Bernoulli **13**(3), 2007, 831–848
DOI: 10.3150/07-BEJ5162

Computable convergence rates for sub-geometric ergodic Markov chains

RANDAL DOUC[1], ERIC MOULINES[2] and PHILIPPE SOULIER[3]

[1]*CMAP, Ecole Polytechnique, 91128 Palaiseau Cedex, France. E-mail: douc@cmapx.polytechnique.fr*
[2]*Ecole Nationale Supérieure des Télécommunications, Département de Traitement du Signal et des Images, 46, rue Barrault, 75634 Paris Cedex 13, France. E-mail: moulines@tsi.enst.fr*
[3]*Department of Mathematics of University Paris X, Bâtiment G, Bureau E15, 200 avenue de la République 92000 Nanterre Cedex, France. E-mail: philippe.soulier@u-paris10.fr*

In this paper, we give quantitative bounds on the f-total variation distance from convergence of a Harris recurrent Markov chain on a given state space under drift and minorization conditions implying ergodicity at a subgeometric rate. These bounds are then specialized to the stochastically monotone case, covering the case where there is no minimal reachable element. The results are illustrated with two examples, from queueing theory and Markov Chain Monte Carlo theory.

Keywords: Markov chains; rates of convergence; stochastic monotonicity

1. Introduction

Let P be a Markov transition kernel on a state space X equipped with a countably generated σ-field $\mathcal{X}$. For a control function $f:\mathsf{X}\to[1,\infty)$, the f*-total variation* or f*-norm* of a signed measure μ on $\mathcal{X}$ is defined as $\|\mu\|_f := \sup_{|g|\le f}|\mu(g)|$. When $f\equiv 1$, the f-norm is the total variation norm, which is denoted $\|\mu\|_{\mathrm{TV}}$. We assume that P is aperiodic positive Harris recurrent with stationary distribution π. Our goal is to obtain *quantitative* bounds on convergence rates, that is,

$$r(n)\|P^n(x,\cdot)-\pi\|_f \le g(x) \qquad \text{for all } x\in\mathsf{X}, \tag{1.1}$$

where $f:\mathsf{X}\to[1,\infty)$ is a control function, $\{r(n)\}_{n\ge 0}$ is a non-decreasing sequence and $g:\mathsf{X}\to[0,\infty]$ is a function which can be computed explicitly. As emphasized in (Roberts and Rosenthal [18], Section 3.5), quantitative bounds have a substantial history in Markov chain theory. Applications are numerous, including convergence analysis of Markov Chain Monte Carlo (MCMC) methods, transient analysis of queueing systems or storage models, etc. These results have since been extended, using similar techniques, by Klokov and Veretennikov [10]. In their work, the authors consider *truly* subgeometric sequences that is, $\{r(n)\}_{n\ge0}\in\Lambda$ satisfying $\lim_{n\to\infty} r(n)n^{-\kappa}=\infty$ for any $\kappa>0$, for a more general class of functional autoregressive process.

In this paper, we study conditions under which (1.1) holds for sequences in the set Λ of subgeometric rate functions from Nummelin and Tuominen [17], defined as the family of sequences $\{r(n)\}_{n\ge0}$ such that $r(n)$ is non-decreasing and $\log r(n)/n\downarrow 0$ as $n\to\infty$. Without loss of generality, we assume that $r(0)=1$ whenever $r\in\Lambda$. These rates of convergence have seldom been

1350-7265

considered in the literature. Let us briefly summarize the results available for convergence at subgeometric rate for general state-space chains. To the best of our knowledge, the first result for subgeometric sequences was obtained by Nummelin and Tuominen [17], who derived sufficient conditions for $\|P^n(x,\cdot)-\pi\|_{\mathrm{TV}}$ to be of order $o(r^{-1}(n))$. The basic condition involved in this work is the existence of a petite set B satisfying $\sup_{x\in B}\mathbb{E}_x[\sum_{k=0}^{\tau_B-1} r(k)] < \infty$, where $\tau_B \stackrel{\text{def}}{=} \inf\{n\ge 1, X_n\in B\}$ (with the convention that $\inf\varnothing=\infty$) is the return time to B. These results were later extended by Tuominen and Tweedie [25] and Dai and Meyn [2] to f-norms for general control functions $f:\mathsf{X}\to[1,\infty)$ under the assumption of the existence of a petite set B satisfying $\sup_{x\in B}\mathbb{E}_x[\sum_{k=0}^{\tau_B-1} r(k)f(X_k)]<\infty$. These contributions do not provide computable expressions for the bounds in (1.1). A direct route to obtaining quantitative bounds for subgeometric sequences has been established by Veretennikov [26,27], based on coupling techniques (see Gulinsky and Veretennikov [8] and Rosenthal [21] for the coupling construction of Harris recurrent Markov chains). This method consists of relating the bounds (1.1) to a moment of the *coupling time* through Lindvall's inequality; see Lindvall [11,12]. Veretennikov [26,27] focus on a particular class of Markov chains, the so-called *functional autoregressive processes*, defined as $X_{n+1}=g(X_n)+W_{n+1}$, where $g:\mathbb{R}^d\to\mathbb{R}^d$ is a Borel function and $(W_n)_{n\ge 0}$ is an i.i.d. sequence, and provides expressions for the bounds in (1.1) with the total variation distance ($f\equiv 1$) and polynomial rate functions $r(n)=n^\beta$, $n\ge 1$. These results have since been extended, using similar techniques, to *truly* subgeometric sequences that is, $\{r(n)\}_{n\ge 0}\in\Lambda$ satisfying $\lim_{n\to\infty} r(n)n^{-\kappa}=\infty$ for any $\kappa>0$, in Klokov and Veretennikov [10], for a more general class of functional autoregressive process.

Fort and Moulines [7] derived quantitative bounds of the form (1.1) for possibly unbounded control functions and polynomial rate functions, also using the coupling method. The bound for the modulated moment of the coupling time is obtained from a particular drift condition introduced by Fort and Moulines [6], later extended by Jarner and Roberts [9]. These results are tailored to the polynomial rate and cannot be adapted to general subgeometric rates (see Fort [5] for comments).

The objective of this paper is to generalize the results mentioned above in two directions. We consider Markov chains over general state spaces and we study general subgeometric rates of convergence instead of polynomial rates. We establish a family of convergence bounds, extending to the subgeometric case the computable bounds obtained in the geometric case by Rosenthal [21] and later refined by Roberts and Tweedie [19] and Douc, Moulines and Rosenthal [4] (see Roberts and Rosenthal [18], Theorem 12, and the references therein). The method, based on coupling techniques, provides a short and nearly self-contained proof of the results presented in Nummelin and Tuominen [17] and Tuominen and Tweedie [25].

This paper is organized as follows. In Section 2, we present our assumptions and state our main results. In Section 2.1, we specialize our results to stochastically monotone Markov chains. Examples from queueing theory and MCMC theory are discussed in Section 3.

2. Statement of the results

The proof is based on the coupling construction (briefly recalled in Section 4). It is assumed that the chain admits a small set:

(A1) There exists a set $C \in \mathcal{X}$, a constant $\varepsilon > 0$ and a probability measure ν such that, for all $x \in C$, $P(x, \cdot) \geq \varepsilon\nu(\cdot)$.

For simplicity, only one-step minorization is considered in this paper. Adaptations to m-step minorization can be carried out, as in Rosenthal [21]. Let $\check{P}$ be a Markov transition kernel on $\mathsf{X} \times \mathsf{X}$ such that, for all $A \in \mathcal{X}$,

$$\check{P}(x, x', A \times \mathsf{X}) = P(x, A)\mathbb{1}_{(C\times C)^c}(x, x') + Q(x, A)\mathbb{1}_{C\times C}(x, x'), \tag{2.1}$$

$$\check{P}(x, x', \mathsf{X} \times A) = P(x', A)\mathbb{1}_{(C\times C)^c}(x, x') + Q(x', A)\mathbb{1}_{C\times C}(x, x'), \tag{2.2}$$

where A^c denotes the complement of the subset A and Q is the so-called *residual kernel* defined, for $x \in C$ and $A \in \mathcal{X}$, by

$$Q(x, A) = \begin{cases} (1-\varepsilon)^{-1}\big(P(x, A) - \varepsilon\nu(A)\big), & 0 < \varepsilon < 1, \\ \nu(A), & \varepsilon = 1. \end{cases} \tag{2.3}$$

One may, for example, set

$$\check{P}(x, x'; A \times A') = P(x, A)P(x', A')\mathbb{1}_{(C\times C)^c}(x, x') + Q(x, A)Q(x', A)\mathbb{1}_{C\times C}(x, x'), \tag{2.4}$$

but, as seen below, this choice is not always the most suitable. For $(x, x') \in \mathsf{X}\times\mathsf{X}$, denote by $\check{\mathbb{P}}_{x,x'}$ and $\check{\mathbb{E}}_{x,x'}$ the law and the expectation, respectively, of a Markov chain with initial distribution $\delta_x \otimes \delta_{x'}$ and transition kernel $\check{P}$.

