_i\}_{i=0}^m$ defined in [HL22b, Cor. 5.5]: this can be established by repeating the proof of Theorem 3.2 based on Lemmas 3.1-3.3. Indeed note that, in the proof of Theorem 3.2, the fact that S is an atom has been only used to prove that the single Markov kernel P satisfies Condition (26) w.r.t. the family $\{V_i\}_{i=0}^m$ in (37) using [HL22b, Cor. 5.4]. Hence the only differences with the atomic case are the following. First the integer $m := \lfloor \eta_0(1-\alpha)^{-1}\rfloor$ may be zero (i.e. $\eta_0 < 1-\alpha$), in which case the construction of the Lyapunov functions V_i is not possible. Second, if $1-\alpha \leq \eta_0 < 1$, then the construction of $\{V_i\}_{i=0}^m$ is possible, but the analogue of the bounds (40)-(41) may be less accurate than in the atomic case since $m := \lfloor \eta_0(1-\alpha)^{-1}\rfloor$ may be smaller than the integer $\lfloor (1-\alpha)^{-1}\rfloor$. Finally note that, under the previous assumptions on P, Assertions (b) and (d) of Theorem 3.1 also apply with respect to the family $\{V_i\}_{i=0}^m$ defined in [HL22b, Cor. 5.5]. Again the bounds (30) can be used for Assertions (b), and we have $\lim_k \gamma_{k,V_j} = 0$, see Remark 3.2.

Remark 3.4 (Convergence of $\{\widehat{\pi}_k\}_{n\geq 0}$ to π in truncation approximation) *As already mentioned, Tweedie proved in [Twe98, Th 3.2] that the convergence in the V-weighted total variation norm takes place for the first-column linear augmentation (see (31) with $\psi_{x,k} = \delta_0$) of V-geometrically ergodic discrete Markov chains. Using regeneration methods, such a convergence is extended to V-geometrically or polynomially ergodic Markov chains with continuous state space in [IG22, Th 2] for a specific linear augmentation (that is the $k-$first columns augmentation, for some k, in the discrete state space case). Finally mention that the weak convergence in the case of general augmentation of continuous state space Markov chains has been recently addressed in [IGL22]. Note that in such context, the weak convergence does not provide the convergence in the total variation norm. The estimation of convergence rates is not discussed in [IG22, IGL22].*

To complete the estimates (40) and (41) of $\|\widehat{\pi}_k - \pi\|_{TV}$, let us provide a bound on $\gamma_k = \pi_k(\Delta_k)$.

Proposition 3.4 *Assume that $P := (P(i,j))_{(i,j)\in\mathbb{N}^2}$ satisfies Assumptions $(\boldsymbol{S})$-$(\boldsymbol{M})$-$(\boldsymbol{DJ_{S^c}})$. Let $k \geq 1$ be such that the matrix P_k in (31) admits a unique P_k-invariant probability measure π_k on B_k, and let γ_k defined in (34). Then*

$$\forall k \geq k_0, \quad \gamma_k \leq \frac{2(cM^\alpha + M)}{c} \times \frac{1}{V(k+1)^\alpha} \tag{48}$$

where α and c are given in $(\boldsymbol{DJ_{S^c}})$.

Proof. Note that $PV^\alpha \leq (PV)^\alpha$ from Jensen's inequality. From (43) we obtain for $k \geq k_0$