Our second condition is a bound on the moment of the hitting time of the bivariate chain to $C \times C$ under the probability $\check{\mathbb{P}}_{x,x'}$. Let $\{r(n)\} \in \Lambda$ be a subgeometric sequence and let $R(n) \stackrel{\text{def}}{=} \sum_{k=0}^{n-1} r(k)$. Let $\sigma_{C\times C} \stackrel{\text{def}}{=} \inf\{n \geq 0, (X_n, X'_n) \in C \times C\}$ and $\tau_{C\times C} \stackrel{\text{def}}{=} \inf\{n \geq 1, (X_n, X'_n) \in C \times C\}$ (the first hitting and return time to $C \times C$) and let

$$U(x, x') \stackrel{\text{def}}{=} \check{\mathbb{E}}_{x,x'}\left[\sum_{k=0}^{\sigma_{C\times C}} r(k)\right]. \tag{2.5}$$

Let $v : \mathsf{X} \times \mathsf{X} \to [0, \infty)$ be a measurable function and set

$$V(x, x') = \check{\mathbb{E}}_{x,x'}\left[\sum_{k=0}^{\sigma_{C\times C}} v(X_k, X'_k)\right]. \tag{2.6}$$

(A2) For any $(x, x') \in \mathsf{X} \times \mathsf{X}$, $U(x, x') < \infty$ and $b_U < \infty$ where

$$b_U \stackrel{\text{def}}{=} \sup_{(x,x')\in C\times C} \check{P}U(x, x') = \sup_{(x,x')\in C\times C} \check{\mathbb{E}}_{x,x'}\left[\sum_{k=0}^{\tau_{C\times C}-1} r(k)\right]. \tag{2.7}$$

(A3) For any $(x, x') \in \mathsf{X} \times \mathsf{X}$, $V(x, x') < \infty$ and $b_V < \infty$ where

$$b_V = \sup_{(x,x')\in C\times C} \check{P}V(x, x') = \sup_{(x,x')\in C\times C} \check{\mathbb{E}}_{x,x'}\left[\sum_{k=1}^{\tau_{C\times C}} v(X_k, X'_k)\right]. \tag{2.8}$$

We will establish that R is the rate of convergence associated with the total variation norm. On the other hand, we will show that the difference $P(x,\cdot)-P(x',\cdot)$ remains bounded in f-norm for any function f satisfying $f(x)+f(x')\le V(x,x')$ for any $(x,x')\in \mathsf{X}\times\mathsf{X}$. Using an interpolation technique, we will derive a rate of convergence $1\le s\le r$ associated with some g-norm, $0\le g\le f$. To construct such an interpolation, we consider a pair of positive functions (α,β) satisfying, for some $0\le\rho\le 1$,

$$\alpha(u)\beta(v)\le \rho u+(1-\rho)v \qquad \text{for all } (u,v)\in\mathbb{R}^+\times\mathbb{R}^+. \tag{2.9}$$

Theorem 2.1. *Assume* (A1), (A2) *and* (A3). *Define*

$$M_U \stackrel{\text{def}}{=} \sup_{k\in\mathbb{N}}\left\{\left(b_U r(k)\frac{1-\varepsilon}{\varepsilon}-R(k+1)\right)_+\right\} \quad \textit{and} \quad M_V\stackrel{\text{def}}{=} b_V\frac{1-\varepsilon}{\varepsilon}, \tag{2.10}$$

where $(x)_+\stackrel{\text{def}}{=}\max(x,0)$. *Then, for any* $(x,x')\in\mathsf{X}\times\mathsf{X}$,

$$\|P^n(x,\cdot)-P^n(x',\cdot)\|_{\mathrm{TV}}\le \frac{U(x,x')+M_U}{R(n)+M_U}, \tag{2.11}$$

$$\|P^n(x,\cdot)-P^n(x',\cdot)\|_f\le V(x,x')+M_V, \tag{2.12}$$

for any non-negative function f satisfying, for any $(x,x')\in\mathsf{X}\times\mathsf{X}$, $f(x)+f(x')\le V(x,x')+M_V$. *Let* (α,β) *be two positive functions satisfying* (2.9) *for some* $0\le\rho\le1$. *Then, for any* $(x,x')\in\mathsf{X}\times\mathsf{X}$ *and* $n\ge1$,

$$\|P^n(x,\cdot)-P^n(x',\cdot)\|_g\le\frac{\rho(U(x,x')+M_U)+(1-\rho)(V(x,x')+M_V)}{\alpha\circ\{R(n)+M_U\}}, \tag{2.13}$$

for any non-negative function g satisfying, for any $(x,x')\in\mathsf{X}\times\mathsf{X}$, $g(x)+g(x')\le\beta\circ\{V(x,x')+M_V\}$.

The proof is postponed to Section 4.

Remark 1. Because the sequence $\{r(k)\}$ is subgeometric, $\lim_{k\to\infty} r(k)/R(k+1)=0$. Therefore, the sequence $\{b_U r(k)(1-\varepsilon)/\varepsilon-R(k)\}$ has only finitely many non-negative terms, which implies that $M_U<\infty$.

Remark 2. When assumption (A2) holds, (A3) is automatically satisfied for some function v. Note that

$$\check{\mathbb{E}}_{x,x'}\left[\sum_{k=0}^{\sigma_{C\times C}} r(k)\right]=\check{\mathbb{E}}_{x,x'}\left[\sum_{k=0}^{\sigma_{C\times C}} r(\sigma_{C\times C}-k)\right].$$

On the other hand, for all $(x,x')\in\mathsf{X}\times\mathsf{X}$,

$$\begin{aligned}&\check{\mathbb{E}}_{x,x'}\big[r(\sigma_{C\times C}-k)\mathbb{1}_{\{\sigma_{C\times C}\ge k\}}\big]\\ &\quad=\check{\mathbb{E}}_{x,x'}\big[\check{\mathbb{E}}_{X_k,X'_k}[r(\sigma_{C\times C})]\mathbb{1}_{\{\sigma_{C\times C}\ge k\}}\big]=\check{\mathbb{E}}_{x,x'}\big[v_r(X_k,X'_k)\mathbb{1}_{\{\sigma_{C\times C}\ge k\}}\big],\end{aligned}$$

where $v_r(x, x') \stackrel{\text{def}}{=} \check{\mathbb{E}}_{x,x'}[r(\sigma_{C\times C})]$. This relation implies that

$$\check{\mathbb{E}}_{x,x'}\left[\sum_{k=0}^{\sigma_{C\times C}} r(k)\right] = \check{\mathbb{E}}_{x,x'}\left[\sum_{k=0}^{\sigma_{C\times C}} v_r(X_k, X'_k)\right] \qquad \text{for all } (x, x') \in \mathsf{X} \times \mathsf{X}.$$

To check assumptions (A2) and (A3), it is often useful to use drift conditions. Drift conditions implying convergence at polynomial rates were recently proposed in Jarner and Roberts [9]. These conditions have since been extended to general subgeometric rates by Douc *et al.* [3]. Define by $\mathcal{C}$ the set of functions

$$\mathcal{C} \stackrel{\text{def}}{=} \Big\{\phi : [1, \infty) \to \mathbb{R}^+,\ \phi \text{ is concave, differentiable and}$$
$$\phi(1) > 0,\ \lim_{v\to\infty} \phi(v) = \infty,\ \lim_{v\to\infty} \phi'(v) = 0\Big\}. \tag{2.14}$$

For $\phi \in \mathcal{C}$, define $H_\phi(v) \stackrel{\text{def}}{=} \int_1^v (1/\phi(x))\,\mathrm{d}x$. Since ϕ is non-decreasing, H_ϕ is a non-decreasing concave differentiable function on $[1, \infty)$ and $\lim_{v\to\infty} H_\phi(v) = \infty$. The inverse, $H_\phi^{-1} : [0, \infty) \to [1, \infty)$, is also an increasing and differentiable function, with derivative $(H_\phi^{-1})' = \phi \circ H_\phi^{-1}$. Note that $(\log\{\phi \circ H_\phi^{-1}\})' = \phi' \circ H_\phi^{-1}$. Since H_ϕ is increasing and ϕ' is decreasing, $\phi \circ H_\phi^{-1}$ is log-concave, which implies that the sequence

$$r_\phi(n) \stackrel{\text{def}}{=} \phi \circ H_\phi^{-1}(n)/\phi \circ H_\phi^{-1}(0) \tag{2.15}$$

belongs to the set of subgeometric sequences Λ. Consider the following assumption:

(A4) There exist functions $W : \mathsf{X} \times \mathsf{X} \to [1, \infty)$ and $\phi \in \mathcal{C}$ such that $\check{P}W(x, x') \le W(x, x') - \phi \circ W(x, x')$ for $(x, x') \notin C \times C$ and $\sup_{(x,x')\in C\times C} \check{P}W(x, x') < \infty$.

It is shown in Douc *et al.* [3] that under assumption (A4), assumptions (A2) and (A3) are satisfied with the rate sequence r_ϕ and the control function $v = \phi \circ W$. More precisely, we have the following.

Proposition 2.2. *Assume* (A4). *Then* (A2) *and* (A3) *hold with* $v = \phi \circ W$, $r = r_\phi$ *and*

$$U(x, x') \le 1 + \frac{r_\phi(1)}{\phi(1)}\{W(x, x') - 1\}\mathbb{1}_{(C\times C)^c}(x, x'), \tag{2.16}$$

$$V(x, x') \le \sup_{C\times C} \phi \circ W + W(x, x')\mathbb{1}_{(C\times C)^c}(x, x'), \tag{2.17}$$

$$b_U \le 1 + \frac{r_\phi(1)}{\phi(1)}\left\{\sup_{C\times C} \check{P}W - 1\right\}, \tag{2.18}$$

$$b_V \le \sup_{C\times C} \phi \circ W + \sup_{C\times C} \check{P}W. \tag{2.19}$$

The proof is in Section 5. Proposition 2.2 is only partially satisfactory because assumption (A4) is formulated on the bivariate kernel $\check{P}$. It is, in general, easier to directly establish the drift condition on the kernel P and to deduce from this condition a drift condition for an appropriately defined kernel $\check{P}$ (see Roberts and Rosenthal [18], Proposition 11). Consider the following assumption:

(A5) There exist a function $W_0 : \mathsf{X} \to [1, \infty)$, a function $\phi_0 \in \mathcal{C}$ and a constant b_0 such that $PW_0 \leq W_0 - \phi_0 \circ W_0 + b_0 \mathbb{1}_C$.

Theorem 2.3. *Suppose that* (A1) *and* (A5) *are satisfied. Let* $d_0 \stackrel{\text{def}}{=} \inf_{x \notin C} W_0(x)$. *Then, if* $\phi_0(d_0) > b_0$, *the kernel* $\check{P}$ *defined in* (2.4) *satisfies the bivariate drift condition* (A4) *with*

$$W(x, x') = W_0(x) + W_0(x') - 1, \tag{2.20}$$

$$\phi = \lambda \phi_0, \qquad \textit{for any } \lambda,\ 0 < \lambda < 1 - b_0/\phi_0(d_0), \tag{2.21}$$

$$\sup_{C \times C} \check{P} W \leq 2(1 - \varepsilon)^{-1} \left\{ \sup_C P W_0 - \varepsilon \nu(W_0) \right\} - 1, \tag{2.22}$$

where the kernel Q *is defined in* (2.3).

The proof is postponed to the Appendix.

2.1. Stochastically ordered chains

Let X be a totally ordered set and denote by $\preceq$ the order relation. For $a \in \mathsf{X}$, denote by $(-\infty, a]$ the set $\{x \in \mathsf{X} : x \preceq a\}$ and by $[a, +\infty)$ the set $\{x \in \mathsf{X} : a \preceq x\}$. A transition kernel P on X is called *stochastically monotone* if for all $a \in \mathsf{X}$, $P(\cdot, (-\infty, a])$ is non-increasing. Stochastic monotonicity has been seen to be crucial in the analysis of queuing networks, Markov Chain Monte Carlo methods, storage models, etc. Stochastically ordered Markov chains have been considered in Lund and Tweedie [14], Lund *et al.* [13], Scott and Tweedie [23] and Roberts and Tweedie [20]. In the first two of these papers, it is assumed that there exists an atom at the bottom of the state space. Lund *et al.* [13] cover only geometric convergence; subgeometric rates of convergence are considered in Scott and Tweedie [23]. Roberts and Tweedie [20] covers the case where the bottom of the space is a small set but restricts its attention to conditions implying a geometric rate of convergence.

For a general stochastically monotone Markov kernel P, it is always possible to define the bivariate kernel $\check{P}$ (see (2.1)) so that the two components $\{X_n\}_{n \geq 0}$ and $\{X'_n\}_{n \geq 0}$ are *pathwise ordered*, that is, their initial order is preserved at all times.

The construction goes as follows. For $x \in \mathsf{X}$, $u \in [0, 1]$ and K a transition kernel on X, denote by $G_K^-(x, u)$ the quantile function associated with the probability measure $K(x, \cdot)$:

$$G_K^-(x, u) = \inf\{y \in \mathsf{X}, K(x, (-\infty, y]) \geq u\}. \tag{2.23}$$

Assume that (A1) holds. For $(x, x') \in \mathsf{X} \times \mathsf{X}$ and $A \in \mathcal{X} \otimes \mathcal{X}$, define the transition kernel $\check{P}$ by

$$\mathbb{1}_{(C\times C)^c(x,x')} \check{P}(x, x'; A) = \int_0^1 \mathbb{1}_A(G_P^-(x, u), G_P^-(x', u))\,\mathrm{d}u + \mathbb{1}_{C\times C}(x, x') \int_0^1 \mathbb{1}_A(G_Q^-(x, u), G_Q^-(x', u))\,\mathrm{d}u,$$

where Q is the residual kernel defined in 2.3. It is easily seen that, by construction, the set $\{(x, x') \in \mathsf{X} \times \mathsf{X} : x \preceq x'\}$ is absorbing for the kernel $\check{P}$.

In the sequel, we assume that (A1) holds for some $C \stackrel{\text{def}}{=} (-\infty, x_0]$ (i.e., that there is a small set at the bottom of the space). Let $v_0 : \mathsf{X} \to [1, \infty)$ be a measurable function and define

$$U_0(x) \stackrel{\text{def}}{=} \mathbb{E}_x\left[\sum_{k=0}^{\sigma_C} r(k)\right] \quad \text{and} \quad V_0(x) = \mathbb{E}_x\left[\sum_{k=0}^{\sigma_C} v_0(X_k)\right]. \tag{2.24}$$

Consider the following assumptions:

(B1) For any $x \in \mathsf{X}$, $U_0(x) < \infty$ and $\sup_C QU_0 = b_{U_0} < \infty$.
(B2) For any $x \in \mathsf{X}$, $V_0(x) < \infty$ and $\sup_C QV_0 = b_{V_0} < \infty$.

Theorem 2.4. *Assume that* (A1), (B2) *and* (B3) *hold for some set* $C \stackrel{\text{def}}{=} (-\infty, x_0]$. *Then* (A2) *and* (A3) *hold with* $U(x, x') = U_0(x \vee x')$, $V(x, x') = V_0(x \vee x')$, $v(x, x') = v_0(x \vee x')$, $b_U = b_{U_0}$ *and* $b_V = b_{V_0}$.

The proof is omitted for brevity. As mentioned above, drift conditions often provide an easy means to establish (B2) and (B3). Consider the following assumption:

(B4) There exists a non-negative function $W_0 : \mathsf{X} \to [1, \infty)$ and a function $\phi \in \mathcal{C}$ such that for $x \notin C$, $PW_0 \le W_0 - \phi \circ W_0$ and $\sup_C PW_0 < \infty$.

Using, as above, Douc *et al.* [3], it may be shown that this assumption implies (B2) and (B3) and allows the constants to be computed explicitly.

Theorem 2.5. *Assume* (A1) *and* (B4). *Then* (B2) *and* (B3) *hold with* $v_0 = \phi \circ W_0$, $r = r_\phi$ *and*

$$U_0(x) \le 1 + \frac{r_\phi(1)}{\phi(1)} \{W_0(x) - 1\} \mathbb{1}_{C^c}(x), \tag{2.25}$$

$$V_0(x) \le \sup_C \phi \circ W_0 + W_0(x) \mathbb{1}_{C^c}(x), \tag{2.26}$$

$$b_{U_0} \le 1 + \frac{r_\phi(1)}{\phi(1)} \left((1 - \varepsilon)^{-1} \left\{ \sup_C PW_0 - \varepsilon \nu(W_0) \right\} - 1 \right), \tag{2.27}$$

$$b_{V_0} \le \sup_C \phi \circ W_0 + (1 - \varepsilon)^{-1} \left\{ \sup_C PW_0 - \varepsilon \nu(W_0) \right\}. \tag{2.28}$$

The proof is analogous to that of Proposition 2.2 and is hence omitted.

3. Applications

3.1. M/G/1 queue

In an M/G/1 queue, customers arrive into a service operation according to a Poisson process with parameter λ. Customers bring jobs requiring service times which are independent of each other and of the inter-arrival time with common distribution B concentrated on $(0, \infty)$ (we assume that the service time distribution has no probability mass at 0). Consider the random variable X_n which counts customers immediately after each service time ends. $\{X_n\}_{n\geq 0}$ is a Markov chain on integers with transition matrix

$$P = \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & \cdots \\ a_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & a_0 & a_1 & a_2 & \cdots \\ 0 & 0 & a_0 & a_1 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}, \tag{3.1}$$

where for each $j \geq 0$, $a_j \stackrel{\text{def}}{=} \int_0^\infty \{\mathrm{e}^{-\lambda t}(\lambda t)^j/j!\}\,\mathrm{d}B(t)$ (see Meyn and Tweedie [15], Proposition 3.3.2). It is known that P is irreducible, aperiodic and positive recurrent if $\rho \stackrel{\text{def}}{=} \lambda m_1 = \sum_{j=1}^\infty j a_j < 1$, where for $u > 0$, $m_u \stackrel{\text{def}}{=} \int t^u\,\mathrm{d}B(t)$. Applying the results derived above, we will compute explicit bounds (depending on λ, x and the moments of the service time distribution) for the convergence bound $\|P^n(x,\cdot) - \pi\|_f$ for some appropriately defined function f.