$$\begin{aligned}
\gamma_k &= 2\sum_{x\in B_k} \pi_k(\{x\})\, P\big(x, {B_k}^c\big) = 2\sum_{x\in B_k} \pi_k(\{x\})\, P\big(x, \{y\in\mathbb{N} : y \geq k+1\}\big) \qquad \text{(from (33))} \\
&= 2\sum_{x\in B_k} \pi_k(\{x\})\, P\big(x, \{y\in\mathbb{N} : V(y)^\alpha \geq V(k+1)^\alpha\}\big) \qquad (\text{since } V^\alpha \nearrow) \\
&\leq 2\sum_{x\in B_k} \pi_k(\{x\})\, \frac{(PV^\alpha)(x)}{V(k+1)^\alpha} \qquad \text{(from Markov's inequality)} \\
&\leq \frac{2}{V(k+1)^\alpha}\Big[M^\alpha + \sum_{x\in S^c\cap B_k} \pi_k(\{x\})\,(PV^\alpha)(x)\Big] \quad (\text{from } PV^\alpha \leq (PV)^\alpha,\ (\boldsymbol{M}),\ \pi_k(B_k)=1) \\
&\leq \frac{2}{V(k+1)^\alpha}\Big[M^\alpha + \sum_{x\in S^c\cap B_k} \pi_k(\{x\})\,V(x)^\alpha\Big] \\
&\quad (\text{since } \forall x\in S^c, (PV^\alpha)(x) \leq (PV(x))^\alpha \leq V(x)^\alpha \text{ from } (\boldsymbol{DJ_{S^c}})) \\
&\leq \frac{2}{V(k+1)^\alpha}\Big[M^\alpha + \sum_{x\in B_k} \pi_k(\{x\})\,V(x)^\alpha\Big] = \frac{2}{V(k+1)^\alpha}\big[M^\alpha + \pi_k({V^\alpha}_k)\big] \qquad (49)
\end{aligned}$$

where ${V^\alpha}_k := {V^\alpha}_{|B_k}$ is the restriction of V^α to B_k. Next, if $V_k := V_{|B_k}$, we have

$$\forall x\in B_k, \quad (P_kV_k)(x) \leq V_k(x) - c{V^\alpha}_k(x) + M$$

from Inequality (44) applied to $W = V$ and $(\boldsymbol{M})$-$(\boldsymbol{DJ_{S^c}})$. Then, by using that $\pi_k P_k = \pi_k$, it follows that

$$c\,\pi_k({V^\alpha}_k) \leq M.$$

Finally, combining (49) and the last inequality, we obtain (48).

□

Let us only discuss the least favourable case of small values of α in $(\boldsymbol{DJ_{S^c}})$, noticing that $m = 1$ when $\alpha \in (0, 1/2)$ and that $m = 2$ when $\alpha \in [1/2, 2/3)$. As we can see from the estimates (40)-(41) of $\|\pi_k - \pi\|_{TV}$ and from the estimate of γ_k in (48), we obtain a quite similar bound to (5) [Liu10]. Note that here P is not assumed to be stochastically monotone.

References

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

[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 Series in Operations Research and Financial Engineering. Springer, 2018.

[HL20] L. Hervé and J. Ledoux. V-geometrical ergodicity of Markov kernels via finite-rank approximations. *Electron. Commun. Probab.*, 25(23):1–12, March 2020.

[HL22a] L. Hervé and J. Ledoux. Explicit bounds for spectral theory of geometrically ergodic Markov kernels and applications. https://hal-cnrs.archives-ouvertes.fr/hal-03819315, 2022.

[HL22b] L. Hervé and J. Ledoux. Quantitative approximation of the invariant distribution of a Markov chain. a new approach. https://hal.archives-ouvertes.fr/hal-03605636, 2022.

[IG22] A. Infanger and P. G. Glynn. On convergence of a truncation scheme for approximating stationary distributions of Continuous state space Markov chains and processes. https://arxiv.org/abs/2206.11738, 2022.

[IGL22] A. Infanger, P. G. Glynn, and Y Liu. On convergence of general truncation-augmentation schemes for approximating stationary distributions of Markov chains. https://arxiv.org/abs/2203.15167, 2022.

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

[Liu10] Y. Liu. Augmented truncation approximations of discrete-time Markov chains. *Oper. Res. Lett.*, 38(3):218–222, 2010.

[LL18] Y. Liu and W. Li. Error bounds for augmented truncation approximations of Markov chains via the perturbation method. *Adv. in Appl. Probab.*, 50(2):645–669, 2018.

[Mas16] H. Masuyama. Error bounds for augmented truncations of discrete-time block-monotone Markov chains under subgeometric drift conditions. *SIAM J. Matrix Anal. Appl.*, 37(3):877–910, 2016.

[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.

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

[Twe98] R. L. Tweedie. Truncation approximations of invariant measures for Markov chains. *J. Appl. Probab.*, 35(3):517–536, 1998.