Because the chain is irreducible and positive recurrent, $\tau_0 < \infty$ $\mathbb{P}_x$-a.s. for $x \in \mathbb{N}$. By construction, for all $x = 1, 2, \ldots$, $\tau_{x-1} \leq \tau_0$, $\mathbb{P}_x$-a.s., which implies that $\mathbb{E}_x[\tau_0] = \mathbb{E}_x[\tau_{x-1}] + \mathbb{E}_{x-1}[\tau_0]$ and, for any $s \in \mathbb{C}$ such that $|s| \leq 1$, $\mathbb{E}_x[s^{\tau_0}] = \mathbb{E}_x[s^{\tau_{x-1}}]\mathbb{E}_{x-1}[s^{\tau_0}]$, where τ_{x-1} is the first return time of the state $x-1$. For all $x = 1, 2, \ldots$, we have $\mathbb{P}_x\{\tau_{x-1} \in \cdot\} = \mathbb{P}_1\{\tau_0 \in \cdot\}$, which shows that $\mathbb{E}_x[\tau_0] = x\mathbb{E}_1[\tau_0]$ and $\mathbb{E}_x[s^{\tau_0}] = e^x(s)$, where $e(s) \stackrel{\text{def}}{=} \mathbb{E}_1[s^{\tau_0}]$. This relation implies that

$$e(s) = s a_0 + \sum_{y=1}^\infty a_y e^y(s) = s\int_0^\infty \mathrm{e}^{\lambda(e(s)-1)t}\,\mathrm{d}B(t).$$

By differentiating the previous relation with respect to$s$ and taking the limit as $s \to 1$, we have $\mathbb{E}_1[\tau_0] = (1-\rho)^{-1}$. Since $\{0, 1\}$ is an atom, we may use Theorem 2.4 with $C = \{0, 1\}$, $r \equiv 1$ and $v_0 \equiv 1$. In this case,

$$U_0(x) = V_0(x) = 1 + \mathbb{E}_x[\sigma_C] = 1 + \mathbb{E}_{x-1}[\tau_0]\mathbb{1}_{\{x\geq 2\}} = 1 + (1-\rho)^{-1}(x-1)\mathbb{1}_{\{x\geq 2\}}.$$

Theorem 2.1 shows that for any $(x, x') \in \mathbb{N}\times\mathbb{N}$ and any functions α and β satisfying (2.9),

$$\alpha(n)\|P^n(x,\cdot) - P^n(x',\cdot)\|_\beta \leq 1 + (1-\rho)^{-1}(x \vee x' - 1)\mathbb{1}_{\{x\vee x'\geq 2\}}.$$

Convergence bounds $\alpha(n)\|P^n(x,\cdot) - \pi\|_\beta$ can be obtained by integrating the previous relation in x' with respect to the stationary distribution π (which can be computed using the Pollaczek–Khinchine formula).

It is possible to choose the set C in a different way, leading to different bounds. One may set, for example, $C = \{0, \dots, x_0\}$, for some $x_0 \geq 2$. For simplicity, assume that the sequence $\{a_j\}_{j\geq 0}$ is non-decreasing. In this case, for all $x \in C$ and $y \in \mathbb{N}$, $P(x, y) = a_{y-x+1}\mathbb{1}_{\{y\geq x-1\}} \geq a_y\mathbb{1}_{\{y\geq x_0-1\}}$ and the set C satisfies (A1) with $\varepsilon \stackrel{\text{def}}{=} \sum_{y=x_0-1}^{\infty} a_y$ and $\nu(y) = \varepsilon^{-1}a_y\mathbb{1}_{\{y\geq x_0-1\}}$. Again taking $r(k) \equiv 1$ and $v_0(x) \equiv 1$, we have

$$\begin{aligned}
U_0(x) = V_0(x) &= 1 + \mathbb{E}_x[\tau_C]\mathbb{1}_{C^c}(x) = 1 + \mathbb{E}_x[\tau_{x_0}]\mathbb{1}_{C^c}(x) \\
&= 1 + \mathbb{E}_{x-x_0}[\tau_0]\mathbb{1}_{C^c}(x) = 1 + (1-\rho)^{-1}(x - x_0)\mathbb{1}_{C^c}(x).
\end{aligned}$$

To apply the results of Theorem 2.4, we finally compute a bound for $b_{U_0} = \sup_C QU_0 = (1-\varepsilon)^{-1}[\sup_C PU_0 - \varepsilon\nu(U_0)]$, which can be obtained by combining a bound for $\sup_C PU_0$ and the expression for $\nu(U_0)$. An expression for $\nu(U_0)$ is computed by a direct application of the definitions. The bound for $\sup_C PU_0$ is obtained by noting that for all $y > x_0$ and $x \in C$, $P(x, y) \leq P(x_0, y) = a_{y-x_0+1}$, which implies that

$$\begin{aligned}
PU_0(x) &= \mathbb{E}_x[\tau_C] = 1 + \mathbb{E}_x\big[\mathbb{E}_{X_1}[\tau_C]\mathbb{1}_{\{\tau_C>1\}}\big] = 1 + \mathbb{E}_x\big[\mathbb{E}_{X_1}[\tau_{x_0}]\mathbb{1}_{\{X_1\notin C\}}\big] \\
&= 1 + (1-\rho)^{-1}\sum_{y=x_0+1}^{\infty}(y - x_0)P(x, y) \leq 1 + (1-\rho)^{-1}\sum_{y=x_0+1}^{\infty}(y - x_0)a_{y-x_0+1}.
\end{aligned}$$

We provide some numerical illustrations of the bounds described above. We use the distribution of service times suggested in Roughan *et al.* [22] and given by

$$b(x) = \begin{cases} \alpha B^{-1}\mathrm{e}^{-(\alpha)/Bx}, & x \leq B, \\ \alpha B^{\alpha}\mathrm{e}^{-\alpha}x^{-\alpha+1}, & x > B, \end{cases} \tag{3.2}$$

where B marks where the tail begins. The mean of the service distribution is $m_1 = B\{1 + \mathrm{e}^{-\alpha}/(\alpha-1)\}/\alpha$ and its Laplace transform, $G(s) = \int_0^{\infty}\mathrm{e}^{-st}\,\mathrm{d}B(t)$, $s \in \mathbb{C}$, $\mathrm{Re}(s) \geq 0$, is given by

$$G(s) = \alpha\frac{1 - \mathrm{e}^{-(sB+\alpha)}}{sB + \alpha} + \alpha B^{\alpha}\mathrm{R}\mathrm{e}^{-\alpha}s^{\alpha}\Gamma(-\alpha, sB),$$

where $\Gamma(x, z)$ is the incomplete Γ function. The probability generating function $P_\pi(z)$ of the stationary distribution is given by the Pollaczek–Khinchine formula

$$P(z) = \frac{(1-\rho)(z-1)G(\lambda(1-z))}{z - G(\lambda(1-z))}.$$

In Figure 1, we display the convergence bound $\|P^n(x, \cdot) - \pi\|_{\mathrm{TV}}$ as a function of the iteration index n, for $x = 10$, $\alpha = 2.5$, different choices of the small set upper limit $x_0 = 1, 3, 6$ and two different values of the traffic, $\rho = 0.5$ (light traffic) and $\rho = 0.9$ (heavy traffic). Perhaps surprisingly, the bound computed using the atom $C = \{0, 1\}$ is not uniformly better in the iteration index n. There is a trade-off between the number of visits to the small set where coupling might occur and the probability that coupling is successful. In the heavy traffic case ($\rho = 0.9$), the queue is not very often empty, so the atom is not frequently visited, explaining why deriving