A Perturbation of V-geometrically ergogic Markov kernels

If the family $\{P_\theta\}_{\theta\in\Theta}$ satisfies Assumption $(\boldsymbol{S_\Theta})$ of Section 1 and the next Assumptions $(\boldsymbol{M_\Theta})$-$(\boldsymbol{D_{\Theta,S^c}})$, then the bounds for $\|\pi_\theta - \pi_{\theta_0}\|_{TV}$ and $\|\pi_\theta - \pi_{\theta_0}\|_{V^{\alpha_0}}$ obtained in [HL22a, Th. 6.1] with $\alpha_0 \in (0,1]$ given in (50) below are more relevant than those derived from Estimates (25a)-(25b) combined with the bound (52) of the next Theorem A.1, see Remark 2.2 and the bound (65) p. 23. Hence Theorem A.1 below is only given for completeness. The goal of this theorem is to show that the results of Section 2 also apply to V-geometrically ergogic Markov kernels and to prove that, when applied to truncation of stochastically monotone discrete Markov kernels, it then provides a bound for $\|\pi_\theta - \pi_{\theta_0}\|_{TV}$ which is similar to Tweedie's Estimate (3), see Remark A.3.

Assume that the family $\{P_\theta\}_{\theta\in\Theta}$ satisfies Assumptions ($\boldsymbol{S_\Theta}$) of Section 1 with respect to some small set $S \in \mathcal{X}$ and some $\nu \in \mathcal{M}_*^+$. Moreover assume that there exists a Lyapunov function $V : \mathbb{X} \to [1, +\infty)$ satisfying $V(0) = 1$ and such that the following conditions hold

$$M_\Theta := \sup_{\theta\in\Theta} \sup_{x\in S} (P_\theta V)(x) < \infty \qquad (\boldsymbol{M_\Theta})$$

$$\exists\delta \in (0,1),\ \forall\theta \in \Theta,\ \forall x \in S^c, \quad (P_\theta V)(x) \leq \delta\, V(x). \qquad (\boldsymbol{D_{\Theta,S^c}})$$

Assumptions ($\boldsymbol{M_\Theta}$)-($\boldsymbol{D_{\Theta,S^c}}$) ensures that the whole family $\{P_\theta\}_{\theta\in\Theta}$ satisfies the so-called geometric drift condition

$$\exists\delta \in (0,1),\ \exists b := M_\Theta > 0,\ \forall\theta \in \Theta,\ P_\theta V \leq \delta\, V + b1_S$$

with respect to the Lyapunov function V in a uniform way in $\theta \in \Theta$. Then for every $\theta \in \Theta$ there exists a unique $P_\theta-$invariant probability measure π_θ on $(\mathbb{X}, \mathcal{X})$ such that $\pi_\theta(V) < \infty$, e.g. see [MT93, RR04]. Also note that $\nu(V) < \infty$ from ($\boldsymbol{S_\Theta}$). We know from [HL22b, Cor. 4.2] that there exists a computable real number $\alpha_0 \in (0, 1]$ (see Remark A.1) such that

$$\forall\theta \in \Theta, \quad P_\theta V^{\alpha_0} \leq \delta^{\alpha_0}\, V^{\alpha_0} + \nu(V^{\alpha_0})1_S. \qquad (50)$$

Define

$$A := \frac{\nu(V^{\alpha_0})}{1-\delta^{\alpha_0}} \quad \text{and} \quad n^* = \max\left(2\,, \left\lfloor \frac{\ln(1-\delta^{\alpha_0}) - \ln\nu(V^{\alpha_0}) - \ln 2}{\alpha_0 \ln\delta} \right\rfloor + 1\right). \qquad (51)$$

The following bounds for $\widetilde{\varepsilon}_{n,\Theta,1_{\mathbb{X}}}$ and $\widetilde{\varepsilon}_{n,\Theta,V^{\alpha_0}}$ are obtained from [HL22b, (24), (27) and Cor. 4.2].