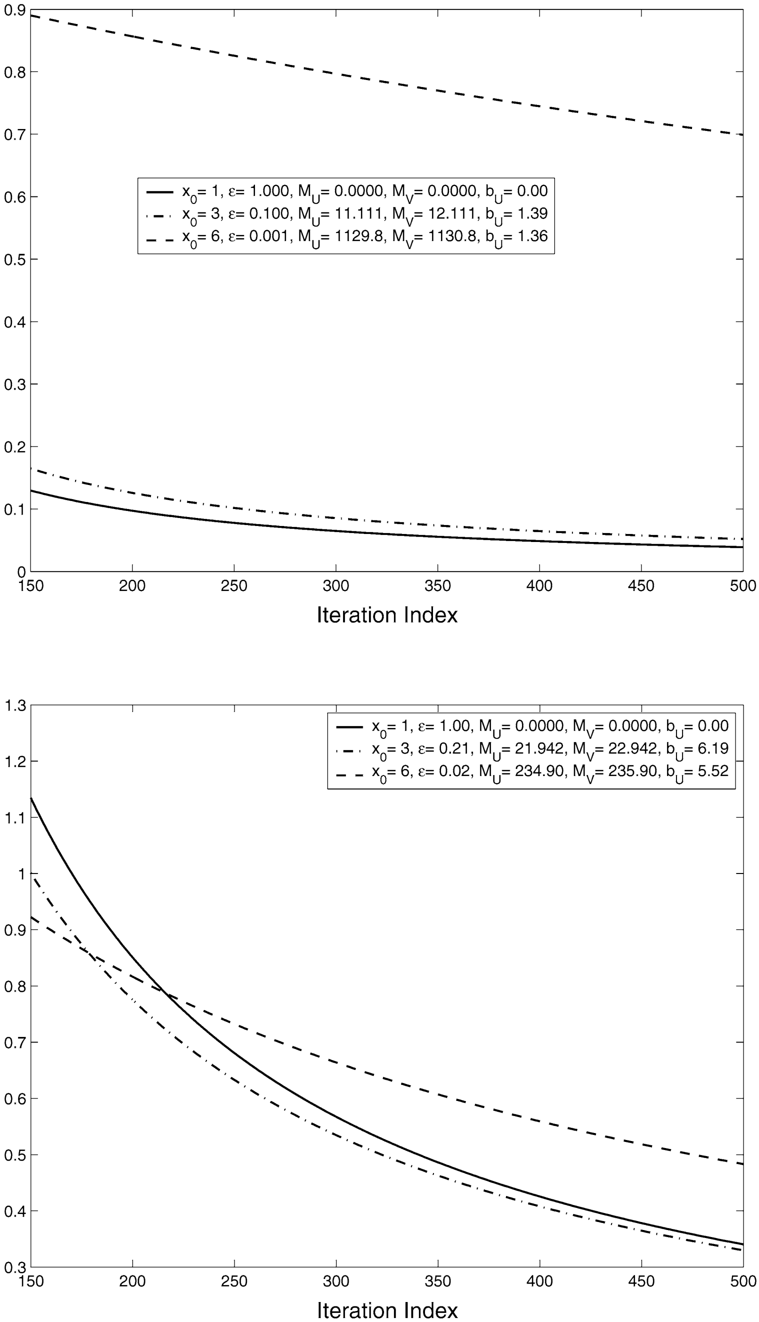

Figure 1. Convergence bound for the total variation distance. Bottom panel: light-traffic case: $\rho = 0.5$, $\alpha = 2.5$; Top panel: heavy-traffic case: $\rho = 0.9$, $\alpha = 2.5$.

the convergence bound from a larger coupling set improves the bound (this effect is even more noticeable for a critically loaded system).

3.2. The independence sampler

This second example is borrowed from Jarner and Roberts [9]. It is an example of a Markov chain which is stochastically monotone with respect to a non-standard ordering of the state and which does not have an atom at the bottom of the state space.

The purpose of the Metropolis–Hastings independence sampler is to sample from a probability density π (with respect to some σ-finite measure μ on X), which is known only up to a scale factor. At each iteration, a move is proposed according to a distribution with density q with respect to μ. The move is accepted with probability $a(x,y) \stackrel{\text{def}}{=} \frac{q(x)}{\pi(x)}\frac{\pi(y)}{q(y)} \wedge 1$. The transition kernel of the algorithm is thus given by

$$P(x,A) = \int_A a(x,y)q(y)\mu(\mathrm{d}y) + \mathbb{1}_A(x)\int_{\mathsf{X}} \bigl(1 - a(x,y)\bigr)q(y)\mu(\mathrm{d}y), \qquad x \in \mathsf{X}, A \in \mathcal{X}.$$

It is well known that the independence sampler is stochastically monotone with respect to the ordering $x' \preceq x \Leftrightarrow \frac{q(x)}{\pi(x)} \leq \frac{q(x')}{\pi(x')}$. Without loss of generality, it is assumed that $\pi(x) > 0$ for all $x \in \mathsf{X}$ and that $q > 0$ π-a.s. For all $\eta > 0$, define the set

$$C_\eta \stackrel{\text{def}}{=} \left\{x \in \mathsf{X}: \frac{q(x)}{\pi(x)} \geq \eta\right\}. \tag{3.3}$$

For any $\eta > 0$, we assume that $0 < \pi(C_\eta) < 1$ and denote by $\nu_\eta(\cdot)$ the probability measure $\nu_\eta(\cdot) = \pi(\cdot \cap C_\eta)/\pi(C_\eta)$. For any $x \in C_\eta$,

$$\begin{aligned} P(x,A) &\geq \int_A \left(\frac{q(x)}{\pi(x)} \wedge \frac{q(y)}{\pi(y)}\right)\pi(y)\mu(\mathrm{d}y) \\ &\geq \int_{A\cap C_\eta} \left(\frac{q(x)}{\pi(x)} \wedge \frac{q(y)}{\pi(y)}\right)\pi(y)\mu(\mathrm{d}y) \geq \eta\pi(A \cap C_\eta) = \eta\pi(C_\eta)\nu_\eta(A), \end{aligned}$$

showing that the set C_η satisfies (A1) with $\nu = \nu_\eta$ and $\varepsilon = \eta\pi(C_\eta)$.

Proposition 3.1. *Assume that there exists a decreasing differentiable function $K : (0,\infty) \to (1,\infty)$, whose inverse is denoted by K^{-1}, satisfying the following conditions*:

(1) *the function $\phi(v) = vK^{-1}(v)$ is differentiable, increasing and concave on $[1,\infty)$, $\lim_{v\to\infty}\phi(v) = \infty$ and $\lim_{v\to\infty}\phi'(v) = 0$;*

(2) *$\int_0^{+\infty} uK(u)\,\mathrm{d}\psi(u) < \infty$, where for $\eta > 0$, $\psi(\eta) \stackrel{\text{def}}{=} 1 - \pi(C_\eta)$.*

Then, for any $\eta^\star$ satisfying $\{1 - \psi(\eta^\star)\}\phi(1) > \int_0^\infty (u \wedge \eta^\star)K(u)\,\mathrm{d}\psi(u)$, (B4) *is satisfied with $W_0 = K \circ (q/\pi)$, $C = C_{\eta^\star}$ and $\phi_0(v) = \{1 - \psi(\eta^\star)\}\phi(v) - \int_0^\infty (u \wedge \eta^\star)K(u)\,\mathrm{d}\psi(u)$. In addition, $\sup_{x\in C_{\eta^\star}} PW_0 \leq \int_0^{+\infty} uK(u)\,\mathrm{d}\psi(u) + K(\eta^\star)$.*

To illustrate our results, we evaluate the convergence bounds in the case where the target density π is the uniform distribution on $[0, 1]$ and the proposal density is $q(x) = (r+1)x^r \mathbb{1}_{[0,1]}(x)$. Proposition 3.1 provides a means to derive a drift condition of the form $PW_0 \le W_0 - \phi \circ W_0$ outside some small set C for functions $\phi \in \mathcal{C}$ of the form $\phi(v) = cv^{1-1/\alpha} + d$ for any $\alpha \in [1, 1 + 1/r)$. In this case, the function ψ is given by $\psi(\eta) = (\eta/(r+1))^{1/r}$ for $\eta \in [0, r+1]$ and $\psi(\eta) = 1$ otherwise. We set, for $u \in [0, r+1]$, $K(u) = (u/(r+1))^{-\alpha}$. The integral $\int uK(u)\,\mathrm{d}\psi(u) = \frac{(r+1)^{-\alpha}}{r(-\alpha+1/r+1)}$ is finite provided that $\alpha < 1 + 1/r$. The function $\phi(u) = uK^{-1}(u) = u^{1-1/\alpha}(r+1)$ belongs to $\mathcal{C}$ provided that $\alpha > 1$.

Using these results, it is now straightforward to evaluate the constants in Theorem 2.1; this approach can be employed to calculate a bound on exactly how many iterations are necessary to get within a prespecified total variation distance of the target distribution. In Figure 2, we have displayed the total variation bounds to convergence for the instrumental densities $q(x) = 3x^2$ ($r = 2$) and $q(x) = (3/2)\sqrt{x}$. We have taken $\alpha = 1.1$ and $\eta^\star = 0.25$ for $r = 2$ and taken $\alpha = 1.5$ and $\eta^\star = 0.5$ for $r = 1/2$. When $r = 2$ and $\alpha = 1.1$, the convergence to stationarity is quite slow, which is not surprising since the instrumental density does not match well the target density at $x = 0$: according to our computable bounds, 500 iterations are required to bring the total variation to the stationary distribution below 0.1. When $r = 1/2$, the degeneracy of the instrumental density at zero is milder and the convergence rate is significantly faster. Less than 50 iterations are required to reach the same bound.