Theorem A.1 *Assume that the family $\{P_\theta\}_{\theta\in\Theta}$ satisfies Assumptions ($\boldsymbol{S_\Theta}$)-($\boldsymbol{M_\Theta}$)-($\boldsymbol{D_{\Theta,S^c}}$) and that for every $\theta \in \Theta$ we have $\pi_\theta(1_S) > 0$, Then*

$$\forall\theta \in \Theta,\ \forall n \geq n^*, \quad \widetilde{\varepsilon}_{n,\Theta,1_{\mathbb{X}}} \leq 2A\,\delta^{\alpha_0 n} \quad \textit{and} \quad \widetilde{\varepsilon}_{n,\Theta,V^{\alpha_0}} \leq A\,\big(1+2A\big)\,\delta^{\alpha_0 n}. \qquad (52)$$

Remark A.1 *The real number $\alpha_0 \in (0, 1]$ in (50) can be easily computed from Assumptions ($\boldsymbol{M_\Theta}$)-($\boldsymbol{D_{\Theta,S^c}}$) by using Jensen's inequality, see [HL22b, (35) and Prop. 4.1]. Actually the real number M_Θ plays an important role in the computation of α_0: roughly speaking, the larger M_Θ is compared to $\nu(V)$, the smaller α_0 is. If S is an atom in ($\boldsymbol{S_\Theta}$) with ν given by $\nu = P(s, \cdot)$ for some $s \in S$, then (50) holds with $\alpha_0 = 1$, see [HL22b, Cor. 4.1]. However the case $\alpha_0 = 1$ is not equivalent to the atomic case, in other words Property (50) may hold with $\alpha_0 = 1$ for non-atomic small set S, see [HL22b, Sec. 6].*

Remark A.2 *Under the assumptions of Theorem A.1, the function $W = V^{\alpha_0}$ satisfies the second condition of ($\boldsymbol{W_\Theta}$) from (50). Consequently Estimates (25a)-(25b) of Corollary 2.1 combined with (52) can be used to control $\|\pi_\theta - \pi_{\theta_0}\|_{V^{\alpha_0}}$, provided that Assumption ($\boldsymbol{\Delta_W}$) holds with $W = V^{\alpha_0}$, that is*

$$\forall x \in \mathbb{X}, \quad \lim_{\theta\to\theta_0} \|P_\theta(x, \cdot) - P_{\theta_0}(x, \cdot)\|_{V^{\alpha_0}} = 0$$

where $\alpha_0 \in (0,1]$ *is given by (50). In particular we have* $\lim_{\theta\to\theta_0}\gamma_\theta = \lim_{\theta\to\theta_0}\gamma_{\theta,V^{\alpha_0}} = 0$ *and*

$$\lim_{\theta\to\theta_0} \|\pi_\theta - \pi_{\theta_0}\|_{V^{\alpha_0}} = 0.$$

Finally note that n^* *given in (51) satisfies (24) since we know from [HL22b, Cor. 4.2] that*

$$\forall\theta\in\Theta, \quad \|\mu^{(\theta)} - \mu_n^{(\theta)}\|_{TV} \le A\,\delta^{\alpha_0 n}.$$

Applications to truncation-augmentation of discrete Markov kernels

Let $P := (P(i,j))_{(i,j)\in\mathbb{N}^2}$ be a Markov kernel on $\mathbb{X} := \mathbb{N}$. For any $k \ge 1$ let $B_k := \{0,\ldots,k\}$. We assume that there exists a finite subset $S \subset \mathbb{N}$ and $\nu \in \mathcal{M}_*^+$ with finite support $\mathrm{Supp}(\nu) \subset \mathbb{N}$ such that P satisfies Condition $(\boldsymbol{S})$, see p. 11. Moreover we assume that there exists an increasing sequence $V := \{V(i)\}_{i\in\mathbb{N}}$ with $V(0) = 1$ such that P satisfies Condition $(\boldsymbol{M})$ (see p. 11) and the following geometric drift condition

$$\exists\delta\in(0,1),\ \forall i\in S^c,\ (PV)(i) \le \delta\,V(i). \qquad (\boldsymbol{D_{S^c}})$$

Under these assumptions we know that there exists a unique $P-$invariant probability measure π on $\mathbb{N}$. We assume that $\pi(1_S) > 0$. For $k \ge 1$ let us consider the k-th truncated-augmented matrix P_k as defined in (31). The goal here is to prove that, if P satisfies Assumptions $(\boldsymbol{S})$-$(\boldsymbol{M})$-$(\boldsymbol{D_{S^c}})$, then its invariant probability measure π can be approximated by the P_k-invariant probability measure π_k, with an explicit error control in function of the integer k. Let $\widehat{P}_k$ be the extended Markov kernel of P_k on $\mathbb{N}$ defined in Section 3. Similarly, $\widehat{\pi}_k$ is the extended probability measure (32) of the P_k-invariant probability measure π_k.

Under Assumptions $(\boldsymbol{S})$-$(\boldsymbol{M})$-$(\boldsymbol{D_{S^c}})$ we know from [HL22b, Cor. 4.2] that there exists a computable real number $\alpha_0 \in (0,1]$ such that

$$\forall i\in S, \quad (PV^{\alpha_0})(i) \le \delta^{\alpha_0}\,V(i)^{\alpha_0} + \nu(V^{\alpha_0}). \qquad (53)$$

Obviously the comments in Remark A.1 apply here to the real number α_0 defined in (53) from the single Markov kernel P. For every $k \ge 1$ and every $i \in \mathbb{N}$, $\Delta_k(i)$ is defined in (33) and we introduce

$$\forall i\in\mathbb{N}, \quad \Delta_{k,V^{\alpha_0}}(i) = \|P(i,\cdot) - \widehat{P}_k(i,\cdot)\|_{V^{\alpha_0}}. \qquad (54)$$

For any $k \ge 1$, if π_k is a P_k-invariant probability measure on B_k, then set

$$\gamma_k := \sum_{i\in B_k}\pi_k(i)\,\Delta_k(i) \quad \text{and} \quad \gamma_{k,V^{\alpha_0}} := \sum_{i\in B_k}\pi_k(i)\,\Delta_{k,V^{\alpha_0}}(i). \qquad (55)$$

Finally let $k_0 \in \mathbb{N}$ be the smallest positive integer such that

$$V(k_0) \ge \frac{1}{\delta}, \quad S \subset B_{k_0} \quad \text{and} \quad \mathrm{Supp}(\nu) \subset B_{k_0}. \qquad (56)$$

Theorem A.2 *Assume that* $P := (P(i,j))_{(i,j)\in\mathbb{N}^2}$ *satisfies Assumptions* $(\boldsymbol{S})$*-*$(\boldsymbol{M})$*-*$(\boldsymbol{D_{S^c}})$ *and Condition (56). Moreover assume that* $\pi(1_S) > 0$ *and that for every* $k \ge k_0$ *(up to pick a larger integer* k_0*) the matrix* P_k *admits a unique* P_k-*invariant probability measure* π_k *on* B_k *such that* $\pi_k(1_S) > 0$*. Then the following assertions holds.*

(a) $\widehat{\pi}_k$ *defined in (32) is the unique* $\widehat{P}_k-$*invariant probability measure.*

(b) We have $\lim_k \gamma_k = \lim_k \gamma_{k,V^{\alpha_0}} = 0$.

(c) Defining A *and* n^* *as in (51), we have for every* $k \geq k_0$ *and for every* $n \geq n^*$*:*

$$\|\widehat{\pi}_k - \pi\|_{TV} \leq 4A\delta^{\alpha_0 n} + 4\,(n-1)\,\gamma_k. \tag{57}$$

(d) Setting $B := 2A(1+2A)$*, we have for every* $k \geq k_0$ *and for every* $n \geq n^*$*:*

$$\|\widehat{\pi}_k - \pi\|_{V^{\alpha_0}} \leq B\delta^{\alpha_0 n} + 2(n-1)\big(\gamma_{k,V^{\alpha_0}} + \ 2\pi_{\theta_0}(V^{\alpha_0})\gamma_k\big). \tag{58}$$

The P_k-invariant probability measure π_k on B_k is supposed to be computable since P_k is a finite matrix. In this case the exact value of the real numbers $\gamma_k := \pi_k(\Delta_k)$ and $\gamma_{k,V^{\alpha_0}} = \pi_k(\Delta_{k,V^{\alpha_0}})$ in (55) are computable from (59) (and from a similar formula for Δ_k). Bounds for γ_k and $\gamma_{k,V^{\alpha_0}}$ are given in Proposition A.3.