4. Proof of Theorem 2.1

The proof is based on the pathwise coupling construction. For $(x, x') \in \mathsf{X} \times \mathsf{X}$ and $A \in \mathcal{X} \otimes \mathcal{X}$, define $\bar{P}$, the coupling kernel, as follows:

$$\begin{aligned}
\bar{P}(x, x', 0; A \times \{0\}) &= \bigl(1 - \varepsilon \mathbb{1}_{C\times C}(x, x')\bigr)\check{P}(x, x', A),\\
\bar{P}(x, x', 0; A \times \{1\}) &= \varepsilon \mathbb{1}_{C\times C}(x, x')\nu\bigl(A \cap \{(x, x') \in \mathsf{X}\times\mathsf{X}, x = x'\}\bigr),\\
\bar{P}(x, x', 1; A \times \{0\}) &= 0,\\
\bar{P}(x, x', 1; A \times \{1\}) &= \int P(x, \mathrm{d}y)\mathbb{1}_A(y, y).
\end{aligned}$$

For any $(x, x') \in \mathsf{X} \times \mathsf{X}$, denote by $\bar{\mathbb{P}}_{x,x'}$ and $\bar{\mathbb{E}}_{x,x'}$ the probability measure and the expectation, respectively, associated to the Markov chain $\{(X_n, X'_n, d_n)\}_{n\ge 0}$ with transition kernel $\bar{P}$ starting from $(X_0, X'_0, 0) = (x, x', 0)$. By construction, for any n, $(x, x') \in \mathsf{X}\times\mathsf{X}$ and $(A, A') \in \mathcal{X} \times \mathcal{X}$, we have

$$\bar{\mathbb{P}}_{x,x',0}(Z_n \in A \times \mathsf{X} \times \{0, 1\}) = \bar{\mathbb{P}}_{x,x',0}(X_n \in A) = P^n(x, A)$$

and

$$\bar{\mathbb{P}}_{x,x',0}(Z_n \in \mathsf{X} \times A' \times \{0, 1\}) = \bar{\mathbb{P}}_{x,x',0}(X'_n \in A') = P^n(x', A').$$

By Douc *et al.* [4], Lemma 1 we may relate the expectations of functionals under the two probability measures $\bar{\mathbb{P}}_{x,x',0}$ and $\check{\mathbb{P}}_{x,x'}$, where $\check{\mathbb{P}}_{x,x'}$ is defined in (2.1): for any non-negative adapted

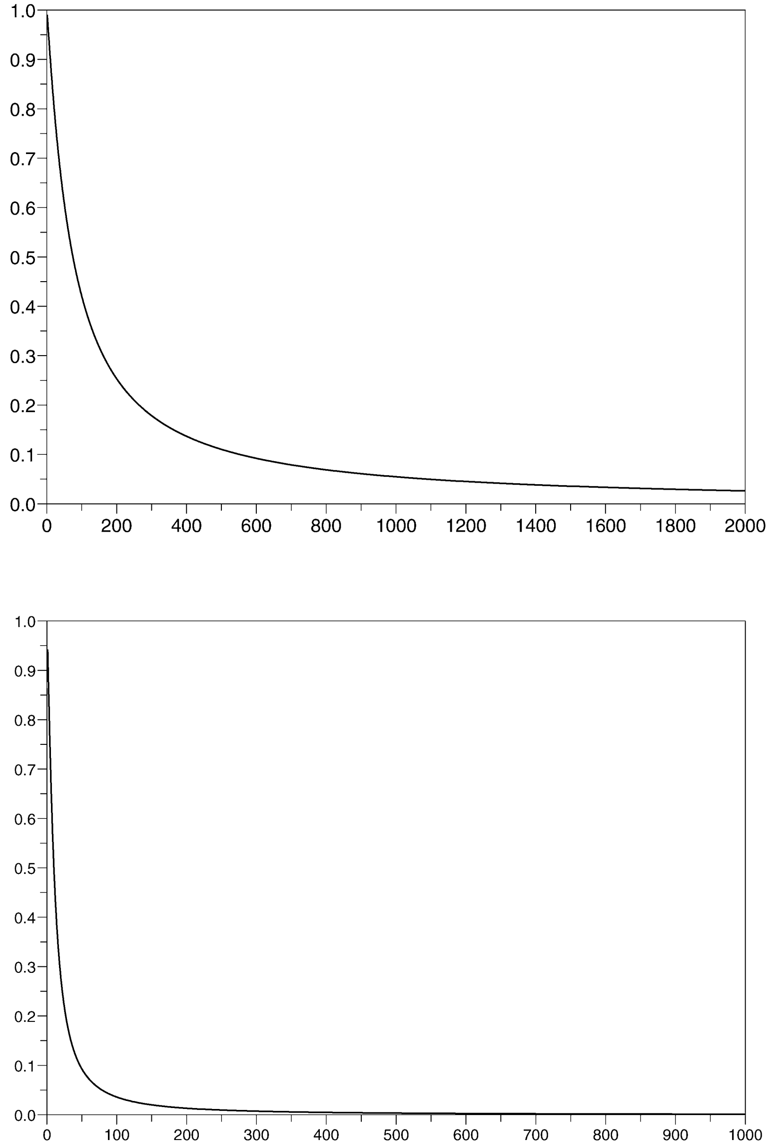

Figure 2. Convergence bound for the total variation distance for the independence sampler. Bottom panel: $q(x) = 3x^2$. Top panel: $q(x) = 1.5\sqrt{x}$.

process $(\chi_k)_{k\geq 0}$ and $(x, x') \in \mathsf{X} \times \mathsf{X}$,

$$\bar{\mathbb{E}}_{x,x',0}\big[\chi_n \mathbb{1}_{\{T>n\}}\big] = \check{\mathbb{E}}_{x,x'}[\chi_n (1-\varepsilon)^{N_{n-1}}], \tag{4.1}$$

where $T \stackrel{\text{def}}{=} \inf\{n \geq 1 : d_n = 1\}$ and $N_n \stackrel{\text{def}}{=} \sum_{i=0}^{n} \mathbb{1}_{C\times C}(X_i, X_i')$ is the number of visits to the set $C \times C$ before time n. Let $f : \mathsf{X} \to [0, \infty)$ and let $g : \mathsf{X} \to \mathbb{R}$ be any Borel function such that $\sup_{x\in\mathsf{X}} |g(x)|/f(x) < \infty$. The classical coupling inequality (see, e.g., Thorisson [24], Chapter 2, Section 3) implies that

$$
\begin{aligned}
|P^n(x, g) - P^n(x', g)| &= |\bar{\mathbb{E}}_{x,x',0}[g(X_n) - g(X_n')]| \\
&\leq \sup_{x\in\mathsf{X}} |g(x)|/f(x) \bar{\mathbb{E}}_{x,x',0}\big[\big(f(X_n) + f(X_n')\big)\mathbb{1}\{d_n = 0\}\big]
\end{aligned}
$$

and (4.1) implies the following key coupling inequality:

$$
\|P^n(x, \cdot) - P^n(x', \cdot)\|_f \leq \check{\mathbb{E}}_{x,x'}\big\{\big(f(X_n) + f(X_n')\big)(1 - \varepsilon)^{N_{n-1}}\big\}. \tag{4.2}
$$

Because $\alpha(u)\beta(v) \leq \rho u + (1 - \rho)v$ for all $(u, v) \in \mathbb{R}^+ \times \mathbb{R}^+$, for any non-negative function f satisfying $f(x) + f(x') \leq \beta \circ V(x, x')$ for all $(x, x') \in \mathsf{X} \times \mathsf{X}$, the coupling inequality (4.2) shows that

$$
\begin{aligned}
&\alpha \circ \{R(n) + M_U\}\|P^n(x, \cdot) - P^n(x', \cdot)\|_f \\
&\quad \leq \alpha \circ \{R(n) + M_U\}\check{\mathbb{E}}_{x,x'}[\{f(X_n) + f(X_n')\}(1 - \varepsilon)^{N_{n-1}}] \\
&\quad \leq \rho\{R(n) + M_U\}\check{\mathbb{E}}_{x,x'}[(1 - \varepsilon)^{N_{n-1}}] + (1 - \rho)\check{\mathbb{E}}_{x,x'}[V(X_n, X_n')(1 - \varepsilon)^{N_{n-1}}].
\end{aligned}
$$

For any $n \geq 0$, let $U_n(x, x') = \check{\mathbb{E}}_{x,x'}[\sum_{k=0}^{\sigma_{C\times C}} r(n + k)]$. It is well known that $\{U_n\}_{n\geq 0}$ satisfies the sequence of drift inequalities

$$
\check{P} U_{n+1} \leq U_n - r(n) + b_U r(n)\mathbb{1}_{C\times C}. \tag{4.3}
$$