Proof of Theorem A.2. Assertion (a) and $\lim_k \gamma_k = 0$ have been proved in Lemmas 3.1-3.2. The proof of Assertion (b) is then complete using Lemma A.1 below and Corollary 2.1. Note that we have $\forall k \geq k_0$, $\widehat{\pi}_k(1_S) = \pi_k(1_S)$ since we have $S \subset B_{k_0}$ from Assumption (56). Also observe that $\gamma_k := \widehat{\pi}_k(\Delta_k)$ and $\gamma_{k,V^{\alpha_0}} = \widehat{\pi}_k(\Delta_{k,V^{\alpha_0}})$. The conclusions (57) and (58) of Theorem 3.2 then follows from Corollary 2.1 combined with Theorem A.1 using the next Lemma A.2. □

Lemma A.1 *Let* $P := (P(i,j))_{(i,j)\in\mathbb{N}^2}$ *be a Markov kernel on* $\mathbb{N}$*. For any* $k \geq 1$*, we have*

$$\forall i \in \mathbb{N}, \quad \Delta_{k,V^{\alpha_0}}(i) = 1_{B_k}(i)\bigg(P(i,{B_k}^c)\sum_{j\in B_k}\psi_{i,k}(j)V(j)^{\alpha_0} + \sum_{j\in {B_k}^c} P(i,j)V(j)^{\alpha_0}\bigg)$$
$$+ 1_{{B_k}^c}(i)\sum_{j=1}^{+\infty} P(i,j)\big(1+V(j)^{\alpha_0}\big). \tag{59}$$

Moreover we have $\forall i \in \mathbb{N}$, $\lim_k \Delta_{k,V^{\alpha_0}}(i) = 0$*, i.e. Assumption* $(\boldsymbol{\Delta_W})$ *holds with* $W = V^{\alpha_0}$*.*

Proof. Let $i \in B_k$. We have

$$\begin{aligned}\|P(i,\cdot) - \widehat{P}_k(i,\cdot)\|_{V^{\alpha_0}} &= \sum_{j\in\mathbb{N}} \big|P(i,j) - \widehat{P}_k(i,j)\big|V(j)^{\alpha_0} \\ &= P(i,{B_k}^c)\sum_{j\in B_k}\psi_{i,k}(j)V(j)^{\alpha_0} + \sum_{j\in {B_k}^c} P(i,j)V(j)^{\alpha_0}.\end{aligned}$$

For any $i \in {B_k}^c$, we easily derive from the definition of $\widehat{P}_k$ that

$$\|P(i,\cdot) - \widehat{P}_k(i,\cdot)\|_{V^{\alpha_0}} = 1 - P(i,0) + \sum_{j=1}^{+\infty} P(i,j)\,V(j)^{\alpha_0} = \sum_{j=1}^{+\infty} P(i,j)\big(1+V(j)^{\alpha_0}\big).$$

We have proved (59). Next, $\sum_{j\in B_k}\psi_{i,k}(j)V(j)^{\alpha_0} \leq \sum_{j\in B_k}\psi_{i,k}(j)V(\ell)^{\alpha_0} = V(\ell)^{\alpha_0}$ for any $\ell > k$ since V is increasing and $\psi_{i,k}(\cdot)$ is a probability distribution. It follows that

$$\forall i \in B_k, \quad \Delta_{k,V^{\alpha_0}}(i) \leq 2\sum_{j\in {B_k}^c} P(i,j)V(j)^{\alpha_0}. \tag{60}$$

Finally fix $i \in \mathbb{N}$. Then it follows from (60) applied to any $k > i$ that $\lim_k \Delta_{k,V^{\alpha_0}}(i) = 0$ since $\sum_{j\in\mathbb{N}} P(i,j)V(j)^{\alpha_0} \leq (PV)(i) < \infty$ from $V^{\alpha_0} \leq V$ and $(\boldsymbol{M})$-$(\boldsymbol{D_{S^c}})$. □

Lemma A.2 *Assume that $P := (P(i,j))_{(i,j)\in\mathbb{N}^2}$ satisfies Assumptions $(\boldsymbol{S})$-$(\boldsymbol{M})$-$(\boldsymbol{D_{S^c}})$ and Condition (56). Set $\widehat{P}_\infty := P$ and $\Theta = ([k_0,+\infty)\cap\mathbb{N})\cup\{\infty\}$. Then the family $\{\widehat{P}_k\}_{k\in\Theta}$ of Markov kernels on $\mathbb{N}$ satisfies Assumptions $(\boldsymbol{S_\Theta})$-$(\boldsymbol{M_\Theta})$-$(\boldsymbol{D_{\Theta,S^c}})$ with Θ as above defined, and*

$$\forall k\in\Theta,\quad \widehat{P}_k V^{\alpha_0} \le \delta^{\alpha_0}\, V^{\alpha_0} + \nu(V^{\alpha_0})1_S \quad \textit{with } \alpha_0 \textit{ given in (53)}. \tag{61}$$

Proof. Let $k \ge k_0$. Then $\widehat{P}_k$ satisfies Condition $(\boldsymbol{S_\Theta})$, see the begining of the proof of Lemma 3.3. Moreover we have

$$\forall i\in B_k,\quad (\widehat{P}_k V)(i) \le (PV)(i),\quad \text{thus}\quad \forall i\in S^c\cap B_k,\quad (\widehat{P}_k V)(i)\le \delta V(i)$$

from Inequality (44) applied to $W = V$ and from $(\boldsymbol{D_{S^c}})$. The first inequality provides $(\boldsymbol{M_\Theta})$ due to $(\boldsymbol{M})$. Next we deduce from the first condition of (56) and from the definition of $\widehat{P}_k$ that

$$\forall i\in B_k^c,\quad (\widehat{P}_k V)(i) = V(0) = 1 \le \delta V(i)$$

since V is increasing and since $i\in B_k^c$ implies that $i > k \ge k_0$. This proves $(\boldsymbol{D_{\Theta,S^c}})$. Finally using [HL22b, Cor. 4.2], Property (61) follows from $(\boldsymbol{S_\Theta})$-$(\boldsymbol{M_\Theta})$-$(\boldsymbol{D_{\Theta,S^c}})$: actually the real number α_0 in (61) is that given in (53) since it only depends on the data in the conditions $(\boldsymbol{S_\Theta})$-$(\boldsymbol{M_\Theta})$-$(\boldsymbol{D_{\Theta,S^c}})$ for $\widehat{P}_k$, which are the same as in $(\boldsymbol{S})$-$(\boldsymbol{M})$-$(\boldsymbol{D_{S^c}})$ for P. □

Proposition A.3 *Assume that $P := (P(i,j))_{(i,j)\in\mathbb{N}^2}$ satisfies Assumptions $(\boldsymbol{S})$-$(\boldsymbol{M})$-$(\boldsymbol{D_{S^c}})$. Let $k\ge 1$ be such that the matrix P_k admits a unique P_k-invariant probability measure π_k on B_k, and let γ_k and $\gamma_{k,V^{\alpha_0}}$ be defined in (55) with $\alpha_0\in(0,1]$ given in (53). Then*

$$\forall k\ge k_0,\quad \gamma_k \le \frac{2M}{(1-\delta)}\times\frac{1}{V(k+1)} \tag{62}$$

where M and δ are given in $(\boldsymbol{M})$-$(\boldsymbol{D_{S^c}})$. Moreover if $\alpha_0 < 1$, then

$$\forall k\ge k_0,\quad \gamma_{k,V^{\alpha_0}} \le \frac{2M}{(1-\delta)}\times\frac{1}{V(k+1)^{1-\alpha_0}}. \tag{63}$$

Finally, if $\alpha_0 = 1$ and if there exists $\eta > 1$, $\delta_\eta\in(0,1)$ and $M_\eta > 0$ such that we have $PV^\eta \le \delta_\eta V^\eta + M_\eta$, then

$$\forall k\ge k_0,\quad \gamma_{k,V} \le \frac{2M_\eta}{(1-\delta_\eta)}\times\frac{1}{V(k+1)^{\eta-1}}. \tag{64}$$

Remark A.3 *To compare our bounds with Tweedie's estimate (3) (stochastically monotone case), we assume that $P := (P(i,j))_{(i,j)\in\mathbb{N}^2}$ satisfies Assumptions $(\boldsymbol{S})$-$(\boldsymbol{M})$-$(\boldsymbol{D_{S^c}})$ with an atome S. Then we have $\alpha_0 = 1$ in (50) (see Remark A.1), and we can see that (57) combined with (62) is similar to (3). If P is not stochastically monotone, (57) and (62) remain true, and this may even provide a better bound than in [Twe98, Sect. 3], since the $V-$geometrical rate of convergence ρ in [Twe98, Th. 3.2, see Eqs (31) and (33)] may be strictly greater than δ in (57), see [HL22a, Cor 2.1] for details. To complete the discussion at the beginning of this appendix and in Remark 2.2, recall that Liu's bound [LL18, Th. 2] obtained in the atomic case, that is*