Similarly, $\check{P} V \leq V - v + b_V \mathbb{1}_{C\times C}$. Define, for $n \geq 0$,

$$
W_n^{(0)} \stackrel{\text{def}}{=} U_n(X_n, X_n') + \sum_{k=0}^{n-1} r(k) + M_U, \qquad W_n^{(1)} \stackrel{\text{def}}{=} V(X_n, X_n') + \sum_{k=0}^{n-1} v(X_k, X_k') + M_V,
$$

with the convention that $\sum_u^v = 0$ when $u > v$. Since, by construction, for any $n \geq 1$, $W_n^{(0)} \geq R(n)$ and $W_n^{(1)} \geq V(X_n, X_n')$, the previous inequality implies that

$$
\begin{aligned}
&\alpha \circ R(n)\|P^n(x, \cdot) - P^n(x', \cdot)\|_f \\
&\quad \leq \rho\check{\mathbb{E}}_{x,x'}\big[W_n^{(0)}(1 - \varepsilon)^{N_{n-1}}\big] + (1 - \rho)\check{\mathbb{E}}_{x,x'}\big[W_n^{(1)}(1 - \varepsilon)^{N_{n-1}}\big].
\end{aligned}
$$

We must now compute bounds for $\check{\mathbb{E}}_{x,x'}[W_n^{(i)}(1 - \varepsilon)^{N_{n-1}}]$, $i = 0, 1$. Define

$$
\begin{aligned}
T_n^{(0)} &\stackrel{\text{def}}{=} \prod_{i=0}^{n-1} \frac{W_i^{(0)} + b_U r(i)\mathbb{1}_{C\times C}(X_i, X_i')}{W_i^{(0)}} \quad \text{and} \\
T_n^{(1)} &\stackrel{\text{def}}{=} \prod_{i=0}^{n-1} \frac{W_i^{(1)} + b_V \mathbb{1}_{C\times C}(X_i, X_i')}{W_i^{(1)}}.
\end{aligned} \tag{4.4}
$$

If $\varepsilon = 1$, $(1-\varepsilon)^{N_{n-1}} = \mathbb{1}_{\{\sigma_0 \geq n\}}$, where $\sigma_0 = \inf\{n \geq 0 \mid (X_n, X'_n) \in C \times C\}$ is the first hitting time of the set $C \times C$: $T_n^{(i)} \mathbb{1}_{\{\sigma_0 \geq n\}} = \mathbb{1}_{\{\sigma_0 \geq n\}} \leq 1$. Now consider the case $\varepsilon < 1$. By construction, for $N_{n-1} = 0$, $T_n^{(i)} = 1$ and for $N_{n-1} > 0$,

$$T_n^{(0)} = \prod_{i=0}^{N_{n-1}-1} \frac{W_{\sigma_i}^{(0)} + b_U r(\sigma_i)}{W_{\sigma_i}^{(0)}} \quad \text{and} \quad T_n^{(1)} = \prod_{i=0}^{N_{n-1}-1} \frac{W_{\sigma_i}^{(1)} + b_V}{W_{\sigma_i}^{(1)}}, \tag{4.5}$$

where σ_i are the successive hitting times of the set $C \times C$ recursively defined by $\sigma_{j+1} = \inf\{n > \sigma_j \mid (X_n, X'_n) \in C \times C\}$. Because $W_n^{(0)} \geq R(n+1) + M_U$ and $1 + b_U r(n)/\{R(n+1) + M_U\} \leq 1/(1-\varepsilon)$, for $N_{n-1} > 0$, we have

$$T_n^{(0)} (1-\varepsilon)^{N_{n-1}} \leq \prod_{i=0}^{N_{n-1}-1} \left(\left\{ 1 + \frac{b_U r(\sigma_i)}{R(\sigma_i + 1) + M_U} \right\} (1-\varepsilon) \right) \leq 1. \tag{4.6}$$

Similarly, because $W_n^{(1)} \geq M_V$ and $1 + b_V/M_V \leq 1/(1-\varepsilon)$, we have $T_n^{(1)}(1-\varepsilon)^{N_{n-1}} \leq 1$. These two relations imply, for $i = 0, 1$, that

$$\check{\mathbb{E}}_{x,x'}\left[W_n^{(0)}(1-\varepsilon)^{N_{n-1}}\right] \leq \check{\mathbb{E}}_{x,x'}\left[W_n^{(0)}\left\{T_n^{(0)}\right\}^{-1}\right],$$
$$\check{\mathbb{E}}_{x,x'}\left[W_n^{(1)}(1-\varepsilon)^{N_{n-1}}\right] \leq \check{\mathbb{E}}_{x,x'}\left[W_n^{(1)}\left\{T_n^{(1)}\right\}^{-1}\right].$$

It now remains to compute a bound for $\check{\mathbb{E}}_{x,x'}[W_n^{(i)}\{T_n^{(i)}\}^{-1}]$. By construction, we have, for $n \geq 1$,

$$\begin{aligned}
&\check{\mathbb{E}}_{x,x'}\left[W_n^{(0)}\left\{T_n^{(0)}\right\}^{-1} \mid \mathcal{F}_{n-1}\right] \\
&\quad = \check{\mathbb{E}}_{x,x'}\left[W_n^{(0)} \mid \mathcal{F}_{n-1}\right] \frac{W_{n-1}^{(0)}}{W_{n-1}^{(0)} + b_U r(n-1)\mathbb{1}_{C \times C}(X_{n-1}, X'_{n-1})} \left\{T_{n-1}^{(0)}\right\}^{-1},
\end{aligned} \tag{4.7}$$

where $\mathcal{F}_n = \sigma\{(X_0, X'_0), \ldots, (X_n, X'_n)\}$. Now, (4.3) yields

$$\check{\mathbb{E}}_{x,x'}\left[W_n^{(0)} \mid \mathcal{F}_{n-1}\right] \leq W_{n-1}^{(0)} + b_U r(n-1)\mathbb{1}_{C \times C}(X_{n-1}, X'_{n-1}). \tag{4.8}$$

Combining (4.7) and (4.8) shows that $\{W_n^{(0)}\{T_n^{(0)}\}^{-1}\}_{n \geq 0}$ is an $\mathcal{F}$-supermartingale. Thus,

$$\check{\mathbb{E}}_{x,x'}\left[W_n^{(0)}(1-\varepsilon)^{N_{n-1}}\right] \leq \check{\mathbb{E}}_{x,x'}\left[W_n^{(0)}\left\{T_n^{(0)}\right\}^{-1}\right] \leq \check{\mathbb{E}}_{x,x'}\left[W_0^{(0)}\right] = U_0(x, x') + M_U.$$

Similarly, $\check{\mathbb{E}}_{x,x'}[W_n^{(1)}(1-\varepsilon)^{N_{n-1}}] \leq V(x, x') + M_V$, which concludes the proof of Theorem 2.1. □

5. Proof of Proposition 2.2, Theorem 2.3

Proof of Proposition 2.2. By applying the comparison theorem (Meyn and Tweedie [15]) and (Douc *et al.* [3], Proposition 2.2) we obtain the following inequalities. For all $(x,x') \in \mathsf{X}\times\mathsf{X}$,

$$\check{\mathbb{E}}_{x,x'}\left[\sum_{k=0}^{\tau_{C\times C}-1} \phi\circ H_\phi^{-1}(k)\right] \le W(x,x') - 1 + b\frac{\phi\circ H_\phi^{-1}(1)}{\phi\circ H_\phi^{-1}(0)}\mathbb{1}_{C\times C}(x,x'), \tag{5.1}$$

$$\check{\mathbb{E}}_{x,x'}\left[\sum_{k=0}^{\tau_{C\times C}-1} \phi\circ W(X_k, X'_k)\right] \le W(x,x') + b\mathbb{1}_{C\times C}(x,x'). \tag{5.2}$$

The sequence $\{\phi\circ H_\phi^{-1}(k)\}_{k\ge 0}$ is log-concave. Therefore, for any $k\ge 0$, $\phi\circ H_\phi^{-1}(k+1)/\phi\circ H_\phi^{-1}(k) \le \phi\circ H_\phi^{-1}(1)/\phi\circ H_\phi^{-1}(0)$. Then, applying (5.1), we obtain

$$\begin{aligned}
&\check{\mathbb{E}}_{x,x'}\left[\sum_{k=0}^{\sigma_{C\times C}} \phi\circ H_\phi^{-1}(k)\right]\\
&\quad= \phi\circ H_\phi^{-1}(0) + \check{\mathbb{E}}_{x,x'}\left[\sum_{k=1}^{\tau_{C\times C}} \phi\circ H_\phi^{-1}(k)\right]\mathbb{1}_{(C\times C)^c}(x,x')\\
&\quad\le \phi\circ H_\phi^{-1}(0) + \frac{\phi\circ H_\phi^{-1}(1)}{\phi\circ H_\phi^{-1}(0)}\check{\mathbb{E}}_{x,x'}\left[\sum_{k=1}^{\tau_{C\times C}} \phi\circ H_\phi^{-1}(k-1)\right]\mathbb{1}_{(C\times C)^c}(x,x'),
\end{aligned}$$

showing (2.16). The proof of (2.17) is along the same lines. □

Proof of Theorem 2.3. Since $d_0 = \inf_{x\notin C} W_0(x)$, if $(x,x')\notin C\times C$, then $W(x,x')\ge d_0$ and $\mathbb{1}_C(x)+\mathbb{1}_C(x')\le 1$ since either $x\notin C$, $x'\notin C$ (or both). The definition of the kernel $\check{P}$ therefore implies that

$$\begin{aligned}
\check{P}W(x,x') &\le W_0(x)+W_0(x') - 1 - \phi_0\circ W_0(x') - \phi_0\circ W_0(x') + b_0\{\mathbb{1}_C(x)+\mathbb{1}_C(x')\}\\
&\le W(x,x') - \phi_0\circ W(x,x') + b_0,
\end{aligned}$$

where we have used the following inequality: for any $u\ge 1$ and $v\ge 1$, $\phi_0(u+v-1)-\phi_0(u)\le \phi_0(v)-\phi_0(1)$. For $(x,x')\notin C$, $b_0\le (1-\lambda)\phi_0(d)\le (1-\lambda)\phi_0\circ W(x,x')$ and the previous inequality implies that $\check{P}W(x,x')\le W(x,x')-\phi\circ W(x,x')$. □