$$\|\widehat{\pi}_k - \pi\|_{TV} \le \frac{1-\delta+M}{2(1-\delta)^2}\,\gamma_k \tag{65}$$

is in any case more accurate than either Tweedie's bound (3) or (57) combined with (62). Finally mention that Liu's bound (65) has been extended in [HL22a, Th. 6.1] to any family $\{P_\theta\}_{\theta\in\Theta}$ of Markov kernels defined on a general state space and satisfying Assumptions ($\boldsymbol{S_\Theta}$)-($\boldsymbol{M_\Theta}$)-($\boldsymbol{D_{\Theta,S^c}}$), provided that δ and M in (65) are replaced with δ^{α_0} and $\nu(V^{\alpha_0})$ respectively, where α_0 is given in (50). In the atomic case where $\alpha_0 = 1$, the bound [HL22a, Th. 6.1] is exactly (65).

Proof. As in the proof of Proposition 3.4 we have from Markov's inequality

$$\begin{aligned}\gamma_k &\le 2\sum_{i\in B_k}\pi_k(i)\,\frac{(PV)(i)}{V(k+1)}\\ &\le \frac{2}{V(k+1)}\sum_{i\in B_k}\pi_k(i)\big(\delta V(i)+M\big)\qquad \text{(from } (\boldsymbol{M})\text{-}(\boldsymbol{D_{S^c}})\text{)}.\end{aligned}$$

Moreover, setting $V_k := V_{|B_k}$ the restriction of V to B_k, we have

$$\forall i\in B_k,\quad (P_kV_k)(i)\le \delta V_k(i)+M$$

from (44) applied to $W=V$ and from ($\boldsymbol{M}$)-($\boldsymbol{D_{S^c}}$). Thus

$$\pi_k(V_k)=\pi_k(P_kV_k)\le \delta\pi_k(V_k)+M,$$

from which we deduce that

$$\pi_k(V_k)\le\frac{M}{1-\delta}.$$

Then (62) easily follows from the previous inequalities.

Next, if $\alpha_0<1$, set $\beta_0 = 1-\alpha_0$ and for every $i\in\mathbb{N}$ define $\eta_i\in\mathcal{M}^+$ by: $\forall A\subset\mathbb{N},\ \eta_i=\sum_{j\in A}P(i,j)V(j)^{\alpha_0}$. Then we can write

$$\begin{aligned}\gamma_{k,V^{\alpha_0}} &\le 2\sum_{i\in B_k}\pi_k(i)\sum_{\{j:V(j)^{\beta_0}\ge V(k+1)^{\beta_0}\}}P(i,j)V(j)^{\alpha_0}\qquad \text{(from (60) and } V\nearrow)\\ &\le 2\sum_{i\in B_k}\pi_k(i)\,\frac{\eta_i(V^{\beta_0})}{V(k+1)^{\beta_0}}\qquad \text{(from Markov's inequality w.r.t. } \eta_i)\\ &= \frac{2}{V(k+1)^{\beta_0}}\sum_{i\in B_k}\pi_k(i)\,(PV)(i)\qquad \text{(from the definitions of } \eta_i \text{ and } \beta_0)\\ &\le \frac{2M}{(1-\delta)\,V(k+1)^{\beta_0}}\qquad \text{(proceeding as for } \gamma_k).\end{aligned}$$

This proves (63). Finally, if $\alpha_0=1$, then setting $\beta=\eta-1$ and applying Markov's inequality to the set $\{j : V(j)^\beta > V(k)^\beta\}$ as above we obtain that

$$\gamma_{k,V}\le\sum_{i\in B_k}\pi_k(i)\,\frac{\eta_i(V^\beta)}{V(k+1)^\beta}=\frac{2}{V(k+1)^\beta}\sum_{i\in B_k}\pi_k(i)\,(PV^\eta)(i)\le\ \frac{2M_\eta}{(1-\delta_\eta)\,V(k+1)^\beta}$$

where the last inequality is derived from the drift condition: $\forall i\in B_k,\ (P_kV^\eta{}_k)(i)\le (PV^\eta)(i)\le\delta_\eta V^\eta(i)+M_\eta$, as in the first part of the proof. □