Appendix A: Proof of Proposition 3.1

Let W be any measurable non-negative function on X. Then, for $\eta>0$ and $x\notin C_\eta$, we have

$$PW(x) - W(x) \le \int_{\mathsf{X}}\left(\eta\wedge\frac{q(y)}{\pi(y)}\right)W(y)\pi(y)\mu(\mathrm{d}y) - W(x)\int_{\mathsf{X}} a(x,y)q(y)\mu(\mathrm{d}y).$$

If $x \notin C_\eta$ and $y \in C_\eta$, then $y \preceq x$ and $a(x, y)q(y) = (q(x)/\pi(x))\pi(y)$. Thus,

$$\int_{\mathsf{X}} a(x, y)q(y)\mu(\mathrm{d}y) \ge \int_{C_\eta} a(x, y)q(y)\mu(\mathrm{d}y) = \frac{q(x)}{\pi(x)}\pi(C_\eta) = \frac{q(x)}{\pi(x)}\bigl(1 - \psi(\eta)\bigr).$$

Altogether, we obtain, for all $x \notin C_\eta$,

$$PW(x) - W(x) \le \int_{\mathsf{X}} \left(\eta \wedge \frac{q(y)}{\pi(y)}\right) W(y)\pi(y)\mu(\mathrm{d}y) - \{1 - \psi(\eta)\}\frac{q(x)}{\pi(x)} W(x). \tag{A.1}$$

The definition of W_0 implies that

$$\int_{\mathsf{X}} \left(\eta \wedge \frac{q(y)}{\pi(y)}\right) W_0(y)\pi(y)\mu(\mathrm{d}y) = \int_0^\infty (\eta \wedge u)K(u)\,\mathrm{d}\psi(u) < \infty.$$

By Lebesgue's bounded convergence theorem, $\lim_{\eta\to 0} \int_0^\infty (\eta \wedge u)K(u)\,\mathrm{d}\psi(u) = 0$. Since, moreover, $\lim_{\eta\to 0} \psi(\eta) = 0$, it follows that for sufficiently small η, $\{1 - \psi(\eta)\}\phi(M) > \int_0^\infty (\eta \wedge u)K(u)\,\mathrm{d}\psi(u)$, hence $\eta^\star$ is well defined. Now, (A.1) and (A) yield, for all $x \notin C_{\eta^\star}$,

$$\begin{aligned} PW_0(x) - W_0(x) &\le \int_0^\infty (\eta^\star \wedge u)K(u)\,\mathrm{d}\psi(u) - \bigl(1 - \psi(\eta^\star)\bigr) W_0(x)K^{-1} \circ W_0(x) \\ &= -\phi_0(W_0(x)). \end{aligned}$$

For $x \in C_{\eta^\star}$, we have $W_0(x) \le K(\eta^\star)$. Finally, we have, for any $x \in C_{\eta^\star}$,

$$\begin{aligned} PW_0(x) &\le \int_{\mathsf{X}} q(y)W_0(y)\mu(\mathrm{d}y) + W_0(x) \\ &= \int_{\mathsf{X}} \frac{q(y)}{\pi(y)} K\left(\frac{q(y)}{\pi(y)}\right)\pi(y)\mu(\mathrm{d}y) + W_0(x) \\ &\le \int_0^\infty uK(u)\,\mathrm{d}\psi(u) + K(\eta^\star). \end{aligned}$$

References

[1] Baxendale, P.H. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.* **15** 700–738. MR2114987

[2] Dai, J.G. and Meyn, S.P. (1995). Stability and convergence of moments for multiclass queueing networks via fluid limit models. *IEEE Trans. Automat. Control* **40** 1889–1904. MR1358006

[3] Douc, R., Fort, G., Moulines, E. and Soulier, P. (2004a). Practical drift conditions for subgeometric rates of convergence. *Ann. Appl. Probab.* **14** 1353–1377. MR2071426

[4] Douc, R., Moulines, E. and Rosenthal, J. (2004b). Quantitative bounds for geometric convergence rates of Markov chains. *Ann. Appl. Probab.* **14** 1643–1665. MR2099647

[5] Fort, G. (2001). Contrôle explicite d'ergodicité de chaînes de Markov: applications à l'analyse de convergence de l'algorithme Monte-Carlo EM. Ph.D. thesis, Université de Paris VI. INIST Number T139824

[6] Fort, G. and Moulines, E. (2000). V-subgeometric ergodicity for a Hastings–Metropolis algorithm. *Statist. Probab. Lett.* **49** 401–410. MR1796485

[7] Fort, G. and Moulines, E. (2003). Polynomial ergodicity of Markov transition kernels. *Stochastic Process. Appl.* **103** 57–99. MR1947960

[8] Gulinsky, O.V. and Veretennikov, A.Y. (1993). *Large Deviations for Discrete-Time Processes with Averaging*. Utrecht: VSP. MR1360713

[9] Jarner, S.F. and Roberts, G.O. (2002). Polynomial convergence rates of Markov chains. *Ann. Appl. Probab.* **12** 224–247. MR1890063

[10] Klokov, S.A. and Veretennikov, A.Y. (2004). Sub-exponential mixing rate for a class of Markov chains. *Math. Commun.* **9** 9–26. MR2076227

[11] Lindvall, T. (1979). On coupling of discrete renewal sequences. *Z. Wahrsch. Verw. Gebiete* **48** 57–70. MR0533006

[12] Lindvall, T. (1992). *Lectures on the Coupling Method*. New York: Wiley. MR1180522

[13] Lund, R.B., Meyn, S.P. and Tweedie, R. (1996). Computable exponential convergence rates for stochastically ordered Markov processes. *Ann. Appl. Probab.* **6** 218–237. MR1389838

[14] Lund, R.B. and Tweedie, R.L. (1996). Geometric convergence rates for stochastically ordered Markov chains. *Math. Oper. Res.* **21** 182–194. MR1385873

[15] Meyn, S.P. and Tweedie, R.L. (1993). *Markov Chains and Stochastic Stability*. London: Springer. MR1287609

[16] Meyn, S.P. and Tweedie, R.L. (1994). Computable bounds for convergence rates of Markov chains. *Ann. Appl. Probab.* **4** 981–1011. MR1304770

[17] Nummelin, E. and Tuominen, P. (1983). The rate of convergence in Orey's theorem for Harris recurrent Markov chains with applications to renewal theory. *Stochastic Process. Appl.* **15** 295–311. MR0711187

[18] Roberts, G.O. and Rosenthal, J.S. (2004). General state space Markov chains and MCMC algorithms. *Probab. Surv.* **1** 20–71. MR2095565

[19] Roberts, G.O. and Tweedie, R.L. (1999). Bounds on regeneration times and convergence rates for Markov chains. *Stochastic Process. Appl.* **80** 211–229. MR1682243

[20] Roberts, G.O. and Tweedie, R.L. (2000). Rates of convergence of stochastically monotone and continuous time Markov models. *J. Appl. Probab.* **37** 359–373. MR1780996

[21] Rosenthal, J.S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **90** 558–566. MR1340509

[22] Roughan, M., Veitch, D. and Rumsewicz, M. (1998). Computing queue-length distributions for power-law queues. In *Proceedings IEEE INFOCOM '98* **1** 356–363. IEEE.

[23] Scott, D.J. and Tweedie, R.L. (1996). Explicit rates of convergence of stochastically ordered Markov chains. *Athens Conference on Applied Probability and Time Series Analysis* **1**. *Applied Probability in Honor of J. M. Gani. Lecture Notes in Statist.* **114** 176–191. New York: Springer. MR1466715

[24] Thorisson, H. (2000). *Coupling, Stationarity and Regeneration*. New York: Springer. MR1741181

[25] Tuominen, P. and Tweedie, R. (1994). Subgeometric rates of convergence of f-ergodic Markov chains. *Adv. in Appl. Probab.* **26** 775–798. MR1285459

[26] Veretennikov, A. (1997). On polynomial mixing bounds for stochastic differential equations. *Stochastic Process. Appl.* **70** 115–127. MR1472961

[27] Veretennikov, A. (2000). On polynomial mixing and the rate of convergence for stochastic differential and difference equations. *Theory Probab. Appl.* **44** 361–374. MR1751475

Received November 2005 and revised February 2007